

CIRCULAR RNA SYNTHESIS

CROSS REFERENCE TO RELATED APPLICATIONS

[0001] This PCT application claims the priority benefit of U.S. Provisional Application No. 63/601,185, filed November 20, 2023; the entire disclosure is incorporated herein by reference for all purposes.

REFERENCE TO SEQUENCE LISTING SUBMITTED ELECTRONICALLY VIA EFS-WEB

[0002] The content of the sequence listing is submitted electronically (Name: ; Size: bytes; and Date of Creation:) and is filed with the application is herein incorporated by reference in its entirety.

TECHNICAL FIELD

[0003] The disclosure relates to methods and compositions for the production of circular RNAs (circRNAs).

BACKGROUND

[0004] Circular RNAs (circRNAs) are covalently closed single-stranded RNA molecules lacking free 5' and 3' ends, and therefore lacking a 5' cap and a 3' polyA tail.

[0005] In nature, circRNAs have a wide range of origins and tissue specificity and play a variety of roles in the development and homeostasis of organisms, in aging, and have also been linked to the occurrence and development of various human diseases (*e.g.*, cardiovascular diseases, diabetes mellitus, and cancer).

[0006] circRNAs are more stable than linear RNAs, do not need a 5' cap or a 3' poly(A) tail to keep them stable, and are not easily degraded by RNA exonucleases (Petkovic, S., and Müller, S. (2013). RNA Self-Processing: Formation of Cyclic Species and Concatemers from a Small Engineered RNA. *FEBS Lett.* 587, 2435–2440.), making circRNAs great candidates as therapeutics.

[0007] Nevertheless, the use of circRNAs as therapeutics has been hampered by technical limitation in their synthesis.

[0008] Linear RNA can be produced by chemical synthesis or enzymatic strategy, and circRNAs can be produced by ligating the ends of linear RNA precursors to produce a covalently closed circle. The use of chemical synthesis is limited by the high cost of purification and low

yield, and can only produce RNAs of less than 50 to 70 nucleotides in length. Enzymatic strategy is usually realized through an *in vitro* transcription (IVT) reaction, which includes a DNA template, a reaction buffer, and a phage RNA polymerase. IVT reaction allows for longer RNA synthesis at a lower cost. However, the run-off nature of phage polymerases may result in incomplete RNA.

[0009] Ribozyme methods include group I intron self-splicing systems, also called permuted introns and exons (PIE) method. Due to the complexity of RNA secondary structure, different RNA sequences lead to differences in the splicing efficiency and therefore yield of the final circRNA, which limits the application of the PIE method. Group II introns can also be used for circRNA synthesis, which involves an inverse splicing reaction. This method forms 2', 5'-phosphodiester bonds at the ligation site instead of the natural 3', 5'-phosphodiester bonds, and the mechanism remains unclear. Hairpin ribozyme (HPR) can produce circRNAs through rolling circle reaction and the self-splicing reaction from circular single-strand DNA template. However, the HPR method can efficiently produce only small circRNAs and suffers from unstable product and exogenous HPR sequences.

[0010] Additionally, the high immunogenicity of the synthesized circRNAs (Chen, Y.G. et al., (2017). Sensing self and foreign circular RNAs by intron identity. *Mol. Cel.* 67, 228–238.e5; Chen, Y.G., et al. (2019). N6-methyladenosine modification controls circular RNA immunity. *Mol. Cel.* 76, 96–109.e), further hinders the applicability of circRNAs as therapeutic molecules.

[0011] Given the strong therapeutic potential of circRNAs, there is an immediate need for methods and compositions for producing circRNAs with high efficiency, and for producing circRNAs having low immunogenicity.

SUMMARY

[0012] Aspects of the present disclosure address the above-referenced problems and/or others. In some aspects, the disclosure provides a genetic construct comprising a formula of:

5'-H-A-E-B-H'-3',

wherein H comprise a 5' homology arm, A comprises a 3' intronic sequence, E comprises an exonic sequence, B comprises a 5' intronic sequence comprising an internal guide sequence (IGS), and H' comprises a 3' homology arm, wherein the 3' intronic sequence comprises a catalytic core, wherein the 3' terminus of A and the 5' terminus of E form a 3' splice site and the 3' terminus of E and 5' terminus of B form a 5' splice site, and wherein the genetic construct is capable of forming a circular RNA that does not contain any exonic sequence from the genome from which the 5' intronic sequence and the 3' intronic sequence are derived ("excision scar").

[0013] In some aspects, the 5' intronic sequence and/or the 3' intronic sequence are derived from a group I intron. In some aspects, the 3' intronic sequence and the 5' intronic sequence together are less than the full intron.

[0014] In some aspects, the 5' intronic sequence and/or the 3' intronic sequence are derived from an intron of *thymidylate synthase* (td) from T4 bacteriophage; of pre-tRNA^{LEU} of the cyanobacteria *Anabaena azollae*, *Anabaena* PCC7120, *Aphanizomenon flos-aquae* strain NIVA-CYA 142, or of *Synechococcus*; of pre-tRNA^{ILE} of the nitrogen-fixing bacteria *Azoarcus*; of 26S ribosomal RNA from *Tetrahymena thermophila*; of *COX2* gene from *Trimorphomyces papilionaceus*; or any combination thereof.

[0015] In some aspects, the 5' intronic sequence and/or the 3' intronic sequence are derived from an intron of *thymidylate synthase* (td) from T4 bacteriophage. In some aspects, the 5' splice site comprises a site between two nucleotides comprising U and N, where N is any nucleotide. In some aspects, the N is A, T, G, or C. In some aspects, the 3' splice site comprises a site between two nucleotides comprising G and N, wherein N is any nucleotide. In some aspects, the N is C. In some aspects, the catalytic core comprises ACG, CU, or both.

[0016] In some aspects, the excision scar comprises SEQ ID NO: 80 or SEQ ID NO: 82. In some aspects, the excision scar comprises SEQ ID NO: 80 and SEQ ID NO: 82. In some aspects, the excision scar comprises SEQ ID NO: 81 or SEQ ID NO: 83. In some aspects, the excision scar comprises SEQ ID NO: 81 and SEQ ID NO: 83.

[0017] In some aspects, the 3' intronic sequence comprises a nucleotide sequence having at least about 70%, at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, at least about 99%, or about 100% sequence identity to a 3' intronic sequence selected from Table 4 (SEQ ID NOs: 11 – 22). In some aspects, the 5' intronic sequence comprises a nucleotide sequence having at least about 70%, at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 95%, at least about 95%, at least about 97%, at least about 98%, at least about 99%, or about 100% sequence identity to a 5' intronic sequence selected from Table 4 (SEQ ID NOs: 23-30).

[0018] In some aspects, the 5' intronic sequence and/or the 3' intronic sequence are derived from an intron of *tRNA^{LEU}* of the cyanobacteria *Anabaena azollae*. In some aspects, the 5' splice site comprises a site between two nucleotides comprising U and N, wherein N is any nucleotide. In some aspects, the N is A. In some aspects, the 3' splice site comprises a site between two nucleotides comprising G and N, wherein N is any nucleotide. In some aspects, the N is C. In some aspects, the catalytic core comprises GAG, CC, or both. In some aspects, the 3' intronic sequence

comprises a nucleotide sequence having at least about 70%, at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 95%, at least about 95%, at least about 97%, at least about 98%, at least about 99%, or about 100% sequence identity with a 3' intronic sequence in Table 4, for example, SEQ ID NO: 14 or 20. In some aspects, the 5' intronic sequence comprises a nucleotide sequence having at least about 70%, at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 95%, at least about 95%, at least about 97%, at least about 98%, at least about 99%, or about 100% sequence identity with a 5' intronic sequence in Table 4, for example, SEQ ID NO: 25 or 29.

[0019] In some aspects, the 5' intronic sequence and the 3' intronic sequence are derived from an intron of *tRNA^{LEU}* of *Azoarcus*. In some aspects, the 5' splice site comprises a site between two nucleotides comprising U and N, wherein the N is any nucleotide. In some aspects, the N is A. In some aspects, the 3' splice site comprises a site between two nucleotides comprising G and N, wherein N is any nucleotide. In some aspects, the N is G. In some aspects, the catalytic core comprises GAG, CC, or both. In some aspects, the 3' intronic sequence comprises a nucleotide sequence having at least about 70%, at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 95%, at least about 95%, at least about 97%, at least about 98%, at least about 99%, or about 100% sequence identity with a 3' intronic sequence in Table 4, for example, SEQ ID NO: 13 or 19. In some aspects, the 5' intronic sequence comprises a nucleotide sequence having at least about 70%, at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 95%, at least about 95%, at least about 97%, at least about 98%, at least about 99%, or about 100% sequence identity with a 5' intronic sequence in Table 4, for example, nucleotide sequence AAA.

[0020] In some aspects, the 5' intronic sequence and the 3' intronic sequence are derived from an intron of 26S ribosomal RNA from *Tetrahymena thermophila*. In some aspects, the 5' splice site comprises a site between two nucleotides comprising U and N, wherein N is any nucleotide. In some aspects, the N is A. In some aspects, the 3' splice site comprises a site between two nucleotides comprising G and N, wherein the N is any nucleotide. In some aspects, the N is U. In some aspects, the catalytic core comprises a catalytic core selected from Table 5. In some aspects, the 3' intronic sequence comprises a nucleotide sequence having at least about 70%, at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 95%, at least about 95%, at least about 97%, at least about 98%, at least about 99%, or about 100% sequence identity with a 3' intronic sequence in Table 4, for example, SEQ ID NO: 12 or 18. In some aspects, the 5' intronic sequence comprises a nucleotide sequence having at least about 70%,

at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 95%, at least about 95%, at least about 97%, at least about 98%, at least about 99%, or about 100% sequence identity with a 5' intronic sequence in Table 4, for example, SEQ ID NO: 24 or 28.

[0021] In some aspects, the disclosure provides a genetic construct comprising a formula of: 5'-H-A-E-B-H'-3',

wherein H comprise a 5' homology arm, A comprises a 3' intronic sequence, E comprises an exonic sequence, B comprises a 5' intronic sequence comprising an internal guide sequence (IGS), and H' comprises a 3' homology arm, wherein the 3' intronic sequence comprises a catalytic core, wherein the 3' terminus of A and the 5' terminus of E form a 3' splice site and the 3' terminus of E and 5' terminus of B form a 5' splice site, wherein the genetic construct is capable of forming a circular RNA that does not contain any exonic sequence from the genome from which the 5' intronic sequence and the 3' intronic sequence are derived ("excision scar"), and wherein the genetic construct comprises a nucleotide sequence having at least about 70%, at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 95%, at least about 95%, at least about 97%, at least about 98%, at least about 99%, or about 100% sequence identity with SEQ ID NO: 113 or 123.

[0022] In some aspects, the disclosure provides a genetic construct comprising a formula of: 5'-H-A-E-B-H'-3',

wherein H comprise a 5' homology arm, A comprises a 3' intronic sequence, E comprises an exonic sequence, B comprises a 5' intronic sequence comprising an internal guide sequence (IGS), and H' comprises a 3' homology arm, wherein the 3' intronic sequence comprises a catalytic core, wherein the 3' terminus of A and the 5' terminus of E form a 3' splice site and the 3' terminus of E and 5' terminus of B form a 5' splice site, wherein the genetic construct is capable of forming a circular RNA that does not contain any exonic sequence from the genome from which the 5' intronic sequence and the 3' intronic sequence are derived ("excision scar"), and wherein the genetic construct comprises a nucleotide sequence having at least about 70%, at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 95%, at least about 95%, at least about 97%, at least about 98%, at least about 99%, or about 100% sequence identity with SEQ ID NO: 114 or 124.

[0023] In some aspects, the disclosure provides a genetic construct comprising a formula of: 5'-H-A-E-B-H'-3',

wherein H comprise a 5' homology arm, A comprises a 3' intronic sequence, E comprises an exonic sequence, B comprises a 5' intronic sequence comprising an internal guide sequence (IGS), and

H' comprises a 3' homology arm, wherein the 3' intronic sequence comprises a catalytic core, wherein the 3' terminus of A and the 5' terminus of E form a 3' splice site and the 3' terminus of E and 5' terminus of B form a 5' splice site, wherein the genetic construct is capable of forming a circular RNA that does not contain any exonic sequence from the genome from which the 5' intronic sequence and the 3' intronic sequence are derived ("excision scar"), and wherein the genetic construct comprises a nucleotide sequence having at least about 70%, at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 95%, at least about 95%, at least about 97%, at least about 98%, at least about 99%, or about 100% sequence identity with SEQ ID NO: 115 or 125.

[0024] In some aspects, the disclosure provides a genetic construct comprising a formula of: 5'-H-A-E-B-H'-3',

wherein H comprise a 5' homology arm, A comprises a 3' intronic sequence, E comprises an exonic sequence, B comprises a 5' intronic sequence comprising an internal guide sequence (IGS), and H' comprises a 3' homology arm, wherein the 3' intronic sequence comprises a catalytic core, wherein the 3' terminus of A and the 5' terminus of E form a 3' splice site and the 3' terminus of E and 5' terminus of B form a 5' splice site, wherein the RNA construct is capable of forming a circular RNA that does not contain any exonic sequence from the genome from which the 5' intronic sequence and the 3' intronic sequence are derived ("excision scar"), and wherein the genetic construct comprises a nucleotide sequence having at least about 70%, at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 95%, at least about 95%, at least about 97%, at least about 98%, at least about 99%, or about 100% sequence identity with SEQ ID NO: 116 or 126.

[0025] In some aspects, the disclosure provides a genetic construct comprising a formula of: 5'-H-A-E-B-H'-3',

wherein H comprise a 5' homology arm, A comprises a 3' intronic sequence, E comprises an exonic sequence, B comprises a 5' intronic sequence comprising an internal guide sequence (IGS), and H' comprises a 3' homology arm, wherein the 3' intronic sequence comprises a catalytic core, wherein the 3' terminus of A and the 5' terminus of E form a 3' splice site and the 3' terminus of E and 5' terminus of B form a 5' splice site, wherein the genetic construct is capable of forming a circular RNA that does not contain any exonic sequence from the genome from which the 5' intronic sequence and the 3' intronic sequence are derived ("excision scar"), and wherein the genetic construct comprises a nucleotide sequence having at least about 70%, at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 95%, at least about 95%,

at least about 97%, at least about 98%, at least about 99%, or about 100% sequence identity with SEQ ID NO: 117 or 127.

[0026] In some aspects, the disclosure provides a genetic construct comprising a formula of 5'-H-A-E-B-H'-3',

wherein H comprises a 3' homology arm, A comprises a 3' homology arm, a 3' intronic sequence, E comprise an exonic sequence, B comprises a 5' intronic sequence comprising an internal guide sequence, and H' comprising a 5' homology arm, wherein the 3' intronic sequence comprises a catalytic core, wherein the 3' terminus of A and the 5' terminus of E form a 3' splice site and the 3' terminus of E and 5' terminus of B form a 5' splice site, and wherein the 5' intronic sequence and/or the 3' intronic sequence are derived from pre-tRNA^{LEU} of the cyanobacteria *Anabaena azollae*, cyanobacteria *Azoarcus*, or *Synechococcus*; or from 26S ribosomal RNA of *Tetrahymena thermophila*, and are capable of forming a circular RNA.

[0027] In some aspects, the 5' intronic sequence and/or the 3' intronic sequences are derived from tRNA^{LEU} of the cyanobacteria *Anabaena azollae*. In some aspects, the 5' splice site comprises a site between two nucleotides comprising U and any nucleotide (N), optionally A. In some aspects, the 3' splice site comprises a site between two nucleotides comprising G and any nucleotide (N), optionally C. In some aspects, the catalytic core comprises GAG, CC, or both. In some aspects, the 5' intronic sequence comprises a nucleotide sequence having at least about 70%, at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 95%, at least about 95%, at least about 97%, at least about 98%, at least about 99%, or about 100% sequence identity with SEQ ID NO: 25 or 29. In some aspects, the 3' intronic sequence comprises a nucleotide sequence having at least about 70%, at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 95%, at least about 95%, at least about 97%, at least about 98%, at least about 99%, or about 100% sequence identity with SEQ ID NO: 14 or 20.

[0028] In some aspects, the 5' intronic sequence and/or the 3' intronic sequences are derived from tRNA^{LEU} of the cyanobacteria *Azoarcus*. In some aspects, the 5' splice site comprises a site between two nucleotides comprising U and any nucleotide, optionally A. In some aspects, the 3' splice site comprises a site between two nucleotides comprising G and any nucleotide, optionally C. In some aspects, the catalytic core comprises GAG, CC, or both. In some aspects, the 5' intronic sequence comprises a nucleotide sequence of AAA. In some aspects, the 3' intronic sequence comprises a nucleotide sequence having at least about 70%, at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 95%, at least about 95%, at least

about 97%, at least about 98%, at least about 99%, or about 100% sequence identity with SEQ ID NO: 13 or 19.

[0029] In some aspects, the 5' intronic sequence and the 3' intronic sequences are derived from tRNA^{LEU} of *Synechococcus*. In some aspects, the 5' splice site comprises a site between two nucleotides comprising U and any nucleotide (N), optionally A. In some aspects, the 3' splice site comprises a site between two nucleotides comprising G and any nucleotide (N), optionally C. In some aspects, the catalytic core comprises GAG, CC, or both. In some aspects, the 5' intronic sequence comprises a nucleotide sequence having at least about 70%, at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 95%, at least about 95%, at least about 97%, at least about 98%, at least about 99%, or about 100% sequence identity to a 5' intronic sequence derived from *Synechococcus* intron as set forth in SEQ ID NO: 140 or 143. In some aspects, the 3' intronic sequence comprises a nucleotide sequence having at least about 70%, at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 95%, at least about 95%, at least about 97%, at least about 98%, at least about 99%, or about 100% sequence identity to a 5' intronic sequence derived from *Synechococcus* intron as set forth in SEQ ID NO: 140 or 143.

[0030] In some aspects, the 5' intronic sequence and the 3' intronic sequences are derived from 26S ribosomal RNA of *Tetrahymena thermophila*. In some aspects, the 5' splice site comprises a site between two nucleotides comprising U and any nucleotide (N), optionally A. In some aspects, the 3' splice site comprises a site between two nucleotides comprising G and any nucleotide (N), optionally C. In some aspects, the catalytic core comprises GAG, CC, or both. In some aspects, the 5' intronic sequence comprises a nucleotide sequence having at least about 70%, at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 95%, at least about 95%, at least about 97%, at least about 98%, at least about 99%, or about 100% sequence identity with SEQ ID NO: 24 or 28. In some aspects, the 3' intronic sequence comprises a nucleotide sequence having at least about 70%, at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 95%, at least about 95%, at least about 97%, at least about 98%, at least about 99%, or about 100% sequence identity with SEQ ID NO: 12 or 18.

[0031] In some aspects, the circular RNA does not comprise an excision scar. In some aspects, the circular RNA has a reduced immunogenicity. In some aspects, the circular RNA has a reduced immunogenicity compared to a circular RNA that contains the excision scar.

[0032] In some aspects, the disclosure provides a genetic construct comprising a formula of 5'-H-A-E-B-H'-3',

wherein H comprises a 3' homology arm, A comprises a 3' homology arm, a 3' intronic sequence, E comprise an exonic sequence, B comprises a 5' intronic sequence comprising an internal guide sequence, and H' comprising a 5' homology arm, wherein the 3' intronic sequence comprises a catalytic core, wherein the 3' terminus of A and the 5' terminus of E form a 3' splice site and the 3' terminus of E and 5' terminus of B form a 5' splice site, and wherein the 5' intronic sequence and/or the 3' intronic sequence are derived from pre-tRNA^{LEU} of the cyanobacteria *Anabaena azollae*, cyanobacteria *Azoarcus*, or *Synechococcus* or from 26S ribosomal RNA of *Tetrahymena thermophila* and are capable of forming a circular RNA and wherein the genetic construct comprises a nucleotide sequence having at least about 70%, at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 95%, at least about 95%, at least about 97%, at least about 98%, at least about 99%, or about 100% sequence identity with SEQ ID NO: 118 or 128.

[0033] In some aspects, the disclosure provides a genetic construct comprising a formula of: 5'-A-E-3',

wherein A comprises a 3' intronic sequence, E comprises an exonic sequence, wherein the 3' intronic sequence comprises a catalytic core, wherein the 3' terminus of A and the 5' terminus of E form a 3' splice site, and wherein the genetic construct is capable of forming a circular RNA.

[0034] In some aspects, the disclosure provides a genetic construct comprising a formula of: 5'-A-E-3',

wherein A comprises a 3' intronic sequence, E comprises an exonic sequence, wherein the 3' intronic sequence comprises a catalytic core, wherein the 3' terminus of A and the 5' terminus of E form a 3' splice site, wherein the 3' terminus of A and the 5' terminus of E form a 3' splice site, wherein the genetic construct comprises a nucleotide sequence having at least about 70%, at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 95%, at least about 95%, at least about 97%, at least about 98%, at least about 99%, or about 100% sequence identity with SEQ ID NO: 119 or 129, and wherein the genetic construct is capable of forming a circular RNA.

[0035] In some aspects, the disclosure provides a genetic construct comprising a formula of 5'-A-E-3',

wherein A comprises a 3' intronic sequence comprising a catalytic core (CC), and E comprises an exonic sequence comprising a first portion of the exonic sequence (E'); a first homology arm (HI'); a second portion of the exonic sequence (E''); a second homology arm (HI''); and a third exonic sequence (E''') comprising a catalytic nucleotide (CN); wherein the 3' terminus of A and the 5'

terminus of E form a 3' splice site, wherein the genetic construct comprises a nucleotide sequence having at least about 70%, at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 95%, at least about 95%, at least about 97%, at least about 98%, at least about 99%, or about 100% sequence identity with SEQ ID NO: 120 or 130, and wherein the genetic construct is capable of forming a circular RNA.

[0036] In some aspects, the disclosure provides a genetic construct comprising a formula of 5'-A-E-3',

wherein A comprises a 3' intronic sequence comprising a catalytic core (CC), and E comprises an exonic sequence comprising a first portion of the exonic sequence (E'); a first homology arm (HI'); a second portion of the exonic sequence (E''); a second homology arm (HI''); and a third exonic sequence (E''') comprising a catalytic nucleotide (CN); wherein the 3' terminus of A and the 5' terminus of E form a 3' splice site, wherein the genetic construct comprises a nucleotide sequence having at least about 70%, at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 95%, at least about 95%, at least about 97%, at least about 98%, at least about 99%, or about 100% sequence identity with SEQ ID NO: 121 or 131, and wherein the genetic construct is capable of forming a circular RNA.

[0037] In some aspects, the disclosure provides a genetic construct comprising a formula of 5'-A-E-3',

wherein H comprises a 3' homology arm, A comprises a 3' intronic sequence comprising a catalytic core (CC), and E comprises an exonic sequence comprising a first portion of the exonic sequence (E'); a first homology arm (HI'); a second portion of the exonic sequence (E''); a second homology arm (HI''); and a third exonic sequence (E''') comprising a catalytic nucleotide (CN); wherein the 3' terminus of A and the 5' terminus of E form a 3' splice site, wherein the genetic construct comprises a nucleotide sequence having at least about 70%, at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 95%, at least about 95%, at least about 97%, at least about 98%, at least about 99%, or about 100% sequence identity with SEQ ID NO: 122 or 132, and wherein the genetic construct is capable of forming a circular RNA.

[0038] In some aspects, the disclosure provides a genetic construct comprising a formula of 5'-A-E-3',

wherein A comprises a 3' intronic sequence comprising a catalytic core (CC), and E comprises an exonic sequence comprising a first portion of the exonic sequence (E'); a first homology arm (HI'); a second portion of the exonic sequence (E''); a second homology arm (HI''); and a third exonic sequence (E''') comprising a catalytic nucleotide (CN); wherein the genetic construct comprises a

nucleotide sequence having at least about 70%, at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 95%, at least about 95%, at least about 97%, at least about 98%, at least about 99%, or about 100% sequence identity with SEQ ID NO: 181 and wherein the genetic construct is capable of forming a circular RNA.

[0039] In some aspects, the disclosure provides a genetic construct comprising a formula of 5'-B-A-E-3',

wherein A comprises a 3' intronic sequence, 3S comprises a 3' splice site, E comprises an exonic sequence, B comprises a 5' intronic sequence, and 5S comprises a 5' splice site, wherein the 3' intronic sequence comprises a catalytic core, wherein B-A comprise the full intron, wherein the genetic construct comprises a nucleotide sequence having at least about 70%, at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 95%, at least about 95%, at least about 97%, at least about 98%, at least about 99%, or about 100% sequence identity with SEQ ID NO: 182 or 183, and wherein the genetic construct is capable of forming a circular RNA.

[0040] In some aspects, the genetic construct is RNA. In some aspects, the genetic construct is DNA.

[0041] In some aspects, the genetic construct further comprises a promoter. In some aspects, the promoter is a T7 promoter, a T3 promoter, or a Sp6 promoter.

[0042] In some aspects, the genetic construct further comprises an IRES.

[0043] In some aspects, the genetic construct further comprises a binding site for a polyA binding protein at the 5' terminus.

[0044] In some aspects, the 5' homology arm comprises a nucleotide sequence at least about 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% identical to SEQ ID NO: 72 (CCGTCGATTGTCCACTGGT). In some aspects, the 3' homology arm comprises a nucleotide sequence at least about 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% identical to SEQ ID NO: 73 (ACCAGTGGACAATCGACGG).

[0045] In some aspects, the exonic sequence is linked to one or more untranslated regions at the 3' and/or 5' of the exonic sequence. In some aspects, the one or more untranslated regions comprise IRES. In some aspects, the IRES comprises a sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to SEQ ID NO: 87-110 or 145-168. In some aspects, the IRES is linked to a polyA. In some aspects, the one or more untranslated regions comprise a polyA. In some aspects, the exonic sequence is linked to a polyA. In some aspects, the polyA comprises a sequence having at least about 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% identity to SEQ ID NO: 85 or 86. In some aspects, the one

or more untranslated regions further comprise a spacer sequence. In some aspects, the spacer comprises a sequence having at least about 70%, at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 95%, at least about 95%, at least about 97%, at least about 98%, at least about 99%, or about 100% sequence identity to SEQ ID NO: 74-75 or 78-79.

[0046] In some aspects, the exonic sequence encodes a protein. In some aspects, the circular RNA exhibits an increased expression of the protein, a prolonged expression of the protein in vitro and/or in vivo, a reduced activation of an immune signaling pathway, an increased immune activation, or any combination thereof. In some aspects, the protein comprises a therapeutic protein. In some aspects, the protein comprises a cytokine, a receptor, a ligand, an immunomodulatory, a growth factor, an antigen, an RNA binding domain, or any combination thereof. In some aspects, the protein comprises an antibody, a fusion protein, or any combination thereof. In some aspects, the protein comprises CD3, CD4, CDS, CD19, CD20, CD22, CD30, CD34, HER2, HER3, HER4, LFA-I, MoI, pl50, 95, VLA-4, ICAM-I, VCAM, alpha v/beta 3 integrin, human macrophage inflammatory protein (MIP-1 -alpha), erythropoietin (EPO), NGF-beta, platelet-derived growth factor (PDGF), fibroblast growth factors, epidermal growth factor (EGF), transforming growth factors (TGF), insulin-like growth factors-I and -II (IGF-I and IGF-II), des(1-3)-IGF-I (brain IGF-1), factor VIII, tissue factor, von Willebrands factor, protein C, alpha-1-antitrypsin, plasminogen activators, such as urokinase and tissue plasminogen activator ("t-PA"), bombazine, thrombin, thrombopoietin, M-CSF, GM-CSF, G-CSF, flk2/flt3 receptor, obesity (OB) receptor, LDL receptor, growth hormone receptors, thrombopoietin receptors ("TPO-R," "c-mpl"), glucagon receptors, interleukin receptors, interferon receptors, T-cell receptors, stem cell factor receptors, OX40L, bone-derived neurotrophic factor (BDNF), neurotrophin-3, -4, -5, or -6 (NT-3, NT-4, NT-5, or NT-6), relaxin A-chain, relaxin B-chain, prorelaxin, interleukins and interleukin receptors, including IL-1 to IL-33 and IL-1 to IL-33 receptors, such as the IL-8 receptor, viral antigens, lipoproteins, calcitonin, glucagon, atrial natriuretic factor, lung surfactant, tumor necrosis factor-alpha and -beta, enkephalinase, RANTES (regulated on activation normally T-cell expressed and secreted), mouse gonadotropin-associated peptide, DNase, inhibin, activin, PUF domain, Cas protein, or any combination thereof.

[0047] In some aspects, the disclosure provides a composition comprising the genetic construct. In some aspects, the composition comprises the genetic construct disclosed herein and a GTP. In some aspects, the GTP in the composition is in a sufficient amount to cleave the 5' intron and the 3' intronic sequences from the genetic construct.

[0048] In some aspects, the disclosure provides a method of making a circular RNA, comprising contacting a GTP and the genetic construct disclosed herein.

[0049] In some aspects, the disclosure provides a circular RNA prepared by the method disclosed herein.

[0050] In some aspects, the disclosure provides a method of treating a disease or condition in a subject in need thereof, comprising administering the circular RNA disclosed herein to the subject.

BRIEF DESCRIPTION OF THE DRAWINGS

[0051] Aspects of the present disclosure may take form in various components and arrangements of components, and in various steps and arrangements of steps. The drawings are only for illustration purpose of preferred embodiments of the present disclosure and are not to be considered as limiting.

[0052] Features of embodiments of the present disclosure will be more readily understood from the following detailed description taken in conjunction with the accompanying drawings.

[0053] **Fig. 1A** is a schematic representation of a genetic construct (DNA or RNA genetic construct) of formula:

5'-P-H-A-3S-E-5S-B-H'-3'

wherein:

P comprises a promoter;

H comprises a 5' homology arm;

A comprises a 3' intronic sequence comprising a catalytic core (CC);

3S comprises a 3' splice site;

E comprises an exonic sequence;

5S comprises a 5' splice site;

B comprises a 5' intronic sequence comprising an internal guide sequence (IGS);

H' comprises a 3' homology arm;

wherein the promoter (P) is to be understood to be comprised in the DNA genetic construct, but not in the RNA genetic construct.

[0054] **Fig. 1B** is a schematic representation of a genetic construct (DNA or RNA genetic construct) of formula:

5'-P-H-A-3S-E-5S-B-H'-3'

wherein:

P comprises a promoter;

H comprises a 5' homology sequence;

A comprises a 3' intronic sequence comprising a catalytic core (CC);

3S comprises a 3' splice site;

E comprises an exonic sequence comprising a first excision scar precursor (ESP) (X) and a second ESP (X');

B comprises a 5' intronic sequence; and

H' comprises a 3' homology arm;

wherein the promoter (P) is to be understood to be comprised in the DNA genetic construct, but not in the RNA genetic construct.

[0055] Fig. 1C is a schematic representation of a genetic construct (DNA or RNA genetic construct) of formula:

5'-P-A-3S-E-3'

wherein:

P comprises a promoter;

A comprises a 3' intronic sequence comprising a catalytic core (CC)

3S comprises a 3' splice site; and

E comprises an exonic sequence comprising a catalytic nucleotide (CN);

wherein the promoter (P) is to be understood to be comprised in the DNA genetic construct, but not in the RNA genetic construct.

[0056] Fig. 1D is a schematic representation of a genetic construct (DNA or RNA genetic construct) of formula:

5'-P-A-3S-E-3'

wherein:

P comprises a promoter;

A comprises a 3' intronic sequence comprising a catalytic core (CC);

3S comprises a 3' splice site;

E comprises an exonic sequence comprising:

a first portion of the exonic sequence (E');

a first homology arm (HI)

a second portion of the exonic sequence (E'')

a second homology arm (HI'); and

a third exonic sequence (E''') comprising a catalytic nucleotide (CN);

wherein the promoter (P) is to be understood to be comprised in the DNA genetic construct, but not in the RNA genetic construct.

[0057] **Fig. 1E** is a schematic representation of a genetic construct (DNA or RNA genetic construct) of formula:

5'-P-B-A-3S-E-3'

wherein:

P comprises a promoter;

B comprises a 5' intronic sequence comprising an internal guide sequence (IGS);

A comprises a 3' intronic sequence comprising a catalytic core (CC)

3S comprises a 3' splice site;

E comprises an exonic sequence comprising a catalytic nucleotide (CN);

wherein the promoter (P) is to be understood to be comprised in the DNA genetic construct, but not in the RNA genetic construct.

[0058] **Fig. 2** is a schematic representation of a circularization mechanism.

[0059] **Fig. 3A** is a plasmid map of the genetic construct depicted in **Fig. 1A**.

[0060] **Fig. 3B** is a plasmid map of the genetic construct depicted in **Fig. 1B**.

[0061] **Fig. 4** is a denaturing RNA gel of 4 group I introns run pre (-) and post (+) GTP circularization. RNA ladder sizes shown to the right of the chart. Fragment analyzer (FA) smear analysis of spliced (circularized), un-spliced, and nicked RNA, as well as concatemers are shown for post GTP samples.

[0062] **Fig. 5A** is a fragment analyzer trace of RNAs circularized in the presence of GTP from RNA genetic constructs comprising sequences containing intronic sequences derived from *Anabaena PCC7120*.

[0063] **Fig. 5B** is a fragment analyzer trace of RNAs circularized in the presence of GTP from RNA genetic constructs comprising sequences containing intronic sequences derived from *Anabaena azollae*.

[0064] **Fig. 5C** is a fragment analyzer trace of RNAs circularized in the presence of GTP from RNA genetic constructs comprising sequences containing intronic sequences derived from *Azoarcus*.

[0065] **Fig. 5D** is a fragment analyzer trace of RNAs circularized in the presence of GTP from RNA genetic constructs comprising sequences containing intronic sequences derived from *bacteriophage T4*.

[0066] Fig. 6 is a chart representing a flow cytometry analysis of GFP payload in HEK293T cells transfected with circRNA derived from different introns as indicated.

DETAILED DESCRIPTION

[0067] The present disclosure is generally directed to improved genetic constructs for synthesizing circular RNAs. Provided herein are RNA genetic constructs capable of forming circular RNAs (*i.e.*, pre-circRNAs), DNA genetic constructs from which the pre-circRNAs can be transcribed, methods of synthesizing circular RNAs (*i.e.*, circRNAs) using the pre-circRNAs, the circRNAs and/or DNA genetic constructs disclosed herein, pharmaceutical compositions comprising the genetic constructs (e.g., pre-circRNAs) and/or circRNAs disclosed herein, and method of treating a disease in a subject in need thereof comprising the pre-circRNAs, the circRNAs and/or DNA genetic constructs or the pharmaceutical compositions disclosed herein.

[0068] Aspects of the present disclosure include synthesizing circRNA from a genetic construct that includes group I introns. The circRNAs discussed in further detail herein may have improved immunogenicity relative to other circRNAs.

I. Definitions

[0069] Unless defined otherwise, all technical and scientific terms used herein have the same meaning as commonly understood by one of ordinary skill in the art to which this disclosure belongs. In case of conflict, the present application, including the definitions, will control. Unless otherwise required by context, singular terms shall include pluralities and plural terms shall include the singular.

[0070] Throughout this disclosure, the term "a" or "an" entity refers to one or more of that entity; for example, "a polynucleotide," is understood to represent one or more polynucleotides. As such, the terms "a" (or "an"), "**one or more**," and "**at least one**" can be used interchangeably herein.

[0071] Furthermore, "**and/or**" where used herein is to be taken as specific disclosure of each of the two specified features or components with or without the other. Thus, the term "and/or" as used in a phrase such as "A and/or B" herein is intended to include "A and B," "A or B," "A" (alone), and "B" (alone). Likewise, the term "and/or" as used in a phrase such as "A, B, and/or C" is intended to encompass each of the following aspects: A, B, and C; A, B, or C; A or C; A or B; B or C; A and C; A and B; B and C; A (alone); B (alone); and C (alone).

[0072] It is understood that wherever aspects are described herein with the language "**comprising**," otherwise analogous aspects described in terms of "**consisting of**" and/or

"**consisting essentially of**" are also provided. As used herein, "**comprising**" is synonymous with "**including**," "**containing**," or "**characterized by**," and is inclusive or open-ended and does not exclude additional, unrecited elements or method steps. As used herein, "**consisting of**" excludes any element, step, or ingredient not specified in the claim element. As used herein, "**consisting essentially of**" does not exclude materials or steps that do not materially affect the basic and novel characteristics of the claim.

[0073] The term "**about**" is used herein to mean approximately, roughly, around, or in the regions of. When the term "about" is used in conjunction with a numerical range, it modifies that range by extending the boundaries above and below the numerical values set forth. In general, the term "about" is used herein to modify a numerical value above and below the stated value by a variance of 10 percent, up or down (higher or lower), unless indicated otherwise. Unless otherwise indicated, all numbers expressing quantities of ingredients, reaction conditions, and so forth used in the specification, including claims, are to be understood as being modified in all instances by the term "about." Accordingly, unless otherwise indicated to the contrary, the numerical parameters are approximations and can vary depending upon the desired properties sought to be obtained. At the very least, and not as an attempt to limit the application of the doctrine of equivalents to the scope of the claims, each numerical parameter should be construed in light of the number of significant digits and ordinary rounding approaches.

[0074] The term "**at least**" prior to a number or series of numbers is understood to include the number adjacent to the term "at least," and all subsequent numbers or integers that could logically be included, as clear from context. For example, the number of nucleotides in a nucleic acid molecule must be an integer. For example, "at least 18 nucleotides of a 21-nucleotide nucleic acid molecule" means that 18, 19, 20, or 21 nucleotides have the indicated property. When at least is present before a series of numbers or a range, it is understood that "at least" can modify each of the numbers in the series or range. "At least" is also not limited to integers (*e.g.*, "at least 5%" includes 5.0%, 5.1%, 5.18% without consideration of the number of significant figures).

[0075] "**Nucleic acid**," "**nucleic acid molecule**," "**nucleotide sequence**," "**nucleic acid sequence**," "**polynucleotide**," and grammatical variants thereof are used interchangeably and refer to the phosphate ester polymeric form of ribonucleosides (adenosine, guanosine, uridine or cytidine; "RNA molecules") or deoxyribonucleosides (deoxyadenosine, deoxyguanosine, deoxythymidine, or deoxycytidine; "DNA molecules"), or any phosphoester analogs thereof, such as phosphorothioates and thioesters, in either single stranded form, or a double-stranded helix. Single stranded nucleic acid sequences refer to single-stranded DNA (ssDNA) or single-stranded

RNA (ssRNA). Double stranded DNA-DNA, DNA-RNA, and RNA-RNA helices are possible. The term nucleic acid molecule, and in particular DNA or RNA molecule, refers only to the primary and secondary structure of the molecule, and does not limit it to any particular tertiary forms. Thus, this term includes double-stranded DNA found, *inter alia*, in linear or circular DNA molecules (e.g., restriction fragments), plasmids, supercoiled DNA and chromosomes. In discussing the structure of particular double-stranded DNA molecules, sequences can be described herein according to the normal convention of giving only the sequence in the 5' to 3' direction along the non-transcribed strand of DNA (*i.e.*, the strand having a sequence homologous to the mRNA). A "recombinant DNA molecule" is a DNA molecule that has undergone a molecular biological manipulation. DNA includes, but is not limited to, cDNA, genomic DNA, plasmid DNA, synthetic DNA, and semi-synthetic DNA. A "nucleic acid composition" of the disclosure comprises one or more nucleic acids as described herein. As described herein, a polynucleotide of the present disclosure comprises DNA, RNA, or both. In some aspects, the term "polynucleotide" includes polydeoxyribonucleotides (containing 2-deoxy-D-ribose), polyribonucleotides (containing D-ribose), including tRNA, rRNA, shRNA, siRNA, miRNA and mRNA, whether spliced or unspliced, any other type of polynucleotide which is an N- or C-glycoside of a purine or pyrimidine base, and other polymers containing normucleotidic backbones, for example, polyamide (e.g., peptide nucleic acids "PNAs") and polymorpholino polymers, and other synthetic sequence-specific nucleic acid polymers providing that the polymers contain nucleobases in a configuration which allows for base pairing and base stacking, such as is found in DNA and RNA.

[0076] It is to be understood that when a DNA nucleotide sequence is disclosed herein, the RNA sequence that can be transcribed therefrom is also to be considered as provided herein. Analogously, when a RNA nucleotide sequence is disclosed herein the DNA sequence from which the RNA sequence can be transcribed is also to be considered as provided herein. Thus, for example, where a nucleotide sequences is provided as a DNA sequence, such as AATCGTA, the corresponding RNA sequence AAUCGUA is also to be considered as provided; where a nucleotide sequences is provided as a RNA sequence, such as AAUCGUA, the corresponding DNA sequence AATCGTA is also to be considered as provided. A person of ordinary skill in the art would know how to infer a DNA sequence form a provided RNA sequence (*i.e.*, substituting the U in the provided RNA sequence with a T to obtain the corresponding DNA sequence), and how to infer a RNA sequence form a provided DNA sequence (*i.e.*, substituting the T in the provided DNA sequence with a U to obtain the corresponding RNA sequence). A person of ordinary skill in the art would be aware of which DNA sequence are generally transcribed into RNA and which DNA

sequences are not generally transcribed into RNA. For example, a person of ordinary skill in the art would know that a promoter sequence is generally not transcribed from DNA into RNA; whereas sequences such as exons and introns are generally transcribed from DNA into RNA. Additionally, a person of ordinary skill in the art would know that some nucleotide sequences, although functionally active solely (or mostly) in their DNA form, are also present in the RNA sequence transcribed therefrom. Analogously, some nucleotide sequences, although functionally active solely (or mostly) in their RNA form, are also present and encoded from the DNA sequence from which the RNA is transcribed. For example, although RNA molecules are spliced, the nucleotide sequence of the splice sites is also present in and encoded from the DNA sequence from which the RNA sequence is transcribed.

[0077] The symbol "-" in the formula of any genetic constructs disclosed herein can be a phosphodiester bond, or one or more nucleotides inserted between two nucleotides. For example, when a formula requires 5'-A-E-B-3', wherein A comprises a 3' intronic sequence, E comprises an exonic sequence (*i.e.*, one or more exons or fragments thereof), and B comprises a 5' intronic sequence, and wherein "-" between A and E and/or "-" between E and B can be a phosphodiester bond or one or more nucleotides between A and E and/or E and B.

[0078] As used herein, the term "**polypeptide**" encompasses both peptides and proteins, unless indicated otherwise.

[0079] The term "**coding region**" refers to a DNA or RNA region (the transcribed region) which "encodes" a particular protein, *e.g.*, such as a payload.

[0080] The term "**RNA**" is used herein to mean a molecule which comprises at least one ribonucleotide residue. "**Ribonucleotide**" relates to a nucleotide with a hydroxyl group at the 2'-position of a β -D-ribofuranosyl group. The term comprises double-stranded RNA, single-stranded RNA, isolated RNA such as partially or completely purified RNA, essentially pure RNA, synthetic RNA, and recombinantly generated RNA such as modified RNA which differs from naturally occurring RNA by addition, deletion, substitution and/or alteration of one or more nucleotides. The term "**mRNA**" means "**messenger-RNA**" and relates to a "**transcript**" which is generated by using a DNA template and encodes a peptide or protein. Typically, an mRNA comprises a 5'-UTR, a protein coding region and a 3'-UTR. mRNA only possesses limited half-life in cells and in vitro. In the context of the present disclosure, mRNA can be generated by *in vitro* transcription from a DNA template. The *in vitro* transcription methodology is known to the skilled person. For example, there is a variety of in vitro transcription kits commercially available. As further described herein,

in some aspects, a RNA is a linear RNA. In some aspects, a RNA is a circular RNA. In some aspects, a RNA is a self-replicating RNA. In some aspects, a RNA is a non-replicating RNA.

[0081] As used herein, the term "**pre-circular RNA**" refers to a RNA (*e.g.*, mRNA) or DNA that is capable of directing its own circularization to form a circular RNA (also referred to herein as "**pre-circRNA**").

[0082] As used herein, the term "**circular RNA**," or "**circRNA**" refers to an RNA (*e.g.*, mRNA) that forms a circular structure through covalent bonds. CircRNAs are single-stranded RNA molecules lacking free 5' and 3' ends, and therefore lacking a 5' cap and a 3' polyA tail.

[0083] As used herein, the term "**payload sequence**" refers to a nucleotide sequence encoding a payload. As used herein, the term "**payload**" refers to any protein that can be encoded by the payload sequence. In some aspects, a payload comprises a therapeutic protein. Non-limiting examples of payloads are provided elsewhere in the present disclosure.

[0084] The term "**sequence identity**" is used herein to mean a relationship between two or more amino acid (polypeptide or protein) sequences or two or more nucleic acid (polynucleotide) sequences, as determined by comparing the sequences. In certain aspects, sequence identity is calculated based on the full length of two given SEQ ID NO or on part thereof. Part thereof can mean at least 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, or 100% of both SEQ ID NO, or any other specified percentage. The term "**identity**" can also mean the degree of sequence relatedness between amino acid or nucleic acid sequences, as the case can be, as determined by the match between strings of such sequences.

[0085] In some aspects, methods to determine identity are designed to give the largest match between the sequences tested. Methods to determine identity and similarity are codified in publicly available computer programs.

[0086] As used herein, the terms "**effective amount**" or "**therapeutically effective amount**" of, *e.g.*, a circRNA disclosed herein, refers to a quantity sufficient to, when administered to the subject, including a human, effect beneficial or desired results, including clinical results, and, as such, an "effective amount" or synonym thereto depends on the context in which it is being applied.

[0087] As used herein, the term "**target cell**" refers to a cell in which a payload (*e.g.*, encoded by the payload sequence) is desired to be expressed. As used herein, the term "**non-target cell**" refers to a cell in which a payload is not intended to be expressed.

[0088] As used herein, the terms "**genetic construct**" refers to an artificially-designed nucleic acid molecule, which can be borne on a vector. A genetic construct can be an RNA nucleic

acid molecule (*e.g.*, a pre-circRNA) or a DNA nucleic acid molecule (*e.g.*, an expression cassette from which the pre-circRNA can be transcribed, or a vector comprising the expression cassette).

[0089] As used herein, the term or "**vector**" refers to any nucleic acid molecule for the cloning of a nucleic acid, such as a plasmid, phage, transposon, cosmid, chromosome, artificial chromosome, virus, virion, etc. A vector can be a replicon to which another nucleic acid segment can be attached so as to bring about the replication of the attached segment. A "replicon" refers to any genetic element (*e.g.*, plasmid, phage, cosmid, chromosome, virus) that functions as an autonomous unit of replication *in vivo*, *i.e.*, capable of replication under its own control. A vector can be a "delivery vector" includes both viral and nonviral vehicles for introducing the nucleic acid into a cell *in vitro*, *ex vivo* or *in vivo*. A vector can be also used *in vitro* in the absence of cells to perform any of the procedures known in the art. For example, a vector can be used for *in vitro* transcription. A large number of vectors are known and used in the art including, for example, plasmids, modified eukaryotic viruses, or modified bacterial viruses. In some aspects, insertion of a polynucleotide into a suitable vector can be accomplished by ligating the appropriate polynucleotide fragments into a chosen vector that has complementary cohesive termini. Vectors can be engineered to encode selectable markers or reporters that provide for the selection or identification of cells that have incorporated the vector. Expression of **selectable markers** or reporters allows identification and/or selection of host cells that incorporate and express other coding regions contained on the vector. Examples of selectable marker genes known and used in the art include: genes providing resistance to ampicillin, streptomycin, gentamycin, kanamycin, hygromycin, bialaphos herbicide, sulfonamide, and the like; and genes that are used as phenotypic markers, *i.e.*, anthocyanin regulatory genes, isopentenyl transferase gene, and the like. Examples of **reporters** known and used in the art include: luciferase (Luc), green fluorescent protein (GFP), chloramphenicol acetyltransferase (CAT), β -galactosidase (LacZ), β -glucuronidase (Gus), and the like. Selectable markers can also be considered to be reporters. In some aspects, the delivery vector is selected from the group consisting of a viral vector (*e.g.*, an AAV vector), a plasmid, a lipid, a protein particle, a bacterial vector, and a lysosome.

[0090] As used herein the term "**splicing**" refers to the process by which exons are joined while introns are removed from primary transcripts (pre-RNAs) to form the mature RNAs.

[0091] "**Exon**" are the coding sections of a DNA molecule, or of an RNA molecule which is transcribed from a DNA molecule that are then translated into protein, or are part of a mature structural RNA (*e.g.*, tRNA or rRNA). In most eukaryotic genes, exons are separated by "**introns**," intervening non-coding sections of DNA. Therefore, introns are segments of nucleic acid that are

transcribed into RNA, but are excised via splicing and therefore are not present in the mature RNA transcript, which comprises only exons. The new, immature strands of messenger or structural RNAs, before splicing occurs are called pre-RNAs, and may contain both introns and exon, the RNA molecules resulting from splicing are called RNAs, or mature RNAs. pre-RNA molecules go through splicing and only the exons remain in the "mature RNA," which is then, in the case of mRNA, translated into a protein. "Structural RNAs" or "non-coding RNAs" are RNA molecules that are not translated into proteins and exert a function in the form of RNA molecules. Exemplary structural RNA molecules are rRNAs and tRNAs. During splicing, introns are removed from the pre-RNA by cleavage at conserved sequences called "**splice sites**," which are located at the 5' and 3' ends of the exons.

[0092] As used herein, the term "**exonic sequence**" refers to a nucleotide sequence which is comprised in the mature RNA after splicing. As used herein, the term "**intronic sequence**" refers to a nucleotide sequence which is removed from the RNA molecule during splicing and is thus not comprised in the mature RNA after splicing. An exonic sequence can comprise one or more exons, or portions thereof, can comprise a nucleotide sequence derived from a naturally occurring nucleotide sequence, or can be an artificial sequence. An intronic sequence can comprise one or more introns, or portions thereof, can comprise a nucleotide sequence derived from a naturally occurring nucleotide sequence, or can be an artificial sequence.

[0093] As used herein, the term "**5' intronic sequence**" and "**3' intronic sequence**" refer to intronic sequences that are located at the 5' and at the 3', respectively, of an intronic sequence. In some aspects, the constructs disclosed herein comprise "5' intronic sequence" and "3' intronic sequence" that are in a reverse position. Thus, in some aspects, a "3' intronic sequence" can be located at the 5', and a "5' intronic sequence" can be located at the 3' of the constructs disclosed herein.

[0094] As used herein, the term "**splice site**" refers to a sequence at the border between an exon and an intron. A splice site comprises a pair of nucleotides linked by a phosphodiester bond that is cleaved during a splicing reaction.

[0095] A 5' splice site is located at the border between the 3' end of an exon and the 5' end of an intron. A 5' splice site comprises a pair of nucleotides between which cleavage of the phosphodiester bond occurs during a splicing reaction. The nucleotide at the 5' of the pair of nucleotides is comprised in the exon, the nucleotide at the 3' of the pair of nucleotides is comprised in the intron.

[0096] A 3' splice site is located at the border between the 5' end of an exon and the 3' end of an intron. A 3' splice site comprises a pair of nucleotides between which cleavage of the phosphodiester bond occurs during a splicing reaction. The nucleotide at the 5' of the pair of nucleotides is comprised in the intron, the nucleotide at the 3' of the pair of nucleotides is comprised in the exon.

[0097] As used herein, the term "**catalytic nucleotide**" refers to a nucleotide that facilitates the circularization of pre-circRNA by interacting with a catalytic core of an intronic sequence.

[0098] "**Linear nucleic acid molecules**" are said to have a "5'-terminus" (5' end) and a "3'-terminus" (3' end) because nucleic acid phosphodiester linkages occur at the 5' carbon and 3' carbon of the sugar moieties of the substituent mononucleotides. The end nucleotide of a polynucleotide at which a new linkage would be to a 5' carbon is its 5' terminal nucleotide. The end nucleotide of a polynucleotide at which a new linkage would be to a 3' carbon is its 3' terminal nucleotide. A terminal nucleotide, as used herein, is the nucleotide at the end position of the 3'- or 5'-terminus.

[0099] "**Transcription**" means the formation or synthesis of an RNA molecule by an RNA polymerase using a DNA molecule as a template. The invention is not limited with respect to the RNA polymerase that is used for transcription. For example, in some embodiments, a T7- type RNA polymerase can be used.

[0100] "**Translation**" means the formation of a polypeptide molecule by a ribosome based upon an RNA template.

[0101] As used herein, the term "**coding sequence**," or "expression sequence" refers to a nucleic acid sequence that encodes a product, e.g., a peptide or polypeptide (*i.e.*, a protein coding sequence), or a regulatory or structural nucleic acid (*i.e.*, a tRNA or a rRNA).

[0102] As used herein, a "**spacer**" refers to a region of a polynucleotide sequence ranging from 1 nucleotide to hundreds or thousands of nucleotides separating two other elements along a polynucleotide sequence. The sequences can be defined or can be random. A spacer is typically non-coding. In some embodiments, spacers include duplex forming regions.

[0103] As used herein, an "**internal ribosome entry site**" or "IRES" refers to an RNA sequence or structural element ranging in size from 10 nt to 1000 nt or more, capable of initiating translation of a polypeptide in the absence of a typical RNA cap structure. An IRES is typically about 500 nt to about 700 nt in length.

II. Pre- circular RNA and DNA Genetic Constructs

[0104] Provided herein are RNA or DNA genetic constructs for the synthesis of circular RNAs (circRNAs). The disclosure also provides for circRNAs produced from such genetic constructs. In some aspects, the genetic constructs for the synthesis of circRNAs are RNA genetic constructs (*i.e.*, pre-circRNAs). In some aspects, the genetic constructs for the synthesis of circRNAs are DNA genetic constructs from which genetic constructs can be transcribed. In some aspects, the genetic constructs disclosed herein, or the genetic constructs transcribed from the DNA genetic constructs disclosed herein, undergo a splicing reaction. In some aspects, the splicing reaction is an auto-catalytic splicing reaction (*i.e.*, self-splicing reaction).

[0105] In some aspects, the genetic constructs disclosed herein comprise one or more intronic sequence (*i.e.*, one or more intron or fragment thereof), one or more exonic sequence (*i.e.*, one or more exon or fragment thereof), and one or more splice sites.

[0106] In some aspects, the genetic constructs disclosed herein comprise one or more intronic sequence (*i.e.*, one or more intron or fragment thereof), one or more exonic sequence (*i.e.*, one or more exon or fragment thereof), and one or more catalytic nucleotides.

[0107] In some aspects, the genetic constructs disclosed herein comprise one or more intronic sequence (*i.e.*, one or more intron or fragment thereof), one or more exonic sequence (*i.e.*, one or more exon or fragment thereof), one or more splice sites, and one or more catalytic nucleotides.

[0108] In some aspects, the genetic constructs disclosed herein are DNA genetic constructs (*e.g.*, comprised in a vector).

[0109] In some aspects, the genetic constructs disclosed herein are RNA genetic constructs. In some aspects, the RNA genetic constructs disclosed herein undergo a splicing reaction. In some aspects, the splicing reaction is an auto-catalytic splicing reaction. In some aspects, the splicing reaction partially removes the one or more intronic sequence (*i.e.*, one or more intron or fragment thereof) from the RNA genetic constructs disclosed herein. In some aspects, the splicing reaction completely removes the one or more intronic sequence (*i.e.*, one or more intron or fragment thereof) from the RNA genetic constructs disclosed herein. In some aspects, the splicing reaction circularizes the one or more exonic sequence (*i.e.*, one or more exon or fragment thereof) comprised in the RNA genetic constructs disclosed herein.

[0110] In some aspects, the circularized one or more exonic sequence (*i.e.*, one or more exon or fragment thereof) is not a protein-coding exonic sequence. In some aspects, the circularized one or more exonic sequence (*i.e.*, one or more exon or fragment thereof) is a structural RNA. In these aspects, the circularized one or more exonic sequence (*i.e.*, one or more exon or fragment

thereof) can be a therapeutic agent (*i.e.*, a therapeutic RNA). In some aspects, the one or more exonic sequence (*i.e.*, one or more exon or fragment thereof) is a protein-coding sequence. In these aspects, the one or more exonic sequence (*i.e.*, one or more exon or fragment thereof) can encode a therapeutic agent (*i.e.*, a therapeutic protein). In these aspects, the one or more exonic sequence (*i.e.*, one or more exon or fragment thereof) are translated into a therapeutic agent (*i.e.*, a therapeutic protein).

[0111] In some aspects, the genetic constructs are linear RNAs produced by *in vivo* transcription of DNA genetic constructs disclosed herein. In some aspects, the genetic constructs (e.g., pre-circRNAs) are linear RNAs produced by *in vitro* transcription of DNA genetic constructs disclosed herein. In some aspects, the genetic constructs (e.g., pre-circRNAs) are linear RNAs produced by chemical synthesis. The genetic constructs (e.g., pre-circRNAs) disclosed herein can be unmodified, partially modified or completely modified.

[0112] In some aspects, the genetic constructs (e.g., pre-circRNAs) disclosed herein can circularize to produce a circular RNA (circRNA). In some aspects, the circRNAs disclosed herein comprises an exonic sequence (*i.e.*, one or more exons) comprised in the DNA genetic constructs or in the RNA genetic constructs disclosed herein. In some aspects, the circRNAs disclosed herein consists of an exonic sequence (*i.e.*, one or more exons) comprised in the DNA genetic constructs or in the RNA genetic constructs disclosed herein. In some aspects, the circRNAs disclosed herein further comprise one or more nucleotides comprised in the DNA genetic constructs or in the RNA genetic constructs disclosed herein that are not part of the exonic sequence (*i.e.*, one or more exons). The one or more nucleotides comprised in the DNA genetic constructs or in the genetic constructs disclosed herein that are not part of the exonic sequence (*i.e.*, one or more exons) can be, for example comprised in the intronic sequence, in the splice sites, in the homology arms of the DNA genetic constructs or in the genetic constructs disclosed herein, or any combination thereof. In these aspects, the circRNAs disclosed herein is described herein as comprising an excision scar. In some aspects, the circular RNA is at least 500, 600, 700, 800, 900, 1000, 1500, 2000, 2500, 3000, 3500, 4000 or 4500 nucleotides in size. The circRNA disclosed herein can be unmodified, partially modified or completely modified.

[0113] In some aspects, the RNA or DNA genetic constructs capable of forming circular RNAs disclosed herein (*i.e.*, pre-circRNAs or DNA genetic constructs from which pre-circRNAs can be transcribed) comprise one or more intronic sequence (*i.e.*, one or more intron or fragment thereof), and/or one or more exonic sequence (*i.e.*, one or more exon or fragment thereof). In some

aspects, the one or more intronic sequence (*i.e.*, one or more intron or fragment thereof) are derived from a nucleotide sequence of a Group I intron.

IIA. Genetic Constructs

IIA1. Genetic Construct A.

[0114] The present disclosure provides a genetic construct capable of forming circular RNAs disclosed herein (*i.e.*, pre-circRNAs or DNA genetic constructs from which pre-circRNAs can be transcribed) comprise a formula of 5'-H-A-E-B-H'-3', wherein H comprises a 5' homology arm, A comprises a 3' intronic sequence, E comprises an exonic sequence (*i.e.*, one or more exons or fragments thereof), B comprises a 5' intronic sequence, and H' comprises a 3' homology arm. In some aspects, the construct comprises one or more splice sites, e.g., between A and E and/or between E and B. In some aspects, the 3' terminus of A and 5' terminus of E form a 3' splice site (3S), and the 3' terminus of E and the 5' terminus of B form a 5' splice site (5S). In some aspects, the construct further comprises a promoter, wherein the promoter (P) is to be understood to be comprised in the DNA genetic construct, but not in the RNA genetic construct. In some aspects, the 5' and 3' intronic sequences are reverse complement of one another.

[0115] In some aspects, the disclosure provides a genetic construct capable of forming circular RNAs disclosed herein (*i.e.*, pre-circRNAs or DNA genetic constructs from which pre-circRNAs can be transcribed) comprise a formula of 5'-P-H-A-3S-E-5S-B-H'-3', wherein: P comprises a promoter (e.g., T7, T3, SP6, etc.), H comprises a 5' homology arm, A comprises a 3' intronic sequence, 3S comprises a 3' splice site, E comprises an exonic sequence (*i.e.*, one or more exons or fragments thereof), 5S comprises a 5' splice site, B comprises a 5' intronic sequence, H' comprises a 3' homology arm, and wherein the promoter (P) is to be understood to be comprised in the DNA genetic construct, but not in the RNA genetic construct. In some aspects, the 5' and 3' intronic sequences are reverse complement of one another. In some aspects, the 3' terminus of A and 5' terminus of E form a 3' splice site (3S), and the 3' terminus of E and the 5' terminus of B form a 5' splice site (5S).

[0116] The present disclosure is also directed to a method of producing a circular RNA comprising adding GTP to a genetic construct (e.g., DNA or RNA) comprising a nucleotide sequence having a formula of 5'-H-A-E-B-H'-3', wherein H comprises a 5' homology arm, A comprises a 3' intronic sequence, E comprises an exonic sequence (*i.e.*, one or more exons or fragments thereof), B comprises a 5' intronic sequence, and H' comprises a 3' homology arm or a formula of 5'-P-H-A-3S-E-5S-B-H'-3', wherein: P comprises a promoter (e.g., T7, T3, SP6, etc.),

H comprises a 5' homology arm, A comprises a 3' intronic sequence, 3S comprises a 3' splice site, E comprises an exonic sequence (*i.e.*, one or more exons or fragments thereof), 5S comprises a 5' splice site, B comprises a 5' intronic sequence, and H' comprises a 3' homology arm, and wherein the promoter (P) is to be understood to be comprised in the DNA genetic construct, but not in the RNA genetic construct. In some aspects, the 3' terminus of A and 5' terminus of E form a 3' splice site (3S), and the 3' terminus of E and the 5' terminus of B form a 5' splice site (5S).

[0117] In some aspects, the disclosure comprises a circular RNA produced by the pre-circRNA or the method herein.

IIA2. Genetic Construct B.

[0118] The present disclosure also provides a genetic construct capable of forming circular RNAs disclosed herein (*i.e.*, pre-circRNAs or DNA genetic constructs from which pre-circRNAs can be transcribed) comprise a formula of: 5'-H-A-E-B-H'-3', wherein H comprises a 5' homology arm, A comprises a 3' intronic sequence, E comprises an exonic sequence comprising a first excision scar precursor (X) and a second excision scar precursor X', B comprises a 5' intronic sequence, and H' comprises a 3' homology arm. In some aspects, the construct further comprises one or more splice sites, e.g., between A and E (3S) and/or between E and B (5S). In some aspects, the 3' terminus of A and 5' terminus of E form a 3' splice site (3S), and the 3' terminus of E and the 5' terminus of B form a 5' splice site (5S). In some aspects, the construct further comprises a promoter, wherein the promoter (P) is to be understood to be comprised in the DNA genetic construct, but not in the RNA genetic construct. In some aspects, the 5' and 3' intronic sequences are reverse complement of one another.

[0119] In some aspects, the present disclosure also provides a genetic construct capable of forming circular RNAs disclosed herein (*i.e.*, pre-circRNAs or DNA genetic constructs from which pre-circRNAs can be transcribed) comprise a formula of: 5'-P-H-A-3S-E-5S-B-H'-3', wherein: P comprises a promoter (e.g., T7, T3, SP6, etc.), H comprises a 5' homology arm, A comprises a 3' intronic sequence, 3S comprises a 3' splice site, E comprises an exonic sequence comprising a first excision scar precursor (X) and a second excision scar precursor X', 5S comprises a 5' splice site, B comprises a 5' intronic sequence, and H' comprises a 3' homology arm; and wherein the promoter (P) is to be understood to be comprised in the DNA genetic construct, but not in the RNA genetic construct. In some aspects, the 5' and 3' intronic sequences are reverse complement of one another. In some aspects, the 3' terminus of A and 5' terminus of E form a 3' splice site (3S), and the 3' terminus of E and the 5' terminus of B form a 5' splice site (5S).

[0120] The present disclosure is also directed to a method of producing a circular RNA comprising adding GTP to a genetic construct (e.g., DNA or RNA) comprising a nucleotide sequence having a formula of 5'-H-A-E-B-H'-3', wherein H comprises a 5' homology arm, A comprises a 3' intronic sequence, E comprises an exonic sequence comprising a first excision scar precursor (X) and a second excision scar precursor X', B comprises a 5' intronic sequence, and H' comprises a 3' homology arm or a formula of: 5'-P-H-A-3S-E-5S-B-H'-3', wherein: P comprises a promoter (e.g., T7, T3, SP6, etc.), H comprises a 5' homology arm, A comprises a 3' intronic sequence, 3S comprises a 3' splice site, E comprises an exonic sequence comprising a first excision scar precursor (X) and a second excision scar precursor X', 5S comprises a 5' splice site, B comprises a 5' intronic sequence, and H' comprises a 3' homology arm; and wherein the promoter (P) is to be understood to be comprised in the DNA genetic construct, but not in the RNA genetic construct. In some aspects, the 3' terminus of A and 5' terminus of E form a 3' splice site (3S), and the 3' terminus of E and the 5' terminus of B form a 5' splice site (5S).

[0121] In some aspects, the disclosure comprises a circular RNA produced by the pre-circRNA or the method herein.

HIA3. Genetic Construct C.

[0122] The present disclosure provides a genetic construct capable of forming circular RNAs disclosed herein (*i.e.*, pre-circRNAs or DNA genetic constructs from which pre-circRNAs can be transcribed) comprises a formula of: 5'-A-E-3', A comprises a 3' intronic sequence, and E comprises an exonic sequence comprising a catalytic nucleotide (CN). In some aspects, the construct further comprises a splice site, e.g., between A and E (3S). In some aspects, the construct further comprises a promoter, wherein the promoter (P) is to be understood to be comprised in the DNA genetic construct, but not in the RNA genetic construct. In some aspects, the 5' and 3' intronic sequences are reverse complement of one another.

[0123] The present disclosure provides a genetic construct capable of forming circular RNAs disclosed herein (*i.e.*, pre-circRNAs or DNA genetic constructs from which pre-circRNAs can be transcribed) comprises a formula of: 5'-P-A-3S-E-3', wherein P comprises a promoter (e.g., T7, T3, SP6, etc.), A comprises a 3' intronic sequence, 3S comprises a 3' splice site, and E comprises an exonic sequence comprising a catalytic nucleotide (CN); and wherein the promoter (P) is to be understood to be comprised in the DNA genetic construct, but not in the RNA genetic construct.

[0124] The present disclosure is also directed to a method of producing a circular RNA comprising allowing a genetic construct (e.g., DNA or RNA) to self-circularize, wherein the genetic construct comprises a nucleotide sequence having a formula of: 5'-A-E-3', A comprises a 3' intronic sequence, and E comprises an exonic sequence comprising a catalytic nucleotide (CN) or a formula of: 5'-P-A-3S-E-3', wherein P comprises a promoter (e.g., T7, T3, SP6, etc.), A comprises a 3' intronic sequence, 3S comprises a 3' splice site, and E comprises an exonic sequence comprising a catalytic nucleotide (CN); and wherein the promoter (P) is to be understood to be comprised in the DNA genetic construct, but not in the RNA genetic construct.

[0125] In some aspects, the disclosure comprises a circular RNA produced by the pre-circRNA or the method herein.

IIA4. Genetic Construct D.

[0126] The present disclosure provides a genetic construct capable of forming circular RNAs disclosed herein (*i.e.*, pre-circRNAs or DNA genetic constructs from which pre-circRNAs can be transcribed) comprises a formula of: 5'-A-E-3', A comprises a 3' intronic sequence comprising a catalytic core (CC); E comprises an exonic sequence comprising: (i) a first portion of the exonic sequence (E'); (ii) a first homology arm (HI); (iii) a second portion of the exonic sequence (E''); (iv) a second homology arm (HI'); and (v) a third exonic sequence (E''') comprising a catalytic nucleotide (CN). In some aspects, the construct further comprises a splice sites, e.g., between A and E (3S), e.g., at the 3' end of A. In some aspects, the construct further comprises a promoter, wherein the promoter (P) is to be understood to be comprised in the DNA genetic construct, but not in the RNA genetic construct. In some aspects, the 5' and 3' intronic sequences are reverse complement of one another. In some aspects, the 3' terminus of A and 5' terminus of E form a 3' splice site (3S).

[0127] In some aspects, the disclosure provides a genetic construct comprising a nucleotide sequence having a formula of: 5'-P-A-3S-E-3', wherein P is a promoter (e.g., T7, T3, SP6, etc.), A comprises a 3' intronic sequence comprising a catalytic core (CC); 3S comprises a 3' splice site; E comprises an exonic sequence comprising: (i) a first portion of the exonic sequence (E'); (ii) a first homology arm (HI); (iii) a second portion of the exonic sequence (E''); (iv) a second homology arm (HI'); and (v) a third exonic sequence (E''') comprising a catalytic nucleotide (CN), and wherein the promoter (P) is to be understood to be comprised in the DNA genetic construct, but not in the RNA genetic construct.

[0128] The present disclosure is also directed to a method of producing a circular RNA comprising allowing a genetic construct (e.g., DNA or RNA) to self-circularize, wherein the genetic construct comprises a nucleotide sequence having a formula of: 5'-A-E-3', A comprises a 3' intronic sequence comprising a catalytic core (CC); E comprises an exonic sequence comprising: (i) a first portion of the exonic sequence (E'); (ii) a first homology arm (HI); (iii) a second portion of the exonic sequence (E''); (iv) a second homology arm (HI'); and (v) a third exonic sequence (E''') comprising a catalytic nucleotide (CN) or a formula of: 5'-P-A-3S-E-3', wherein P is a promoter (e.g., T7, T3, SP6, etc.), A comprises a 3' intronic sequence comprising a catalytic core (CC); 3S comprises a 3' splice site; E comprises an exonic sequence comprising: (i) a first portion of the exonic sequence (E'); (ii) a first homology arm (HI); (iii) a second portion of the exonic sequence (E''); (iv) a second homology arm (HI'); and (v) a third exonic sequence (E''') comprising a catalytic nucleotide (CN), and wherein the promoter (P) is to be understood to be comprised in the DNA genetic construct, but not in the RNA genetic construct. In some aspects, the 3' terminus of A and 5' terminus of E form a 3' splice site (3S).

[0129] In some aspects, the disclosure comprises a circular RNA produced by the pre-circRNA or the method herein.

IIA5. Genetic Construct E.

[0130] The present disclosure provides a genetic construct capable of forming circular RNAs disclosed herein (*i.e.*, pre-circRNAs or DNA genetic constructs from which pre-circRNAs can be transcribed) comprise a formula of: 5'-B-A-E-3', wherein B comprises a 5' intronic sequence comprising an internal guide sequence (IGS), A comprises a 3' intronic sequence, and E comprises an exonic sequence comprising a catalytic nucleotide. In some aspects, the construct comprises a splice site between A and E. In some aspects, the 3' terminus of A and 5' terminus of E form a 3' splice site (3S). In some aspects, the construct comprises a promoter, wherein the promoter (P) is to be understood to be comprised in the DNA genetic construct, but not in the RNA genetic construct.

[0131] The present disclosure provides a genetic construct capable of forming circular RNAs disclosed herein (*i.e.*, pre-circRNAs or DNA genetic constructs from which pre-circRNAs can be transcribed) comprise a formula of: 5'-P-B-A-3S-E-3' wherein P comprises a promoter (e.g., T7, T3, SP6, etc.), B comprises a 5' intronic sequence comprising an internal guide sequence (IGS), A comprises a 3' intronic sequence, 3S comprises a 3' splice site, and E comprises an exonic

sequence comprising a catalytic nucleotide; and wherein the promoter (P) is to be understood to be comprised in the DNA genetic construct, but not in the RNA genetic construct.

[0132] The present disclosure is also directed to a method of producing a circular RNA comprising allowing a genetic construct (e.g., DNA or RNA) to self-circularize, wherein the genetic construct comprises a nucleotide sequence having a formula of: 5'-B-A-E-3', wherein B comprises a 5' intronic sequence comprising an internal guide sequence (IGS), A comprises a 3' intronic sequence, and E comprises an exonic sequence comprising a catalytic nucleotide or a formula of: 5'-P-B-A-3S-E-3', wherein P comprises a promoter (e.g., T7, T3, SP6, etc.), B comprises a 5' intronic sequence comprising an internal guide sequence (IGS), A comprises a 3' intronic sequence, 3S comprises a 3' splice site, and E comprises an exonic sequence comprising a catalytic nucleotide; and wherein the promoter (P) is to be understood to be comprised in the DNA genetic construct, but not in the RNA genetic construct.

[0133] In some aspects, the disclosure comprises a circular RNA produced by the pre-circRNA or the method herein.

II.B. Components of Genetic Constructs

II.B.1. Splice Sites

[0134] In some aspects, the genetic constructs capable of forming circular RNAs disclosed herein (*i.e.*, pre-circRNAs or DNA genetic constructs from which pre-circRNAs can be transcribed) comprise or form a 5' splice site, a 3' splice site, a catalytic nucleotide, or a combination thereof.

[0135] In some aspects, the genetic constructs capable of forming circular RNAs disclosed herein (*i.e.*, pre-circRNAs or DNA genetic constructs from which pre-circRNAs can be transcribed) comprise a 5' splice site and a 3' splice site.

[0136] In some aspects, the genetic constructs capable of forming circular RNAs disclosed herein (*i.e.*, pre-circRNAs or DNA genetic constructs from which pre-circRNAs can be transcribed) comprise a 3' splice site and a catalytic nucleotide.

[0137] In some aspects, the 5' splice site and the 3' splice site are transposed with respect to the position of the same splice sites would they have been in a naturally occurring configuration. In a naturally occurring configuration, a 5' splice site is located upstream of an exon, and a 3' splice site is located downstream of an exonic sequence. In the genetic constructs disclosed herein, a 5' splice site can be located downstream of an exonic sequence, and a 3' splice site can be located upstream of an exonic sequence, *i.e.*, the 5' splice site and the 3' splice site are transposed with

respect to the position of the same splice sites would they have been in a naturally occurring configuration.

[0138] In some aspects, the 3' splice site is located upstream of an exonic sequence and a catalytic nucleotide is located downstream of an exonic sequence.

[0139] A splice site is a nucleotide sequence located at the border between an intronic sequence and an exonic sequence, and comprises a pair of nucleotides linked by a phosphodiester bond that is cleaved during a splicing reaction. Thus, each splice site comprises a portion that is located in the intronic sequence and a portion that is located in the exonic sequence. In some aspects, the splice site is a 5' splice site. A 5' splice site is located at the border between the 3' end of an exonic sequence and the 5' end of an intronic sequence. A 5' splice site comprises a pair of nucleotides between which cleavage of the phosphodiester bond occurs during a splicing reaction. The nucleotide at the 5' of the pair of nucleotides (herein also referred to as "nucleotide at the 5' of the 5' splice site" or "5S(5'") is comprised in the exonic sequence, the nucleotide at the 3' of the pair of nucleotides (herein also referred to as "nucleotide at the 3' of the 5' splice site" or "5S(3'") is comprised in the intronic sequence.

[0140] In some aspects, the 5S(5') is a U or a T and the 5S(3') is a N, wherein N is any one of A, C, G, or U.

[0141] In some aspects, the 5' splice site comprises a sequence of formula 5'-U(or T)-N-3', wherein the site of splice is between U (or T) and N, and wherein N is any one of A, C, G, T, or U.

[0142] In some aspects, the splice site is a 3' splice site. A 3' splice site is located at the border between the 5' end of an exonic sequence and the 3' end of an intronic sequence. A 3' splice site comprises a pair of nucleotides between which cleavage of the phosphodiester bond occurs during a splicing reaction. The nucleotide at the 5' of the pair of nucleotides (herein also referred to as "nucleotide at the 5' of the 3' splice site" or "3S(5'") is comprised in the intronic sequence, the nucleotide at the 3' of the pair of nucleotides (herein also referred to as "nucleotide at the 3' of the 3' splice site" or "3S(3'") is comprised in the exonic sequence.

[0143] In some aspects, the 3S(5') is a G, and the 3S(3') is a N, wherein N is any one of A, C, G, T, or U.

[0144] In some aspects, the 3' splice site comprises a sequence of formula 5'-G-N-3', wherein the site of splice is between G and N, and wherein N is any one of A, C, G, T, or U.

[0145] In some aspects, the genetic constructs capable of forming circular RNAs disclosed herein (*i.e.*, pre-circRNAs or DNA genetic constructs from which pre-circRNAs can be transcribed) comprise a 5' splice site and a 3' splice.

[0146] In some aspects, the genetic constructs capable of forming circular RNAs disclosed herein (*i.e.*, pre-circRNAs or DNA genetic constructs from which pre-circRNAs can be transcribed) comprise a splice site and a catalytic nucleotide.

[0147] In some aspects, the genetic constructs capable of forming circular RNAs disclosed herein (*i.e.*, pre-circRNAs or DNA genetic constructs from which pre-circRNAs can be transcribed) comprise a 3' splice site and a catalytic nucleotide.

[0148] In some aspects, the 3' splice site comprises a nucleotide sequence having at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, or 100% identity with a nucleotide sequence of SEQ ID NOs: 31-36. In some aspects, the 3' splice site comprises a nucleotide sequence having at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, or 100% identity with a nucleotide sequence of SEQ ID NO: 33. In some aspects, the 3' splice site consists of a nucleotide sequence of SEQ ID NOs: 31-36. In some aspects, the 3' splice site consists of a nucleotide sequence of SEQ ID NO: 33.

[0149] In some aspects, the 5' splice site comprises a nucleotide sequence having at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, or 100% identity with a nucleotide sequence of SEQ ID NOs: 37-41 or a nucleotide sequence of NAGU[^]AAA, NCUU[^], NAGU[^], NCUU[^], or NGGU[^], wherein N is A, U, C, or G, and wherein "[^]" indicates the position of the phosphodiester bond that is cleaved during a splicing reaction. In some aspects, the 5' splice site comprises a nucleotide sequence of NAGU[^]AAA, wherein N is A, U, C, or G, and wherein "[^]" indicates the position of the phosphodiester bond that is cleaved during a splicing reaction. In some aspects, the 5' splice site consists of a nucleotide sequence of SEQ ID NOs: 37-41, or of a nucleotide sequence of NAGU[^]AAA, NCUU[^], NAGU[^], NCUU[^], or NGGU[^], wherein N is A, U, C, or G, and wherein "[^]" indicates the position of the phosphodiester bond that is cleaved during a splicing reaction. In some aspects, the 5' splice site consists of a nucleotide sequence of NAGU[^]AAA, wherein N is A, U, C, or G, and wherein "[^]" indicates the position of the phosphodiester bond that is cleaved during a splicing reaction. Non-limiting examples of 5' splice sites and 3' splice sites that are useful for the present genetic constructs are shown in **Table 1**.

Table 1. Exemplary splice sites ("[^]" indicates the position of the phosphodiester bond that is cleaved during a splicing reaction).

SEQ ID NO	DESCRIPTION	SEQUENCE
-----------	-------------	----------

31	Anabaena PCC7120 3' Splice site	UGCAAGAGAAUG [^] N
32	Tetrahymena thermophila 3' Splice site	UUGGAGUACUCG [^] UAAGGUAN
33	Azoarcus 3' Splice site	UCACACAAACCGG [^] N
34	Anabaena azollae 3' Splice site	UGCGAGAGAAUG [^] N
35	Bacteriophage T4 Td 3' Splice site	GAACAUAUUG [^] CUN
36	Bacteriophage T4 Td 3' Splice site	GAACAUAUUG [^] CUACCGUUUAAUAUUGCGU CAGGUAGUAAACUACUACUACAACCGUCU GAAGCAN
37	Anabaena PCC7120 5' Splice site	NCUU [^] AAAUAUUGAG
38	Tetrahymena thermophila 5' Splice site	NCCCUCU [^] AAUAGCAAU
	Azoarcus 5' Splice site	NAGU [^] AAA
39	Anabaena azollae 5' Splice site	NCUU [^] AGAAAACUGAGCCU
40	Bacteriophage T4 Td 5' Splice site	NGGU [^] UAAUUGAGGCUGAGU
41	Bacteriophage T4 Td 5' Splice site	NCAGAGAUGUUUCUUGGGU [^] UAAUUGAGG CUGAGU
	Anabaena PCC7120 5' Splice site	NCUU [^]
	Azoarcus 5' Splice site	NAGU [^]
	Anabaena azollae 5' Splice site	NCUU [^]
	Bacteriophage T4 Td 5' Splice site	NGGU [^]

II.B.2. Introns

[0150] Splicing is a process by which a precursor RNA transcript (*i.e.*, primary transcript) is transformed into a mature RNA molecule (*e.g.*, a mature mRNA, tRNA, or rRNA). During splicing, introns are removed from the precursor RNA molecule (*i.e.*, primary transcript) and exons are joined together. For many eukaryotic RNAs, splicing occurs in a series of reactions catalyzed by the spliceosome, a complex of small nuclear ribonucleoproteins (snRNPs), but self-splicing introns, can catalyze their own excision from their parent RNA molecule. For example, Group I introns can catalyze their own excision from their parent RNA molecule without accessory proteins.

[0151] Group I introns are self-splicing introns, which can catalyze their own excision from primary transcripts (*e.g.*, pre-mRNA, pre-tRNA and pre-rRNA) in a wide range of organisms. The ability of group I introns to self-splice renders them capable of acting as ribozymes.

[0152] Group I introns are found in bacterial, phage, viral, organellar genomes and often nuclear rDNA genes of fungi, plants, and algae.

[0153] Group I introns are highly variable at the primary sequence level yet possess characteristic conserved secondary and tertiary structures. Structural investigations have revealed a well-defined and highly conserved RNA core responsible for the catalysis, despite the fact that almost no universally conserved nucleotide residues are present among group I introns. The secondary structure of group I introns consists of paired (P) elements designated P1 to P13 and single-stranded loop regions. Short, conserved sequences can be recognized in some intronic sequences. The P1 and the P10 helices form the substrate-binding domain wherein the 5' and 3' splice sites are juxtaposed to each other. In some group I introns, P2 is absent. The active core of the group I ribozyme is assembled by two helical domains P4/P6 (P4, P5 and P6), which is considered the scaffolding domain, and P3/P9 (P3, P7, P8 and P9) that form the catalytic domain. The P3-P7-P9 helix contains the guanosine-5'-triphosphate (GTP) binding pocket and the exogenous GTP docks onto the G-binding site located in P7. Here the 3'-OH of an exogenous GTP is positioned so that it can attack the 5'-3' phosphodiester bond at the 5' splice site located within the P1 fold. In some aspects, the P10 comprises a nucleotide sequence of TAAGGTA (or UAAGGUA), or CT (or CU). In some aspects, the P10 consists of a nucleotide sequence of TAAGGTA (or UAAGGUA), or CT (or CU).

[0154] Group I introns are removed from precursor RNA by an autocatalytic RNA splicing event that is mediated by the intron's RNA tertiary structure. Base-pairing interactions between the 5'-end of the intron and flanking exonic sequences define the location of the 5' and 3' splice sites. The Internal Guide Sequence (IGS), which is a short intronic sequence near the 5'-end that pairs with sequences of the upstream exon to form P1, determines the 5' splice site. The 3' splice site is determined by pairing of a short sequence of the downstream exon with a portion of the IGS, forming P10 and mediating interactions between P9 and the P3/P8 helices that form the catalytic core.

[0155] Splicing of the group I intron RNA is by a two-step transesterification reaction with an exogenous GTP (α G) with its 3'-OH acting as an initiating nucleophile. Binding of the α G in the G-binding site in P7 positions the 3'-OH of GTP to attack the 5' splice site. During the first transesterification step the α G is attached to the 5'-end of the intron RNA by a 3'-5' phosphodiester bond. This step is followed by conformational changes allowing the upstream exon's terminal 3' guanosine (ω G) to trade position with the α G and occupy the G-binding site to initiate the second transesterification reaction. The 3'-OH of the upstream exon attacks the 3' splice site (an interaction facilitated by the formation of P10) promoting the ligation of upstream and downstream exons and the release of the intron RNA. Non limiting examples of P10 are shown below:

Table 2.

P10 (P2007) DNA T. thermophila	TAAGGTA
P10 (P2077) DNA T4 Td	CT
P10 (P2007) RNA T. thermophila	UAAGGUA
P10 (P2077) RNA T4 Td	CU

[0156] In some aspects, the one or more intronic sequence (*i.e.*, one or more intron or fragment thereof) is derived from a single intron. In some aspects, the one or more intronic sequence (*i.e.*, one or more intron or fragment thereof) is derived from two or more different introns.

[0157] In some aspects, the one or more intronic sequence (*i.e.*, one or more intron or fragment thereof) is an artificial sequences comprising a nucleotide sequence having at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, or 100% identity with a nucleotide sequence of a naturally occurring intron. In some aspects, the intron is a group I.

[0158] In some aspects, the one or more intronic sequence (*i.e.*, one or more intron or fragment thereof) is derived from an intron derived from a bacterial genome, a phage genome, a viral genome, a eukaryotic organellar genome, or from a nuclear genome of fungi, plants, or algae, or from an RNA transcribed therefrom. In some aspects, the one or more intronic sequence (*i.e.*, one or more intron or fragment thereof) is derived from a phage genome or from an RNA transcribed therefrom. In some aspects, the intronic sequence is derived from a bacteriophage T4 genome or from an RNA transcribed therefrom. In some aspects, the one or more intronic sequence (*i.e.*, one or more intron or fragment thereof) is derived from a bacterial genome or from an RNA transcribed therefrom. In some aspects, the one or more intronic sequence (*i.e.*, one or more intron or fragment thereof) is derived from a cyanobacterium genome or from an RNA transcribed therefrom. In some aspects, the one or more intronic sequence (*i.e.*, one or more intron or fragment thereof) is derived from an *Anabaena azollae*, *Anabaena PCC7120*, or *Aphanizomenon flos-aquae* (*e.g.*, strain NIVA-CYA 142), or *Synechococcus* genome or from an RNA transcribed therefrom. In some aspects, the one or more intronic sequence (*i.e.*, one or more intron or fragment thereof) is derived from a nitrogen-fixing bacterium genome or from an RNA transcribed therefrom. In some aspects, the one or more intronic sequence (*i.e.*, one or more intron or fragment thereof) is derived from an *Azoarcus* genome or from an RNA transcribed therefrom. In some aspects, the one or more intronic sequence (*i.e.*, one or more intron or fragment thereof) is derived from a protozoan genome or from an RNA transcribed therefrom. In some aspects, the one or more intronic sequence (*i.e.*, one or more intron or fragment thereof) is derived from a ciliophoran

genome or from an RNA transcribed therefrom. In some aspects, the one or more intronic sequence (*i.e.*, one or more intron or fragment thereof) is derived from a *Tetrahymena thermophila* genome or from an RNA transcribed therefrom. In some aspects, the one or more intronic sequence (*i.e.*, one or more intron or fragment thereof) is derived from a fungus genome or from an RNA transcribed therefrom. In some aspects, the one or more intronic sequence (*i.e.*, one or more intron or fragment thereof) is derived from a trimorphomyces genome or from an RNA transcribed therefrom. In some aspects, the one or more intronic sequence (*i.e.*, one or more intron or fragment thereof) is derived from a *Trimorphomyces papilionaceus* genome or from an RNA transcribed therefrom.

[0159] In some aspects, the one or more intronic sequence (*i.e.*, one or more intron or fragment thereof) is derived from a genomic region encoding a protein-coding RNA or a structural RNA (e.g., tRNA, rRNA), or from an RNA transcribed therefrom.

[0160] In some aspects, the one or more intronic sequence (*i.e.*, one or more intron or fragment thereof) is derived from an intron of a *thymidylate synthase* (td) from bacteriophage T4; from an intron of a pre-tRNA^{LEU} from *Anabaena azollae*, *Anabaena PCC7120*, *Aphanizomenon flos-aquae* (e.g., strain NIVA-CYA 142), or *Synechococcus*; from an intron of a pre-tRNA^{ILE} from *Azoarcus*; from an intron of a 26S ribosomal RNA from *Tetrahymena thermophila*; from an intron of a cytochrome C oxidase subunit 2 (*COX2*) from *Trimorphomyces papilionaceus*, or from any combination thereof. In some aspects, the one or more intronic sequence (*i.e.*, one or more intron or fragment thereof) is derived from an intron comprised in the GISSD database (<http://www.rna.whu.edu.cn/gissd/>).

[0161] In some aspects, the one or more intronic sequence (*i.e.*, one or more intron or fragment thereof) is derived from an intron of a *thymidylate synthase* (td) from bacteriophage T4.

[0162] In some aspects, the one or more intronic sequence (*i.e.*, one or more intron or fragment thereof) is derived from an intron of a pre-tRNA^{LEU} from *Anabaena azollae*.

[0163] In some aspects, the one or more intronic sequence (*i.e.*, one or more intron or fragment thereof) is derived from an intron of a pre-tRNA^{LEU} from *Anabaena PCC7120*.

[0164] In some aspects, the one or more intronic sequence (*i.e.*, one or more intron or fragment thereof) is derived from an intron of a pre-tRNA^{LEU} from *Aphanizomenon flos-aquae* (e.g., strain NIVA-CYA 142).

[0165] In some aspects, the one or more intronic sequence (*i.e.*, one or more intron or fragment thereof) is derived from an intron of a pre-tRNA^{LEU} from *Synechococcus*.

[0166] In some aspects, the one or more intronic sequence (*i.e.*, one or more intron or fragment thereof) is derived from an intron of a pre-tRNA^{ILE} from *Azoarcus*.

[0167] In some aspects, the one or more intronic sequence (*i.e.*, one or more intron or fragment thereof) is derived from an intron of a 26S ribosomal RNA from *Tetrahymena thermophila*.

[0168] In some aspects, the one or more intronic sequence (*i.e.*, one or more intron or fragment thereof) is derived from an intron of a cytochrome C oxidase subunit 2 (*COX2*) from *Trimorphomyces papilionaceus*.

[0169] In some aspects, the one or more intronic sequence (*i.e.*, one or more intron or fragment thereof) comprises a nucleotide sequence derived from an intron having at least 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or 100% identity with any one of SEQ ID NOs: 1-5, 6-10, 139-141, or 142-144. In some aspects, the one or more intronic sequence (*i.e.*, one or more intron or fragment thereof) comprises a nucleotide sequence derived from an intron having at least 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or 100% identity with SEQ ID NOs: 3 or 8. In some aspects, the one or more intronic sequence (*i.e.*, one or more intron or fragment thereof) comprises a nucleotide sequence derived from an intron consisting of any one of SEQ ID NOs: 1-5, 6-10, 139-141, or 142-144. In some aspects, the one or more intronic sequence (*i.e.*, one or more intron or fragment thereof) comprises a nucleotide sequence derived from an intron consisting of SEQ ID NOs: 3 or 8.

[0170] Non-limiting examples of intron that are useful for the present genetic constructs are shown in **Table 2**.

Table 3. Exemplary Introns

SEQ ID NO	DESCRIPTION	SEQUENCE
1	Anabaena PCC7120 full intron DNA	AAATAATTGAGCCTTAAAGAAGAAATTCCTTAAAGTGGATGCTCTCAAACCTCA GGGAAACCTAAATCTAGTTATAGACAAGGCAATCCTGAGCCAAGCCGAAGTA GTAATTAGTAAACAATAGATGACTTACAACCTAATCGGAAGGTGCAGAGACT CGACGGGAGCTACCCCTAACGTCAAGACGAGGGTAAAGAGAGAGTCCAATTCT CAAAGCCAATAGGCAGTAGCGAAAGCTGCAAGAGAATG
2	Tetrahymena thermophila full intron DNA	AAATAGCAATATTTACCTTTGGAGGGAAAAGTTATCAGGCATGCACCTGGTA GCTAGTCTTTAAACCAATAGATTGCATCGGTTTAAAAGGCAAGACCGTCAAAA TTGCGGGAAAGGGTCAACAGCCGTTTCACTACCAAGTCTCAGGGGAACTTT GAGATGGCCTTGCAAAGGGTATGGTAATAAGCTGACGGACATGGTCCTAACC ACGCAGCCAAGTCCTAAGTCAACAGATCTTCTGTTGATATGGATGCAGTTCA CAGACTAAATGTCGGTCGGGGAAGATGTATCTTCTCATAAGATATAGTCGG ACCTCTCCTTAATGGGAGCTAGCGGATGAAGTGATGCACACTGGAGCCGCT GGGAACCTAATTTGTATGCGAAAGTATATTGATTAGTTTGGAGTACTCG
3	Azoarcus full intron DNA	aaatttGctCCTTGCGCCGGGAAACCACGCAAGGGATGGTGTCAAATTCGGC GAAACCTAAGCGCCCGGGCGGTATGGCAACGCCGAGCCAAGCTTCGGCG CCTGCGCCGATGAAGGTGTAGAGACTAGACGGCACCCACCTAAGGCAACGC TATGGTGAAGGCATAGTCCAGGGAGTGGCGAAAGTCACACAAACCGG

4	Anabaena azollae full intron DNA	AGAAAACTGAGCCTTGATAGAGAAATCTTTTAAGTGGAGCTCTCAAATTCA GGGAAACCTAAATCTGTTTACAGATATGGCAATCCTGAGCCAAGCCAGAAA ATTAGACTTGAGATTTGATTTTGGAGTTAAACTCAAAATTTAAAATCCCAAA TTCAAAATTCGGGAAGGTGCAGAGACTCGACGGGAGCTACCTTAACGTAAA GCCGAGGGTAAAGAGAGAGTCCAATTCTCAAAGCCTGAAGTTGCTGAAGCAA CAAGGCAGTAGTGAAAGCTGCGAGAGAATG
5	Bacteriophage T4 Td full intron DNA	TAATTGAGGCTGAGTATAAGGTGACTTATACTTGTAATCTATCTAAACGGG GAACCTCTCTAGTAGACAATCCCGTGCTAAATTGTAGGACTAATTCCATTTA TCAGATTTCTAGGGGAGACCCTCGAATGGAATTGGTTCTACATAAATGCCTA ACGACTATCCCTTTGGGGAGTAGGGTCAAGTGACTCGAAACGATAGACAACT TGCTTTAACAAGTTGGAGATATAGTCTGCTCTGCATGGTGACATGCAGCTGG ATATAATTCCGGGTAAGATTACGACCTTATCTGAACATAATG
139	Aphanizomenon flos-aquae strain NIVA- CYA 142 Leucine Trna - intron DNA	cttaaAAAGGAATTGAGCCTTGATTGAGAAATCTTTCAAGTGGCAGCTCTCA AACTCAGGGAAACCTAAATCTATTCTGATAGATACGGCAATCCTGAGCCAAGCC CAGAGAAAATTTCCCTGGGAaGGTGCAGAGACTCGACGGGAGCTACCTAAC GTGAAGACGAGGGTAAAGGGAGAGTCCAATTCTCAAAACTTATGATTGCTTT AGCAACTGAGTAGTAGTGAAAAGTGAAGAGAATG
140	Synechococcus PCC 6301 tRNA (LEU) - intron DNA	AGAAAACTGGGCTCGATCGCGAAAGGGATCGAGTGGCAGCTCTCAAACCTCA GGGAAACCTAAACTTTAAACATTAAAGTCATGGCAATCCTGAGCCAAGCTA AAGCTAAACAATAACAGCTTTAGAAGGTGCAGAGACTAGACGGGAGCTACC CTAACGGATTCAAGCCGAGGGTAAAGGGATAGTCCAATTCTCAACATCGCGAT TGTTGATGGCAGCGAAAGTTGCAAAGAGAATG
141	Trimorphomyce s papilionaceus COX2 IA1 - intron DNA	TACTTTATGGTTGAGACGTAATGAAGTTAAATGCATGAACACCTAACATTA GGAAATATGCACTGGATGAATTTATAGATTTATCTACTCAACGACTAAGGTA TATTATACCTTTAAAAAACTTCACGAATTATCGAAGACATAGTCTGACCCTA TGTGAAAATATAGGAAGCTTAATACGGCGATAACATATAA
6	Anabaena PCC7120 full intron RNA	AAAUAAUUGAGCCUUAAGAAGAAAUUCUUUAAGUGGAUGCUCUCAAAACUCA GGGAAACCUAAAUCUAGUUUAUGACAAGGCAUCCUGAGCCAAGCCGAAGUA GUAAUUAGUAAAACAUAUGACUUACACUAUUCGGAAGGUGCAGAGACU CGACGGGAGCUACCCUAACGUCAAGACGAGGGUAAAGAGAGAGUCCAAUUCU CAAAGCCAAUAGGCAGUAGCGAAAGCUGCAAGAGAAUG
7	Tetrahymena thermophila full intron RNA	AAAUAGCAAUUAUUACCUUUGGAGGGAAAAGUUAUCAGGCAUGCACCUGGUA GCUAGUCUUUAAAACCAUAGAUUGCAUCGGUUUAAAAGGCAAGACCGUCAA UUGCGGGAAAGGGGUCAACAGCCGUUCAGUACCAAGUCAGGGGAAACUUU GAGAUGGCCUUGCAAAGGGUUAUGGUAAUAGCUGACGGAUGGUCCUAAAC ACGCAGCCAAGUCCUAAAGUCAACAGAUUCUUGUUAUUGGAUGCAGUUA CAGACUAAAUGUCGGUCGGGGAAGAUGUAUUCUUCUAUAGAUAUAGUCGG ACCUCUCCUAAUGGGAGCUAGCGGAUGAAGUGAUGCAACACUGGAGCCGCU GGGAACUAAUUGUAUGCGAAAGUAUAUUGAUUAGUUUUGGAGUACUCG
8	Azoarcus full intron RNA	aaauuuGcuCCUUGCGCCGGGAAACACGCAAGGGAUGGUGUCAAAUUCGGC GAAACCUAAGCGCCCGCCGGGCGUAUGGCAACGCCGAGCCCAAGCUUCGGCG CCUGCGCCGAUGAAGGUGUAGAGACUAGACGGCACCCACCUAAGGCAAACGC UAUGGUGAAGGCAUAGUCCAGGGAGUGGCGAAAGUCACACAAACCGG
9	Anabaena azollae full intron RNA	AGAAAACUGAGCCUUGAUAGAGAAAUCUUUUAAGUGGAAGCUCUCAAAUUCA GGGAAACCUAAAUCUGUUUACAGAUUAGGCAUCCUGAGCCAAGCCAGAAA AUUAGACUUGAGAUUUGAUUUUGGAGUUAAACUCAAAAUUUAAAUCCCAA UUCAAAAUUCGGGAAGGUGCAGAGACUCGACGGGAGUACCCUAAACGUAAA GCCGAGGGUAAAGAGAGAGUCCAAUUCUCAAAAGCCUGAAGUUGCUGAAGCAA CAAGGCAGUAGUGAAAGCUGCGAGAGAAUG
10	Bacteriophage T4 Td full intron RNA	UAAUUGAGGCCUGAGUAUAAGGUGACUUUAUCUUGUAAUCUAUCUAAACGGG GAACCUUCUAGUAGACAUCCCGUGCUAAAUUGUAGGACUAAUUCUUAUUA UCAGAUUUCUAGGGGAGACCUCGAAUGGAAUUGGUUCUACAUAUAAUGCCUA ACGACUAUCCCUUUGGGGAGUAGGGUCAAGUGACUCGAAACGAUAGACAACU UGC UUUAACAAGUUGGAGAUUAGUCUGCUCUGCAUGGUGACAUGCAGCUGG AUUAUAAUUCGGGGUUAAGAUUAACGACCUUAUCUGAACAUAAUG
142	Aphanizomenon flos-aquae strain NIVA-	cuuaaAAAGGAAUUGAGCCUUGAUUGAGAAAUCUUUCAAGUGGCAGCUCUCA AACUCAGGGAAACCUAAAUCUAUUCGUAGAUACGGCAAUCCUGAGCCAAGCC CAGAGAAAAUUUCCUGGGAAaGGUGCAGAGACUCGACGGGAGCUACCCUAAAC

	CYA 142 Leucine Trna - intron RNA	GUGAAGACGAGGGUAAAGGGAGAGUCCAAUUCUCAAACUUAUGAUUGCUIU AGCAACUGAGUAGUAGUGAAAACUGCAAGAGAAUG
143	Synechococcus PCC 6301 tRNA (LEU) - intron RNA	AGAAAACUGGGCCUCGAUCGCGAAAGGGAUCGAGUGGCAGCUCUCAAACUCA GGGAAACCUAAAACUUUAAACAUUAAAGUCAUGGCAAUCCUGAGCCAAGCUA AAGCUAAACAACUAAACAGCUUUAGAAGGUGCAGAGACUAGACGGGAGCUACC CUAACGGAUUCAGCCGAGGGUAAAGGGAUAGUCCAAUUCUCAAACUAGCGAU UGUUGAUGGCAGCGAAAGUUGCAAAGAGAAUG
144	Trimorphomyce s papilionaceus COX2 IA1 - intron RNA	UACUUUAUGGUUGAGACGUAAUGAAGUUAAAUGCAUGAACACCUAACAAUUA GGAAAUAUGCAGUGGAUGAAUUUAUAGAUAUUAUCUACUACGACUAGGUA UAUUAUACCUUUAAAAACUUCACGAAUUAUCGAAGACAUAGUCUGACCCUA UGUGAAAUAUAGGAAGCUAAUACGGCGAUAAACAUUUA

[0171] When referring to pre-circRNAs "derived from an intron of" means comprising an intronic sequence derived from the RNA encoded by a certain gene. When referring to a DNA genetic construct (from which pre-circRNAs can be transcribed) "derived from an intron of" means comprising an intronic sequence derived from the DNA sequence of a certain gene.

[0172] In some aspects, the intronic sequence (*i.e.*, one or more intron or fragment thereof) comprised in the genetic constructs capable of forming circular RNAs disclosed herein (*i.e.*, pre-circRNAs or DNA genetic constructs from which pre-circRNAs can be transcribed) comprises a nucleotide sequence of a full length intron. In some aspects, the intronic sequence (*i.e.*, one or more intron or fragment thereof) comprised in the genetic constructs capable of forming circular RNAs disclosed herein (*i.e.*, pre-circRNAs or DNA genetic constructs from which pre-circRNAs can be transcribed) does not comprise a nucleotide sequence of a full length intron.

[0173] In some aspects, the intronic sequence comprises less than 95%, less than 90%, less than 85%, less than 80%, less than 75%, less than 70%, less than 65%, less than 60%, less than 55%, less than 50%, less than 45%, less than 40%, less than 35%, less than 30%, less than 25%, less than 20%, less than 15%, less than 10%, less than 5%, less than 4%, less than 3%, less than 2%, less than 1% of a sequence of a full length intron.

[0174] In some aspects, the intronic sequence comprises a nucleotide sequence having at least 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or 100% identity with any one of SEQ ID NOs: 1-5, 6-10, 139-141, or 142-144. In some aspects, the intronic sequence comprises a nucleotide sequence having at least 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or 100% identity with any one of SEQ ID NOs: 3 or 8. In some aspects, the intronic sequence consists of a nucleotide sequence of any one of SEQ ID NOs: 1-5, 6-10, 139-141, or 142-144. In some aspects, the intronic sequence consists of a nucleotide sequence of SEQ ID NOs: 3 or 8.

[0175] In some aspects, the intronic sequence (*i.e.*, one or more intron or fragment thereof) comprised in the genetic constructs capable of forming circular RNAs disclosed herein (*i.e.*, pre-circRNAs or DNA genetic constructs from which pre-circRNAs can be transcribed) is split in two segments.

[0176] In some aspects, the intronic sequence is split in a 5' segment and a 3' segment. The 5' segment of the intronic sequence is herein also referred to as "5' intronic sequence," and the 3' segment of the intronic sequence is herein also referred to as "3' intronic sequence."

[0177] It is to be understood, that the 5' and 3' descriptors of the intron segments refer solely to the position of the 2 intron segments with respect to the full length sequence of the intron, and that both the 5' intronic sequence and the 3' intronic sequence can comprise any portion of the intronic sequence, so long as the intronic sequences required for circularization (e.g., the IGS, the omega G, or the catalytic core) are present in the intronic sequence. For example, an intronic sequence comprised in genetic constructs disclosed herein (*i.e.*, pre-circRNAs or DNA genetic constructs from which pre-circRNAs can be transcribed) can comprise a sequence comprising the 80% of the sequence of a full length intron (*i.e.*, an intron fragment comprising 80% of the sequence of the full length intron), wherein the full length intron has a length of 1000 nucleotides (from nucleotide 1 at the 5' to nucleotide 1000 at the 3'). Such exemplary intronic sequence comprises 800 nucleotides. The 800 nucleotides can correspond, for example, to nucleotides 1-300 and 501-1000 of the 1000 nucleotides of the full length intron, *i.e.*, from nucleotide 1 to nucleotide 300 and from nucleotide 501 to nucleotide 1000 of the full length intron. Such exemplary intronic sequence does not comprise nucleotides 301-500 of the full length intron. In this exemplary a case, the 5' intronic sequence may comprise, for example, 300 nucleotides, those nucleotides correspond to nucleotides 1-300 of the 1000 nucleotides of the full length intron, and the 3' intronic sequence may comprise, for example, 500 nucleotides corresponding to nucleotides 501-1000 of the 1000 nucleotides of the full length intron.

[0178] Additionally, both the 5' intronic sequence and the 3' intronic sequence can comprise a nucleotide sequence comprising nucleotides that are not contiguous in the sequence of the full length intron. For example, an exemplary intronic sequence comprising the 80% of the sequence of a full length intron (*i.e.*, an intron fragment comprising 80% of the sequence of the full length intron), wherein the full length intron has a length of 1000 nucleotides (from nucleotide 1 at the 5' to nucleotide 1000 at the 3'), comprises 800 nucleotide. In this exemplary case, the 5' intronic sequence may comprise, for example, 200 nucleotides, those nucleotides may correspond to nucleotides 1-50 and 100-250 of the 1000 nucleotides of the full length intron, *i.e.*, from

nucleotide 1 to nucleotide 50 and from nucleotide 100 to nucleotide 250 of the full length intron, and the 3' intronic sequence may comprise, for example, 600 nucleotides corresponding to nucleotides 401-1000 of the 1000 nucleotides of the full length intron.

[0179] In some aspects, the genetic constructs capable of forming circular RNAs disclosed herein (*i.e.*, pre-circRNAs or DNA genetic constructs from which pre-circRNAs can be transcribed) comprise a 5' intronic sequence and a 3' intronic sequence. In these aspects, the 5' intronic sequence and a 3' intronic sequence are comprised *in cis* in a same genetic construct.

[0180] In some aspects, the genetic constructs capable of forming circular RNAs disclosed herein (*i.e.*, pre-circRNAs or DNA genetic constructs from which pre-circRNAs can be transcribed) comprise a 5' intronic sequence and do not comprise and a 3' intronic sequence. In some aspects, the genetic constructs capable of forming circular RNAs disclosed herein (*i.e.*, pre-circRNAs or DNA genetic constructs from which pre-circRNAs can be transcribed) comprise a 3' intronic sequence and do not comprise and a 5' intronic sequence.

[0181] In some aspects, the 5' intronic sequence and the 3' intronic sequence are derived from a same intron. In some aspects, the 5' intronic sequence and the 3' intronic sequence are derived from at least two different introns. In some aspects a 5' intronic sequence can be a chimeric 5' intronic sequence, *i.e.*, a 5' intronic sequence comprising portions derived from one, two, or more introns. In some aspects a 3' intronic sequence can be a chimeric 3' intronic sequence, *i.e.*, a 3' intronic sequence comprising portions derived from one, two, or more introns. For example, an intronic sequence can comprise a P1 domain from a first intron, a catalytic core P7 domain from a second intron, and a 3' terminal end from a third intron.

[0182] In some aspects, the 5' intronic sequence and the 3' intronic sequence together correspond to a full length intron. In some aspects, the 5' intronic sequence and the 3' intronic sequence together are less than a full length intron. In some aspects, the 5' intronic sequence comprises less than 95%, less than 90%, less than 85%, less than 80%, less than 75%, less than 70%, less than 65%, less than 60%, less than 55%, less than 50%, less than 45%, less than 40%, less than 35%, less than 30%, less than 25%, less than 20%, less than 15%, less than 10%, less than 5%, less than 4%, less than 3%, less than 2%, less than 1% of a sequence of a full length intron.

[0183] In some aspects, the 3' intronic sequence comprises a nucleotide sequence having at least 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or 100% identity with any one of SEQ ID NOs: 11-16, 17-22, or 177-178. In some aspects, the 3' intronic sequence comprises a nucleotide sequence having at least 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or 100% identity with SEQ ID NOs: 13 or 19. In some aspects, the 3'

intronic sequence consists of a nucleotide sequence of any one of SEQ ID NOs: 11-16, 17-22, or 177-178. In some aspects, the 3' intronic sequence consists of a nucleotide sequence of SEQ ID NOs: 13 or 19.

[0184] In some aspects, the 5' intronic sequence comprises a nucleotide sequence having at least 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or 100% identity with any one of SEQ ID NOs: 23-26, 27-30, or 175-176; or a nucleotide sequence of AAA. In some aspects, the 5' intronic sequence comprises a nucleotide sequence of AAA. In some aspects, the 5' intronic sequence consists of a nucleotide sequence of any one of SEQ ID NOs: 23-26, 27-30, or 175-176. In some aspects, the 5' intronic sequence consists of nucleotide sequence of AAA.

[0185] In some aspects, the 3' intronic sequence comprises a nucleotide sequence having at least 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or 100% identity with SEQ ID NO: 11 or 17. In some aspects, the 3' intronic sequence consists of a nucleotide sequence of SEQ ID NO: 11 or 17. In some aspects, the 5' intronic sequence comprises a nucleotide sequence having at least 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or 100% identity with SEQ ID NO: 23 or 27. In some aspects, the 5' intronic sequence consists of a nucleotide sequence of SEQ ID NO: 23 or 27. In some aspects, the 3' intronic sequence comprises a nucleotide sequence having at least 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or 100% identity with SEQ ID NO: 11 or 17, and the 5' intronic sequence comprises a nucleotide sequence having at least 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or 100% identity with SEQ ID NO: 23 or 27. In some aspects, the 3' intronic sequence consists of a nucleotide sequence of SEQ ID NO: 11 or 17, and the 5' intronic sequence consists of a nucleotide sequence of SEQ ID NO: 23 or 27.

[0186] In some aspects, the 3' intronic sequence comprises a nucleotide sequence having at least 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or 100% identity with SEQ ID NO: 16 or 22. In some aspects, the 3' intronic consists of a nucleotide sequence of SEQ ID NO: 16 or 22. In some aspects, the 5' intronic sequence comprises a nucleotide sequence having at least 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or 100% identity with SEQ ID NO: 23 or 27. In some aspects, the 5' intronic sequence consists of a nucleotide sequence of SEQ ID NO: 23 or 27. In some aspects, the 3' intronic sequence comprises a nucleotide sequence having at least 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or 100% identity with SEQ ID NO: 16 or 22, and the 5' intronic sequence comprises a nucleotide sequence having at least 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or 100% identity with SEQ ID NO: 23 or 27. In some aspects, the 3'

intronic sequence consists of a nucleotide sequence of SEQ ID NO: 16 or 22, and the 5' intronic sequence consists of a nucleotide sequence of SEQ ID NO: 23 or 27.

[0187] In some aspects, the 3' intronic sequence comprises a nucleotide sequence having at least 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or 100% identity with SEQ ID NO: 12 or 18. In some aspects, the 3' intronic sequence consists of a nucleotide sequence of SEQ ID NO: 12 or 18. In some aspects, the 5' intronic sequence comprises a nucleotide sequence having at least 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or 100% identity with SEQ ID NO: 24 or 28. In some aspects, the 5' intronic sequence consists of a nucleotide sequence of SEQ ID NO: 24 or 28. In some aspects, the 3' intronic sequence comprises a nucleotide sequence having at least 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or 100% identity with SEQ ID NO: 12 or 18, and the 5' intronic sequence comprises a nucleotide sequence having at least 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or 100% identity with SEQ ID NO: 24 or 28. In some aspects, the 3' intronic sequence consists of a nucleotide sequence of SEQ ID NO: 12 or 18, and the 5' intronic sequence consists of a nucleotide sequence of SEQ ID NO: 24 or 28.

[0188] In some aspects, the 3' intronic sequence comprises a nucleotide sequence having at least 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or 100% identity with SEQ ID NO: 13 or 19. In some aspects, the 3' intronic sequence consists of a nucleotide sequence of SEQ ID NO: 13 or 19. In some aspects, the 5' intronic sequence comprises a nucleotide sequence of a nucleotide sequence of AAA. In some aspects, the 5' intronic sequence consists of a nucleotide sequence of AAA. In some aspects, the 3' intronic sequence comprises a nucleotide sequence having at least 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or 100% identity with SEQ ID NO: 13 or 19, and the 5' intronic sequence comprises a nucleotide sequence of AAA. In some aspects, the 3' intronic sequence consists of a nucleotide sequence of SEQ ID NO: 13 or 19, and the 5' intronic sequence consists of a nucleotide sequence of AAA.

[0189] In some aspects, the 3' intronic sequence comprises a nucleotide sequence having at least 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or 100% identity with SEQ ID NO: 14 or 20. In some aspects, the 3' intronic sequence consists of a nucleotide sequence of SEQ ID NO: 14 or 20. In some aspects, the 5' intronic sequence comprises a nucleotide sequence having at least 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or 100% identity with SEQ ID NO: 25 or 29. In some aspects, the 5' intronic sequence consists of a nucleotide sequence of SEQ ID NO: 25 or 29. In some aspects, the 3' intronic sequence comprises a nucleotide sequence having at least 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%,

97%, 98%, 99% or 100% identity with SEQ ID NO: 14 or 20, and the 5' intronic sequence comprises a nucleotide sequence having at least 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or 100% identity with SEQ ID NO: 25 or 29. In some aspects, the 3' intronic sequence consists of a nucleotide sequence of SEQ ID NO: 14 or 20, and the 5' intronic sequence consists of a nucleotide sequence of SEQ ID NO: 25 or 29.

[0190] In some aspects, the 3' intronic sequence comprises a nucleotide sequence having at least 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or 100% identity with SEQ ID NO: 15 or 21. In some aspects, the 3' intronic sequence consists of a nucleotide sequence of SEQ ID NO: 15 or 21. In some aspects, the 5' intronic sequence comprises a nucleotide sequence having at least 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or 100% identity with SEQ ID NO: 26 or 30. In some aspects, the 5' intronic sequence consists of a nucleotide sequence of SEQ ID NO: 26 or 30. In some aspects, the 3' intronic sequence comprises a nucleotide sequence having at least 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or 100% identity with SEQ ID NO: 15 or 21, and the 5' intronic sequence comprises a nucleotide sequence having at least 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or 100% identity with SEQ ID NO: 26 or 30. In some aspects, the 3' intronic sequence consists of a nucleotide sequence of SEQ ID NO: 15 or 21, and the 5' intronic sequence consists of a nucleotide sequence of SEQ ID NO: 26 or 30.

[0191] In some aspects, the 3' intronic sequence comprises a nucleotide sequence having at least 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or 100% identity with SEQ ID NO: 177 or 178. In some aspects, the 3' intronic sequence consists of a nucleotide sequence of SEQ ID NO: 177 or 178. In some aspects, the 5' intronic sequence comprises a nucleotide sequence having at least 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or 100% identity with SEQ ID NO: 175 or 176. In some aspects, the 5' intronic sequence consists of a nucleotide sequence of SEQ ID NO: 175 or 176. In some aspects, the 3' intronic sequence comprises a nucleotide sequence having at least 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or 100% identity with SEQ ID NO: 177 or 178, and the 5' intronic sequence comprises a nucleotide sequence having at least 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or 100% identity with SEQ ID NO: 175 or 176. In some aspects, the 3' intronic sequence consists of a nucleotide sequence of SEQ ID NO: 177 or 178, and the 5' intronic sequence consists of a nucleotide sequence of SEQ ID NO: 175 or 176.

[0192] Non-limiting examples of 5' intronic sequences and 3' intronic sequences that are useful for the present genetic constructs are shown in **Table 4**.

Table 4. Exemplary 5' intronic sequences and 3' intronic sequences.

SEQ ID NO	DESCRIPTION	SEQUENCE
11	Anabaena PCC7120 3' intron (DNA)	AACAATAGATGACTTACAATAATCGGAAGGTGCAGAGACTCGACGGGAGCT ACCCTAACGTCAAGACGAGGGTAAAGAGAGAGTCCAATTCTCAAAGCCAATA GGCAGTAGCGAAAGCTGCAAGAGAATG
12	Tetrahymena thermophila 3' intron (DNA)	ATTTACCTTTTGGAGGGGAAAAGTTATCAGGCATGCACCTGGTAGCTAGTCTTT AAACCAATAGATTGCATCGGTTTAAAAGGCAAGACCGTCAAATTGCGGGAAA GGGGTCAACAGCCGTTTCACTACCAAGTCTCAGGGGAACTTTGAGATGGCCT TGCAAAGGGTATGGTAATAAGCTGACGGACATGGTCCTAACACGCAGCCAA GTCTTAAGTCAACAGATCTTCTGTTGATATGGATGCGATTTCACAGACTAAAT GTCGGTTCGGGGAAGATGTATTCTTCTCATAAGATATAGTCGGACCTCTCCTT AATGGGAGCTAGCGGATGAAGTGATGCAACACTGGAGCCGCTGGGAACTAAT TTGTATGCGAAAGTATATTGATTAGTTTTGGAGTACTCG
13	Azoarcus 3' intron (DNA)	tttGctCCTTGCGCCGGGAAACCACGCAAGGGATGGTGTCAAATTCGGCGAA ACCTAAGCGCCCGCCCGGGCGTATGGCAACGCCGAGCCAAGCTTCGGCGCCT GCGCCGATGAAGGTGTAGAGACTAGACGGCACCCACCTAAGGCAAAACCTAT GGTGAAGGCATAGTCCAGGGAGTGGCGAAAGTCACACAAACCGG
14	Anabaena azollae 3' intron (DNA)	AACTCAAAATTTAAAATCCCAAATTCAAAATTCGGGAAGGTGCAGAGACTC GACGGGAGCTACCCTAACGTAAAGCCGAGGGTAAAGAGAGAGTCCAATTCTC AAAGCCTGAAGTTGCTGAAGCAACAAGGCAGTAGTGAAAGCTGCGAGAGAAAT G
15	Bacteriophage T4 Td 3' intron (DNA)	GGGAGACCTCGAATGGAATTGGTTCTACATAAATGCCTAACGACTATCCCT TTGGGGAGTAGGGTCAAGTGACTCGAAACGATAGACAACCTTGCTTTAACAAAG TTGGAGATATAGTCTGCTCTGCATGGTGACATGCAGCTGGATATAATTCCGG GGTAAGATTAAAGACCTTATCTGAACATAATG
16	Anabaena PCC7120 3' intron (DNA)	GGAGACCCTCGACCGTCGATTGTCCACTGGTCAACAATAGATGACTTACAAC TAATCGGAAGGTGCAGAGACTCGACGGGAGCTACCTAACGTCAAGACGAGG GTAAAGAGAGAGTCCAATTCTCAAAGCCAATAGGCAGTAGCGAAAGCTGCAA GAGAATG
17	Anabaena PCC7120 3' intron (RNA)	AACAAUAGAUGACUUACAACUAAUCGGAAGGUGCAGAGACUCGACGGGAGCU ACCCUAACGUCACAGACGAGGUAAAGAGAGAGUCCAAUUCUCAAAGCCAUA GGCAGUAGCGAAAGCUGCAAGAGAAUG
18	Tetrahymena thermophila 3' intron (RNA)	AUUUACCUUUGGAGGGGAAAAGUUUAUCAGGCAUGCACCUGGUAGUCUAGUCUUU AAACCAAUAGAUUUGCAUCGGUUUAAAAGGCAAGACCGUCAAUUGCGGGAAA GGGGUCAACAGCCGUUCAGUACCAAGUCUCAGGGGAAACUUUGAGAUGGCCU UGCAAAGGGUAUGGUAAUAAGCUGACGGACAUGGUCCUAAACACGCAGCCAA GUCCUAAGUCAACAGAUUCUUGUUGAUUGGAUGCAGUUCACAGACUAAAU GUCGGUCGGGGAAGAUGUAUUCUUCUCAUAAGAUAUAGUCGGACCUCUCCUU AAUGGGAGCUAGCGGAUGAAGUGAUGCAACACUGGAGCCGUGGGAACUAAU UUGUAUGCGAAAGUAUAUUGAUUAGUUUUGGAGUACUCG
19	Azoarcus 3' intron (RNA)	uuuGcuCCUUGCGCCGGGAAACCACGCAAGGGAUGGUGUCAAUUCGGCGAA ACCUAAGCGCCCGCCCGGGCGUAUGGCAACGCCGAGCCAAGCUUCGGCGCCU GCGCCGAUGAAGGUGUAGAGACUAGACGGCACCCACCUAAGGCAAACGCUAU GGUGAAGGCAUAGUCCAGGGAGUGGCGAAAGUCACACAAACCGG
20	Anabaena azollae 3' intron (RNA)	AACUCAAAUUUUAAAUCCCAAUUCAAAAUUCCGGGAAGGUGCAGAGACUC GACGGGAGCUACCCUAACGUAAAGCCGAGGGUAAAGAGAGAGUCCAAUUCUC AAAGCCUGAAGUUGCUGAAGCAACAAGGCAGUAGUGAAAGCUGCGAGAGAAU G
21	Bacteriophage T4 Td 3' intron (RNA)	GGGAGACCCUCGAAUGGAUUUGGUUCUACAUAAUGCCUAACGACUAUCCCU UUGGGGAGUAGGGUCAAGUGACUCGAAACGAUAGACAACUUGCUUUAAACAG UUGGAGAUUAUGUCUGCUCUGCAUGGUGACAUGCAGCUGGAUAUAUUCGG GGUAAGAUAUACGACCUUAUCUGAACAUAAUG
22	Anabaena PCC7120 3' intron (RNA)	GGAGACCCUCGACCGUCGAUUGUCCACUGGUCAACAAUAGAUGACUUACAAC UAAUCGGAAGGUGCAGAGACUCGACGGGAGCUACCCUAACGUCAAGACGAGG GUAAAGAGAGAGUCCAAUUCUCAAAGCCAUAAGGCAGUAGCGAAAGCUGCAA GAGAAUG

23	Anabaena PCC7120 5' intron (DNA)	AAATAATTGAGCCTTAAAGAAGAAATCTTTAAGTGGATGCTCTCAAACCTCA GGGAAACCTAAATCTAGTTATAGACAAGGCAATCCTGAGCCAAGCCGAAGTA GTAATTAGTAA
24	Tetrahymena thermophila 5' intron (DNA)	AAATAGCAAT
	Azoarcus 5' intron (DNA)	aaa
25	Anabaena azollae 5' intron (DNA)	AGAAAACCTGAGCCTTGATAGAGAAATCTTTAAGTGGAGCTCTCAAATTCA GGGAAACCTAAATCTGTTTACAGATATGGCAATCCTGAGCCAAGCCAGAAA ATTAGACTTGAGATTTGATTTTGGAGTTA
26	Bacteriophage T4 Td 5' intron (DNA)	TAATTGAGGCCTGAGTATAAGGTGACTTATACTTGTAATCTATCTAAACGGG GAACCTCTCTAGTAGACAATCCCGTGCTAAATTGTAGGACTAATTCCATTTA TCAGATTTCTAG
27	Anabaena PCC7120 5' intron (RNA)	AAAUAAUUGAGCCUUAAGAAGAAAUUCUUUAAGUGGAUGCUCUCAACUCA GGGAAACCUAAAUUCUAGUUAUAGACAAGGCAAUCCUGAGCCAAGCCGAAGUA GUAAUUAGUAA
28	Tetrahymena thermophila 5' intron (RNA)	AAAUAGCAAU
	Azoarcus 5' intron (RNA)	aaa
29	Anabaena azollae 5' intron (RNA)	AGAAAACUGAGCCUUGAUAGAGAAAUUUUAAGUGGAAGCUCUCAAAUUCA GGGAAACCUAAAUUCUGUUUACAGAUUAUGGCAAUCCUGAGCCAAGCCAGAAA AUUAGACUUGAGAUUUGAUUUUGGAGUUA
30	Bacteriophage T4 Td 5' intron (RNA)	UAAUUGAGGCCUGAGUUAUAGGUGACUUAUCUUGUAAUUCUAUCUAAACGGG GAACCUCUCUAGUAGACAAUCCCGUGCUAAAUUGUAGGACUAAUUCUUA UCAGAUUUCUAG

[0193] In some aspects, the 5' intronic sequence comprises an internal guide sequence. In some aspects, the 5' intronic sequence consists of an internal guide sequence flanked by any number of additional intronic nucleotides at the 5' (e.g., 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 30, 35, 40, 45, or 50, etc.), and/or any number of additional intronic nucleotides at the 3' (e.g., 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 30, 35, 40, 45, or 50, etc.).

[0194] In some aspects, the 5' intronic sequence consists of an internal guide sequence flanked by any number of additional intronic nucleotides at the 5' (e.g., 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 30, 35, 40, 45, or 50, etc.), and/or any number of additional intronic nucleotides at the 3' (e.g., 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 30, 35, 40, 45, or 50, etc.). In some aspects, the 5' intronic sequence consists of an internal guide sequence.

[0195] As use herein, the term "intronic nucleotides" refers to nucleotides that are part of the intron from which the intronic sequences are derived.

[0196] In some aspects, the 3' intronic sequence comprises less than 95%, less than 90%, less than 85%, less than 80%, less than 75%, less than 70%, less than 65%, less than 60%, less than

55%, less than 50%, less than 45%, less than 40%, less than 35%, less than 30%, less than 25%, less than 20%, less than 15%, less than 10%, less than 5%, less than 4%, less than 3%, less than 2%, less than 1% of a sequence of a full length intron.

[0197] In some aspects, the 3' intronic sequence comprises a catalytic core sequence. In some aspects, the 3' intronic sequence consists of a catalytic core sequence flanked by any number of additional intronic nucleotides at the 5' (e.g., 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 30, 35, 40, 45, or 50, etc.), and/or any number of additional intronic nucleotides at the 3' (e.g., 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 30, 35, 40, 45, or 50, etc.). In some aspects, the 3' intronic sequence consists of a catalytic core sequence flanked by any number of additional intronic nucleotides at the 5' (e.g., 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 30, 35, 40, 45, or 50, etc.), and/or any number of additional intronic nucleotides at the 3' (e.g., 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 30, 35, 40, 45, or 50, etc.). In some aspects, the 3' the intronic sequence consists of a catalytic core sequence.

II.B.3. Catalytic Core

[0198] In some aspects, the intronic sequence comprises a catalytic core. As used herein, the term "catalytic core" refers to a nucleotide sequence comprising sequences and structures indispensable for catalytic function. The catalytic core interacts with the substrates used for the steps of the self-splicing reaction. The substrates used for the steps of the self-splicing reaction comprise an exogenous guanosine, a 5'-splice site, a 3' splice site, or a combination thereof.

[0199] In some aspects, the catalytic core comprises a sequence of GAG and CC, wherein the GAG and the CC are separated by (N)_n, wherein N is any one of A, C, G, T, or U, and n is an integer. In some aspects, the catalytic core comprises a sequence of ACG and CT (or U), wherein the ACG and the CT(or U) are separated by (N)_n, wherein N is any one of A, C, G, T, or U, and n is an integer. In some aspects, an exogenous G binds to the catalytic core.

[0200] In some aspects, the intronic sequence does not comprise a catalytic core.

[0201] In some aspects, the catalytic core comprises a nucleotide sequence of any one of AGAGA[N]₆₋₄₁A[N]₄₃₋₄₅GUCC, ACAGA[N]₆₋₄₁A[N]₄₃₋₄₅GUCG, or ACGACU[N]₇₋₆₈UAGUCU, wherein N is any one of A, U, C, or G, and wherein each N in each one of the chains of [N]₆₋₄₁, [N]₄₃₋₄₅, or [N]₇₋₆₈, can be independently any one of A, U, C, or G. In some aspects, the catalytic core consists of a nucleotide sequence of AGAGA[N]₆₋₄₁A[N]₄₃₋₄₅GUCC, ACAGA[N]₆₋₄₁A[N]₄₃₋₄₅GUCG, or ACGACU[N]₇₋₆₈UAGUCU, wherein N is any one of A, U, C, or G, and wherein each

N in each one of the chains of [N]₆₋₄₁, [N]₄₃₋₄₅, or [N]₇₋₆₈, can be independently any one of A, U, C, or G.

[0202] Non-limiting examples of catalytic cores that are useful for the present genetic constructs are shown in **Table 5**.

Table 5. Exemplary catalytic cores

DESCRIPTION	SEQUENCE
Anabaena (P852) -A150-A154, A195, G198-C201	AGAGA [N] ₆₋₄₁ A [N] ₄₃₋₄₅ GUCC
T. thermophila (P2007) - A260-A264, A305, G308-G311	ACAGA [N] ₆₋₄₁ A [N] ₄₃₋₄₅ GUCC
Azoarcus (P2008) - A124-A128, A169, G172-C175	AGAGA [N] ₆₋₄₁ A [N] ₄₃₋₄₅ GUCC
Anabaena azollae (P2076) - A178-A182, A224, G227-C230	AGAGA [N] ₆₋₄₁ A [N] ₄₃₋₄₅ GUCC
T4 Td (P2077) A157-U162, U230-U235	ACGACU [N] ₇₋₆₈ UAGUCU

II.B.4. Internal Guide Sequence (IGS)

[0203] In some aspects the intronic sequence comprises an internal guide sequence. As used herein, the term "**internal guide sequence**" or "IGS" refers to a nucleic acid sequence comprised in the intronic sequence that pairs with the exon located at its 5' during the splicing reaction. In some aspects, the internal guide sequence pairs with a nucleotide comprised in the exon. In some aspects, the internal guide sequence is fully or substantially complementary to a nucleotide sequence comprised in the exon.

[0204] In some aspects, the internal guide sequence comprises a nucleotide sequence of any one of CTAA (or CUAA), CCCTCT (or CCCUCU), agT (or agU), CTT (or CUU), GGT (or GGU), TTGAG (or UUGAG), GGAGGG, tttGct (or uuuGcu), CTGAG (or CUGAG) or GGCC. In some aspects, the internal guide sequence comprises a nucleotide sequence of agT (or agU) or tttGct (or uuuGcu). In some aspects, the internal guide sequence consists of a nucleotide sequence of CTAA (or CUAA), CCCTCT (or CCCUCU), agT (or agU), CTT (or CUU), GGT (or GGU), TTGAG (or UUGAG), GGAGGG, tttGct (or uuuGcu), CTGAG (or CUGAG) or GGCC. In some aspects, the internal guide sequence consists of a nucleotide sequence of agT (or agU) or tttGct (or uuuGcu).

[0205] Non-limiting examples of internal guide sequences that are useful for the present genetic constructs are shown in **Table 6**.

Table 6. Exemplary internal guide sequences

DESCRIPTION	SEQUENCE
IGS' (P852) DNA Anabaena	CTTAA
IGS' (P2007) DNA T. thermophila	CCCTCT
IGS' (P2008) DNA Azoarcus	agT

IGS' (P2076) DNA <i>Anabaena azollae</i>	CTT
IGS' (P2077) DNA T4 Td	GGT
IGS (P852) DNA <i>Anabaena</i>	TTGAG
IGS (P2007) DNA <i>T. thermophila</i>	GGAGGG
IGS (P2008) DNA <i>Azoarcus</i>	tttGct
IGS (P2076) DNA <i>Anabaena azollae</i>	CTGAG
IGS (P2077) DNA T4 Td	GGCC
IGS' (P852) RNA <i>Anabaena</i>	CUUAA
IGS' (P2007) RNA <i>T. thermophila</i>	CCCUCU
IGS' (P2008) RNA <i>Azoarcus</i>	agU
IGS' (P2076) RNA <i>Anabaena azollae</i>	CUU
IGS' (P2077) RNA T4 Td	GGU
IGS (P852) RNA <i>Anabaena</i>	UUGAG
IGS (P2007) RNA <i>T. thermophila</i>	GGAGGG
IGS (P2008) RNA <i>Azoarcus</i>	uuuGcu
IGS (P2076) RNA <i>Anabaena azollae</i>	CUGAG
IGS (P2077) RNA T4 Td	GGCC

II.B.5. Homology Arms

[0206] In some aspects, the genetic constructs capable of forming circular RNAs disclosed herein (*i.e.*, pre-circRNAs or DNA genetic constructs from which pre-circRNAs can be transcribed) further comprise one or more homology arms. As used herein, the term "homology arm" refers to a sequence that is capable of paring (*i.e.*, hybridizing) with another sequence.

[0207] In some aspects, the homology arms are at the 5' and/or at the 3' end of the genetic constructs disclosed herein. For example, the first homology arm can be located at the 5' end of the genetic construct, and a second homology arm can be located at the 3' end of the genetic construct. In some aspects, the homology arms are interposed between any two other sequences comprised in the genetic constructs disclosed herein. For example, a first homology arm can be between the 3' splice site and the exonic sequence, and a second homology arm can be between the 5' splice site and the exonic sequence. In some aspects, the homology arms are comprised within any other sequence comprised in the genetic constructs disclosed herein. For example, the homology arms can be comprised in the intronic sequences, in the exonic sequences, in the splice sites, or any combination thereof. In some aspects, the homology arms can be comprised in a spacer sequence. As used herein, a spacer sequence is any sequence that is designed to be located between the 5'/3' splice sites and/or functional regions of the exonic sequence (e.g., IRES, payload, etc.). In some aspects, the homology arms are excised during circularization and are not comprised in the circRNAs. In some aspects, the homology arms are not excised during circularization and are comprised in the circRNAs.

[0208] The first and the second homology arms are the reverse complement of the other and are capable of pairing with one another. In some aspect, the pairing of the first and the second homology arms changes the spatial conformation of the genetic construct. For example, the pairing of the first and the second homology arms can bring the 3' splice site and the 5' splice site in proximity.

[0209] In some aspects, the homology arms comprise a nucleotide sequence having at least 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or 100% identity with a nucleotide sequence of any one of SEQ ID NOs: 72-73, 76-77, or 171-174. In some aspects, the homology arms consist of a nucleotide sequence of any one of SEQ ID NOs: 72-73, 76-77, or 171-174.

[0210] Non-limiting examples of homology arms that are useful for the present genetic constructs are shown in **Table 7**.

Table 7. Exemplary homology arms

SEQ ID NO	DESCRIPTION	SEQUENCE
72	5' homology (all) DNA	CCGTCGATTGTCCACTGGTC
73	3' homology (all) DNA	GACCAAGTGGACAATCGACGG
76	5' homology (all) RNA	CCGUCGAUUGUCCACUGGUC
77	3' homology (all) RNA	GACCAGUGGACAAUCGACGG
171	H of construct D - DNA	GGTTGTTGGAGATCTCTGGC
172	H of construct D - RNA	GGUUGUUGGAGAUUCUGGC
173	H' of construct D -DNA	GCCAGAGATCTCCAACAACC
174	H' of construct D -RNA	GCCAGAGAUCUCCAACAACC

[0211] In some aspects, the 5' the intronic sequence, the 3' intronic sequence, the 5' splice site, the 3' splice site, the exonic sequence, and optionally the homology arms are on a same polynucleotide molecule (e.g., on a same DNA or RNA molecule). Thus the 5' intronic sequence, the 3' intronic sequence, the 5' splice site, the 3' splice site, the exonic sequence, and optionally the homology arms can be in cis (*i.e.*, comprised in a same polynucleotide molecule).

[0212] In some aspects, the one or more exonic sequence is bookended by the 5' and the 3' intronic sequences. In some aspects, the 5' intronic sequence is upstream of the one or more exonic sequence. In some aspects, the 3' intronic sequence is downstream of the one or more exonic sequence.

II.B.6. Exons

[0213] In some aspects the exonic sequences (*i.e.*, one or more exon or fragment thereof) comprised in the RNA or DNA genetic constructs capable of forming circular RNAs disclosed herein (*i.e.*, pre-circRNAs or DNA genetic constructs from which pre-circRNAs can be transcribed) comprise one exon, or two or more exons.

[0214] In some aspects, the exonic sequence is (*i.e.*, in the pre-circRNA construct) or encodes for (*i.e.*, in the DNA genetic constructs from which the pre-circRNAs can be transcribed) a functional RNA. In some aspects, the functional RNA is not a protein-coding RNA. In some aspects, the functional RNA is a therapeutic RNA. In some aspects, the functional RNA is a protein-coding RNA. In some aspects, the protein-coding RNA is translated into a protein. In some aspects, the protein is a therapeutic protein.

[0215] In some aspects, an RNA or DNA genetic construct capable of forming circular RNAs disclosed herein (*i.e.*, a pre-circRNAs or a DNA genetic construct from which the pre-circRNAs can be transcribed) encodes one protein. In some aspects, an RNA or DNA genetic construct capable of forming circular RNAs disclosed herein (*i.e.*, a pre-circRNAs or a DNA genetic construct from which the pre-circRNAs can be transcribed) encodes two or more proteins. In some aspects, the sequences encoding the two or more proteins can be separated by a ribosomal skipping element or a nucleotide sequence encoding a protease cleavage site. Non limiting examples of self-cleaving 2A peptides are: thosea-asigna virus 2 A peptide (T2A), porcine teschovirus-1 2 A peptide (P2A), foot-and-mouth disease virus 2 A peptide (F2A), equine rhinitis A virus 2A peptide (E2A), cytoplasmic polyhedrosis virus 2A peptide (BmCPV 2A), or flacherie virus of B. mori 2A peptide (BmIFV 2A).

[0216] Non-limiting examples of proteins that can be encoded by the RNA or DNA genetic constructs capable of forming circular RNAs disclosed herein (*i.e.*, pre-circRNAs or DNA genetic constructs from which pre-circRNAs can be transcribed) are chimeric antigen receptor (CAR), T-cell receptor (TCR), B-cell receptor (BCR), immune cell activation or inhibitory receptor, recombinant fusion protein, chimeric mutant protein, or fusion protein, or a combination thereof. In some embodiments, the therapeutic protein comprises an antibody, nanobody, non-antibody protein, immune modulatory ligand, receptor, structural protein, growth factor ligand or receptor, hormone or hormone receptor, transcription factor, checkpoint inhibitor or agonist, Fc fusion protein, anticoagulant, blood clotting factor, chaperone protein, antimicrobial protein, structural protein, biochemical enzyme, tight junction protein, mitochondrial stress response, cytoskeletal protein, metal-binding protein, or small molecule. In some embodiments, the immune modulatory ligand comprises an interferon, cytokine, chemokine, or interleukin. In some embodiments, the

structural protein is a channel protein or nuclear pore protein, a cytokine, an immune checkpoint inhibitor, an agonist, an antagonist, a chimeric antigen receptor, one or more TCR chains, a secreted T cell or immune cell engager, a transcription factor, an immunosuppressive enzyme.

[0217] In some aspects, the protein is a cytokine. Non limiting examples of cytokines are, but are not limited to, IL-2, IL-4, IL-7, IL-10, IL-12, IL-15, IL-18, IL-21, IL-35, IL-12p70, IL-27beta, IFN- α , IFN- β , IFN γ , TGFbeta, or a functional fragment thereof.

[0218] In some aspects, the protein is an antigen binding protein, or a functional fragments thereof. In some aspects, the antigen binding protein is a trispecific antigen-binding protein (TRITEs), or bispecific antigen-binding protein (BITEs), or a functional fragments thereof. In some aspects, the antigen binding protein is an antibody, or a functional fragments thereof. Non limiting examples of antigens are, but are not limited to, Ebola virus glycoprotein Ebola virus glycoprotein, $\alpha 4$ integrin, Amyloid beta, Amyloid beta protofibrils, Amyloid beta, N3pG (N-terminal truncated), Angiopoietin-like 3, B. anthraxis PA, B7-H3, B-cell maturation antigen, BLYS, C1s, C5, CCR4, CD11a, CD19, CD20, CD22, CD3, CD30, CD33, CD38, CD4, CD52, CD62 (aka P-selectin), CD79b, CGRP, CGRP receptor, Clostridium difficile enterotoxin B, Complement 5, CTLA-4, Dabigatran, Ebola virus, EGFR, EGFR, cMET, Endotoxin, EpCAM, EPCAM/CD3, Factor Ixa, X, FcRn, FGF23, Folate receptor alpha, G protein-coupled receptor 5D, CD3, GD2, gp100, CD3, GPIIb/IIIa, HER2, IFNAR1, IFN γ , IgE, IGF-1R, IL-12/23, IL-13, IL-17a, IL-17A, F, IL-17R, IL-1 β , IL-23 p19, IL-23p19, IL-2R, IL-36 receptor, IL-4R α , IL-5, IL-5R α , IL-6, IL-6R, LAG-3, MASP-2, Nectin-4, PCSK9, PD-1, PDGFR α , PD-L1, Plasma kallikrein, RANK-L, RSV, SARS-CoV-2, Sclerostin, SLAMF7, Thymic stromal lymphopoietin, Tissue factor, Tissue factor pathway inhibitor, TNF, TROP-2, VEGF, VEGF-A, VEGF-A, Ang-2, VEGFR2, von Willebrand factor, and $\alpha 4\beta 7$ integrin. Non limiting examples of antibodies are, but are not limited to, Muromonab-CD3, Efalizumab, Tositumomab-I131, Nebacumab, Edrecolomab, Catumaxomab, Daclizumab, Olaratumab, Abciximab, Rituximab, Basiliximab, Palivizumab, Infliximab, Trastuzumab, Adalimumab, Ibritumomab tiuxetan, Omalizumab, Cetuximab, Bevacizumab, Natalizumab, Panitumumab, Ranibizumab, Eculizumab, Certolizumab pegol, Ustekinumab, Canakinumab, Golimumab, Ofatumumab, Tocilizumab, Denosumab, Belimumab, Ipilimumab, Brentuximab vedotin, Pertuzumab, Ado-trastuzumab emtansine, Raxibacumab, Obinutuzumab, Siltuximab, Ramucirumab, Vedolizumab, Nivolumab, Pembrolizumab, Blinatumomab, Alemtuzumab, Evolocumab, Idarucizumab, Necitumumab, Dinutuximab, Secukinumab, Mepolizumab, Alirocumab, Daratumumab, Elotuzumab, Ixekizumab, Reslizumab, Bezlotoxumab, Atezolizumab, Obiltoxaximab, Brodalumab, Dupilumab, Inotuzumab ozogamicin, Guselkumab,

Sarilumab, Avelumab, Emicizumab, Ocrelizumab, Benralizumab, Durvalumab, Gemtuzumab ozogamicin, Erenumab, erenumab-aooe, Galcanezumab, galcanezumab-gnlm, Burosumab, burosumab-twza, Lanadelumab, lanadelumab-flyo, Mogamulizumab, mogamulizumab-kpkc, Tildrakizumab, tildrakizumab-asmn, Fremanezumab, fremanezumab-vfrm, Ravulizumab, ravulizumab-cwvz, Cemiplimab, cemiplimab-rwlc, Ibalizumab, ibalizumab-uiyk, Emapalumab, emapalumab-lzsg, Moxetumomab pasudotox, moxetumomab pasudotox-tdfk, Caplacizumab, caplacizumab-yhdp, Risankizumab, risankizumab-rzaa, Polatuzumab vedotin, polatuzumab vedotin-piiq, Romosozumab, romosozumab-aqqg, Brolucizumab, brolucizumab-dbl, Crizanlizumab, crizanlizumab-tmca, Enfortumab vedotin, enfortumab vedotin-ejfv, [fam-]trastuzumab deruxtecan, fam-trastuzumab deruxtecan-nxki, Isatuximab, isatuximab-irfc, Belantamab mafodotin, belantamab mafodotin-blmf, Sacituzumab govitecan, sacituzumab govitecan-hziy, Tafasitamab, tafasitamab-cxix, Satralizumab, satralizumab-mwge, Eptinezumab, eptinezumab-jjmr, Inebilizumab, inebilizumab-cdon, Teprotumumab, teprotumumab-trbw, Evinacumab, Dostarlimab, dostarlimab-gxly, Amivantamab, amivantamab-vmjw, Tralokinumab, tralokinumab-ldrm, Anifrolumab, anifrolumab-fnia, Loncastuximab tesirine, loncastuximab tesirine-lpyl, Atoltivimab, maftivimab, odesivimab-ebgn, Naxitamab-gqgk, Margetuximab-cmkb, Ansuvimab-zykl, Aducanumab, aducanumab-avwa, Regdanvimab, Sotrovimab, Tisotumab vedotin, tisotumab vedotin-tftv, Bimekizumab, Casirivimab + imdevimab, Tezepelumab, tezepelumab-ekko, Faricimab, faricimab-svoa, Sutimlimab, sutimlimab-jome, Tixagevimab, cilgavimab, Spesolimab, Nirsevimab, Teplizumab, teplizumab-mzwv, Ublituximab, Tebentafusp, tebentafusp-tebn, Relatlimab, Mosunetuzumab, Teclistamab, Tremelimumab, Mirvetuximab soravtansine, mirvetuximab soravtansine-gynx, Lecanemab, Toripalimab, Trastuzumab duocarmazine, Epcoritamab, Mirikizumab, Glofitamab, Pozelimab, Lebrikizumab, Talquetamab, Rozanolixizumab, Cosibelimab, Concizumab, Elranatamab, Sugemalimab, Penpulimab, Donanemab, Sintilimab, Tislelizumab, Retifanlimab, Narsoplimab, and Omburtamab.

[0219] In some aspects, the protein is a chimeric antigen receptors (CARs or CAR-Ts). Non limiting examples of CARs are, but are not limited to, CAR comprising an antigen binding domain specific for an antigen selected from the group consisting of CD19, CD123, CD22, CD30, CD171, CS-1, C-type lectin-like molecule- 1, CD33, epidermal growth factor receptor variant III (EGFRvIII), ganglioside G2 (GD2), ganglioside GD3, TNF receptor family member B cell maturation (BCMA), Tn antigen ((Tn Ag) or (GalNAc-Ser/Thr)), prostate-specific membrane antigen (PSMA), Receptor tyrosine kinase-like orphan receptor 1 (ROR1), Fms-Like Tyrosine

Kinase 3 (FLT3), Tumor-associated glycoprotein 72 (TAG72), CD38, CD44v6, Carcinoembryonic antigen (CEA), Epithelial cell adhesion molecule (EPCAM), B7H3 (CD276), KIT (CD117), Interleukin-13 receptor subunit alpha-2, mesothelin, Interleukin 11 receptor alpha (IL-11R α), prostate stem cell antigen (PSCA), Protease Serine 21, vascular endothelial growth factor receptor 2 (VEGFR2), Lewis(Y) antigen, CD24, Platelet-derived growth factor receptor beta (PDGFR-beta), Stage-specific embryonic antigen-4 (SSEA-4), CD20, Folate receptor alpha, HER2, HER3, Mucin 1, cell surface associated (MUC1), epidermal growth factor receptor (EGFR), neural cell adhesion molecule (NCAM), Prostase, prostatic acid phosphatase (PAP), elongation factor 2 mutated (ELF2M), Ephrin B2, fibroblast activation protein alpha (FAP), insulin-like growth factor 1 receptor (IGF -I receptor), carbonic anhydrase IX (CAIX), Proteasome (Prosome, Macropain) Subunit, Beta Type, 9 (LMP2), glycoprotein 100 (gp100), oncogene fusion protein consisting of breakpoint cluster region (BCR) and Abelson murine leukemia viral oncogene homolog 1 (Abl) (bcr-abl), tyrosinase, ephrin type-A receptor 2 (EphA2), Fucosyl GM1, sialyl Lewis adhesion molecule (sLe), ganglioside GM3, transglutaminase 5 (TGS5), high molecular weight-melanoma-associated antigen (HMWMAA), o-acetyl-GD2 ganglioside (OAcGD2), Folate receptor beta, tumor endothelial marker 1 (TEM1/CD248), tumor endothelial marker 7-related (TEM7R), claudin 6 (CLDN6), thyroid stimulating hormone receptor (TSHR), G protein-coupled receptor class C group 5, member D (GPRC5D), chromosome X open reading frame 61 (CXORF61), CD97, CD179a, anaplastic lymphoma kinase (ALK), Polysialic acid, placenta-specific 1 (PLAC1), hexasaccharide portion of globoH glycosphingolipid (GloboH), mammary gland differentiation antigen (NY-BR-1), uroplakin 2 (UPK2), Hepatitis A virus cellular receptor 1 (HAVCR1), adrenoceptor beta 3 (ADRB3), pannexin 3 (PANX3), G protein-coupled receptor 20 (GPR20), lymphocyte antigen 6 complex, locus K 9 (LY6K), Olfactory receptor 51E2 (OR51E2), TCR Gamma Alternate Reading Frame Protein (TARP), Wilms tumor protein (WT1), Cancer/testis antigen 1 (NY-ESO-1), Cancer/testis antigen 2 (LAGE-1a), MAGE family members (including MAGE-A1, MAGE-A3 and MAGE-A4), ETS translocation-variant gene 6, located on chromosome 12p (ETV6-AML), sperm protein 17 (SPA17), X Antigen Family, Member 1 A (XAGE1), angiopoietin-binding cell surface receptor 2 (Tie 2), melanoma cancer testis antigen-1 (MAD-CT-1), melanoma cancer testis antigen-2 (MAD-CT-2), Fos-related antigen 1, tumor protein p53 (p53), p53 mutant, prostatein, surviving, telomerase, prostate carcinoma tumor antigen-1, melanoma antigen recognized by T cells 1, Rat sarcoma (Ras) mutant, human Telomerase reverse transcriptase (hTERT), sarcoma translocation breakpoints, melanoma inhibitor of apoptosis (ML-IAP), ERG (transmembrane protease, serine 2 (TMPRSS2) ETS fusion gene), N-

Acetyl glucosaminyl-transferase V (NA17), paired box protein Pax-3 (PAX3), Androgen receptor, Cyclin B1, v-myc avian myelocytomatosis viral oncogene neuroblastoma derived homolog (MYCN), Ras Homolog Family Member C (RhoC), Tyrosinase-related protein 2 (TRP-2), Cytochrome P450 1B1 (CYP1B1), CCCTC-Binding Factor (Zinc Finger Protein)-Like, Squamous Cell Carcinoma Antigen Recognized By T Cells 3 (SART3), Paired box protein Pax-5 (PAX5), proacrosin binding protein sp32 (OY-TES1), lymphocyte-specific protein tyrosine kinase (LCK), A kinase anchor protein 4 (AKAP-4), synovial sarcoma, X breakpoint 2 (SSX2), Receptor for Advanced Glycation Endproducts (RAGE-1), renal ubiquitous 1 (RU1), renal ubiquitous 2 (RU2), legumain, human papilloma virus E6 (HPV E6), human papilloma virus E7 (HPV E7), intestinal carboxyl esterase, heat shock protein 70-2 mutated (mut hsp70-2), CD79a, CD79b, CD72, Leukocyte-associated immunoglobulin-like receptor 1 (LAIR1), Fc fragment of IgA receptor (FCAR or CD89), Leukocyte immunoglobulin-like receptor subfamily A member 2 (LILRA2), CD300 molecule-like family member f (CD300LF), C-type lectin domain family 12 member A (CLEC12A), bone marrow stromal cell antigen 2 (BST2), EGF-like module-containing mucin-like hormone receptor-like 2 (EMR2), lymphocyte antigen 75 (LY75), Glypican-3 (GPC3), Fc receptor-like 5 (FCRL5), MUC16, 5T4, 8H9, α 5 β 1 integrin, α 5 β 1 integrin, alphafetoprotein (AFP), B7-H6, ca-125, CA9, CD44, CD44v7/8, CD52, E-cadherin, EMA (epithelial membrane antigen), epithelial glycoprotein-2 (EGP-2), epithelial glycoprotein-40 (EGP-40), ErbB4, epithelial tumor antigen (ETA), folate binding protein (FBP), kinase insert domain receptor (KDR), k-light chain, L1 cell adhesion molecule, MFJ18, NKG2D, oncofetal antigen (h5T4), tumor/testis-antigen IB, GAGE, GAGE-1, BAGE, SCP-1, CTZ9, SAGE, CAGE, CT10, MART-1, immunoglobulin lambda-like polypeptide 1 (IGLL1), Hepatitis B Surface Antigen Binding Protein (HBsAg), viral capsid antigen (VC A), early antigen (EA), EBV nuclear antigen (EBNA), HHV-6 p41 early antigen, HHV-6B U94 latent antigen, HHV-6B p98 late antigen, cytomegalovirus (CMV) antigen, large T antigen, small T antigen, adenovirus antigen, respiratory syncytial virus (RSV) antigen, haemagglutinin (HA), neuraminidase (NA), parainfluenza type 1 antigen, parainfluenza type 2 antigen, parainfluenza type 3 antigen, parainfluenza type 4 antigen, Human Metapneumovirus (HMPV) antigen, hepatitis C virus (HCV) core antigen, HIV p24 antigen, human T-cell lymphotropic virus (HTLV-1) antigen, Merkel cell polyoma virus small T antigen, Merkel cell polyoma virus large T antigen, Kaposi sarcoma-associated herpesvirus (KSHV) lytic nuclear antigen and KSHV latent nuclear antigen.

[0220] In some aspects, the protein is a transcription factor. Non limiting examples of transcription factors are, but are not limited to, Forkhead box P3 (Foxp3), and signal transducer

and activator of transcription (STAT, e.g., STAT1, STAT2, STAT3, STAT4, STAT 5 (including STAT5 A and STAT5B), and STATE).

II.B.7. IRES

[0221] In some aspects, the RNA or DNA constructs disclosed herein further comprise an internal ribosome entry site (IRES). In some aspects, the IRES is comprised in an exon.

[0222] First identified as a feature Picorna virus RNA, IRES plays an important role in initiating protein synthesis in absence of the 5' cap structure. An IRES can act as the sole ribosome binding site, or can serve as one of multiple ribosome binding sites of an mRNA. Nucleic acids or mRNA containing more than one functional ribosome binding site can encode several peptides or polypeptides that are translated independently by the ribosomes ("multicistronic nucleic acid molecules"). When nucleic acids or mRNA are provided with an IRES, further optionally provided is a second translatable region.

[0223] Examples of IRES sequences that can be used according to the disclosure include without limitation, those from picornaviruses (e.g., FMDV), pest viruses (CFFV), polio viruses (PV), encephalomyocarditis viruses (ECMV), foot-and-mouth disease viruses (FMDV), hepatitis C viruses (HCV), classical swine fever viruses (CSFV), murine leukemia virus (MLV), simian immune deficiency viruses (SIV) or cricket paralysis viruses (CrPV). In some aspects, the genetic constructs disclosed herein comprise an IRES. In some embodiments, the genetic constructs disclosed herein comprises an IRES having at least 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or 100% identity with a nucleotide sequence of SEQ ID NOs: 87-110 or 145-168. Modifications of IRES and accessory sequences are disclosed herein to increase or reduce IRES activities, for example, by truncating the 5' and/or 3' ends of the IRES, adding a spacer 5' to the IRES, modifying the 6 nucleotides 5' to the translation initiation site (Kozak sequence), modification of alternative translation initiation sites, and creating chimeric/hybrid IRES sequences. In some aspects, the IRES sequence in the genetic constructs disclosed herein comprises one or more of these modifications relative to a native IRES (e.g., a native IRES disclosed in **Table 8** (SEQ ID NOs: 87-110 or 145-168)).

Table 8. Exemplary IRES

SEQ ID NO	DESCRIPTION	SEQUENCE
87	CVB3	TTAAACAGCCTGTGGGTTGATCCACCCACAGGCCCATTTGGGCGCTAGCACTCTGGTA TCACGGTACCTTTGTGCGCCTGTTTATACCCCTCCCCCAACTGTAACCTTAGAAGTAA CACACACCGATCAACAGTCAGCGTGGCACACCAGCCACGTTTGTATCAAGCACTTCTGT TACCCCGGACTGAGTATCAATAGACTGCTCACGCGGTTGAAGGAGAAAGCGTTCGTTAT CCGGCCAACTACTTCGAAAAACCTAGTAACACCGTGGAAGTTGCAGAGTGTTCGCTCA GCACTACCCAGTGTAGATCAGGTCGATGAGTCACCGCATTCCCCACGGGCGACCGTGG

		CGGTGGCTGCGTTGGCGGCCTGCCCATGGGGAAACCCATGGGACGCTCTAATACAGACA TGGTGCGAAGAGTCTATTGAGCTAGTTGGTAGTCCTCCGGCCCCCTGAATGCGGCTAATC CTAACTGCGGAGCACACACCCTCAAGCCAGAGGGCAGTGTGTCGTAACGGGCAACTCTG CAGCGGAACCGACTACTTTGGGTGTCCGTGTTTCATTTTATTCCTATACTGGCTGCTTA TGGTGACAATTGAGAGATCGTTACCATATAGCTATTGGATTGGCCATCCGGTGACTAAT AGAGCTATTATATATCCCTTTGTTGGGTTTATACCACTTAGCTTGAAAGAGGTTAAAC ATTACAATTCAATTGTTAAGTTGAATACAGCAA
88	EV71	TTAAACAGCTGTGGGTTGTCAACCCACCCACAGGGTCCACTGGGCGCTAGTACACTGGT ATCTCGGTACCTTTGTACGCCGTGTTTATACCCCTCCCTGATTTGCAACTTAGAAGCA ACGCAAACAGATCAATAGTAGGTGTGACATACCAGTCGCATCTTGATCAAGCACTTCT GTATCCCCGGACCGAGTATCAATAGACTGTGCACACGGTTGAAGGAGAAAACGTCCGTT ACCCGGCTAACTACTTTCGAGAAGCCTAGTAACGCCATTGAAGTTGCAGAGTGTTTCGCT CAGCACTCCCCCGTGTAGATCAGGTGCGATGAGTCACCGCATTTCCACACGGGCGACCGT GGCGGTGGCTGCGTTGGCGGCCCTGCCTATGGGGTAACCCATAGGACGCTCTAATACGGA CATGGCGTGAAGAGTCTATTGAGCTAGTTAGTAGTCCTCCGGCCCCCTGAATGCGGCTAA TCCTAACTGCGGAGCACATACCCCTTAATCCAAAGGGCAGTGTGTCGTAACGGGCAACTC TGCAGCGGAACCGACTACTTTGGGTGTCCGTGTTTCTTTTATTCTTGATTGGCTGCT TATGGTGACAATTAAAGAATTGTTACCATATAGCTATTGGATTGGCCATCCAGTGTCAA ACAGAGCTATTGTATATCTCTTTGTTGGATTTCACACCTCTCACTCTTGAAACGTTACAC ACCCCAATTACATTATACTGCTGAACACGAAGCG
89	CVB5	TTAAACAGCCTGTGGGTTGTTCCACCCGACGGGCCACTGGGCGCCAGCACTCTGGT ATCACGGTACCTTTGTGCGCCTGTTTAAACCCCTCTCCCAATTTGAAACTTAGAAGC AATACACCTCGATCAATAGTAGGCATGACACGCCAGCCATGTCTTGATCAAGCACTTCT GTTTCCCGGACTGAGTATCAATAAAGTGTTCGCGCGGTGGAAGGAGAAAACGTCCGTT ACCCGACTAACTACTTCGAGAAAACCCAGTAACACCATGGAAATTGCGGAGTGTTTCACT CAGCACATTTCCAGTGTAGATCAGGTGCGATGAGTCACCGCATTTCCACACGGGTGACCGT GGCGGTGGCTGCGCTGGCGGCCCTGCCCATGGGGCAACCCATGGGACGCTTCAATATGGA CATGGTGTGAAGAGTCTATTGAGCTAGTTAGTAGTCCTCCGGCCCCCTGAATGCGGCTAA TCCTAACTGCGGAGCACGTGCCTCCATTCCAGGGGGTGGCGTGTGTAACGGGCAACTC TGCAGCGGAACCGACTACTTTGGGTGTCCGTGTTTCTTTTAAATTTTATACTGGCTGCTT ATGGTGACAATTGAAAGATTGTTGCCATATAGCTATTGGATTGGCCATCCGGTATCCAA CAGAGCAATTGTGTACCTTTTGTGGATTGTACCATTACCAGAACAAGTTTTCATA CACTGTGTACATTATTAGACTAAACACAGAAAA
90	EchoV11	TTAAACAGCCTGTGGGTTGTACCCACCCACAGGGGCCACTGGGCGCTAGCACACTGGT ATCACGGTACCTTTGTGCGCCTGTTTATACCCCTTCCCGCAACCGCAAATTTAGAAG CAAAGCTAACCCGATCGATAGCGGATGCGCATGCCAGCCGCAATTTGATCAAGTACTTC TGTTCCTCCCGGACCGAGTATCAATAGACTGCTCACGCGGTTGAAGGAGAAAACGTCCGT TACCCGACCAACTACTTCGAGAAACCTAGTAACATCATGAATGTTGCAGGGCGTTTCGA TCAGCACGACCCCTGGTGTAGATCAGGCTGATGAGTCACCGCATTTCCACACGGGTGACCG TGGCGGTGGCTGCGTTGGCGGCCCTGCCTATGGGGTGACCCATAGGACGCTCTAATACGG ACATGGTGCGAAGAGTCTATTGAGCTAGTTGGTAGTCCTCCGGCCCCCTGAATGCGGTTA ATCCTAACTGCGGACGACATACCCCTAATCCAGGGGCAGTGTGTCGTAACGGGCAACT CTGCAGCGGAACCGACTACTTTGGGTGTCCGTGTTTCTTTTATTTTATACTGGCTGC TTATGGTGACAATCTCAGAGTTGTTACCATATAGCTATTGGATTGGCCATCCGGTGAGC AACAGAGCTGTCAATTATCAGTTGTTGGCTTTATACCTCTAAATCACACGGTTTTTTTT TTTTTGGAACGCTTGATTATCATCTTAACCCCTCAATAAGGCAA
91	EVB107	TTAAACAGCCTGTGGGTTGTTCCACCCGACGGGCCACTGGGCGCTAGCACACTGGT ATCCCGGTACCCCTTGTGCGCCTGTTTATATACCCCTCCCCCTTATGTAACCTTAGAAGTA TGATTCAAACGGTCGACAGGCGGCTCAGTGCACCAACTGAGTCATGACCAAGCACTTCT GTTACCCCGGACTGAGTATCAATAAGCTGTTTACACGGCTGAAGGAGAAAACGTTCGTT ACCCGGCCAATTACTTCGAGAAAACCTAGTACCACCATGAAGGTTGCGCAGTGTTTCGCT CCACACAACCCAGTGTAGATCAGGTGCGATGAGTCACCGCATTTCCACACGGGCGACCGT GGCGGTGGCTGCGTTGGCGGCCCTGCCCATGGGGCAACCCATGGGACGCTTCAATACGA CATGGTGTGAAGAGTCTATTGAGCTAATTGGTAGTCCTCCGGCCCCCTGAATGCGGCTAA TCCCACTGTGGAGCAGATACTCACAACACAGTGAGCGGTCTGTGTAACGGGCAACTC CGCAGCGGAACCGACTACTTTGGGTGTCCGTGTTTCTTTTATTCTTACATTGGCTGCT TATGGTGACAATTGACAAATTGTTACCATATAGCTATTGGATTGGCCATCCGGTGACAA ACAGAGCTATTGTTTACTTGTGTTGTTGTTTTCATACCATTAATTAAGGTCTTAGAA ACTCTCAACTTTATTTTGACACTCAATACAGCAA
92	PV3	TTAAACAGCTCTGGGTTGCTCCCAACCCGACGGGCCACGTGGGCGGCCAGTACTCTGG TATTGCGGTACCTTTGTACGCCGTGTTTATACCTCCCTCCCCCGTGCAACTTTAGAAGA

		AATCCACAAAGTTCAATAGAGGGGGTGCAAACCAGCACCACCACGAACAAGCACTTCTGT TTTCCCCGGTGATGTCGTATAAGCTGTACCCACGGCTGAAGGCGACGGATCCGTTATCC GCTTGAGTACTTCGAGAAGCCTAGTATTACCTTGGGAATCTTCGACGCGTTGCGCTCAGC ACTCTACCCCGAGTGTAGCTTAGGTCGATGAGTCTGGGCACGCCCCACCAGGCGACGGTG GCCCAGGCTGCGTTGGCGGCCCTACCCATGGCTATCACCATGGGACGCTAGTTGTGAACA AGGTGTGAAGAGCCTATTGAGCTACCCAAGAGTCCCTCCGGCCCCCTGAATGCGGCTAATC CTAACCCAGGAGCAAGTGTCTCAACCCAGGGGATGGCTTGTCTGAACGCGAAAGTCTG TGGCGGAACCGACTACTTTGGGTGTCCGTGTTTCTTTTATTTTATGTATGGCTGCTT ATGGTGACAATCAAAGGTTGTTACCATAAAGCAATTTGGATTGGCCATCCGGTGAGAAT CAAACATATTATCTACCTGTTTGTGGGTTTCTTCTTTTACTTGAACAATACCTCTAA TAATAACTGCTATATTGTCAATAAGACATTATCATCACA
93	CVB1	TTAAAACAGCCTGTGGGTTGTTCCCAACCACAGGCCCATTTGGGCGCTAGCACTCTGGTA TCACGGTACCTTTGTGCGCCTGTTTACATCCCTCCCCAAATTGTAATTTAGAAGTTT CACACACCGATCATTAGCAAGCGTGGCACACCAGCCATGTTTGTATCAAGCACTTCTGT TACCCCGGACTGAGTATCAATAGACCGCTAACGCGGTTGAAGGAGAAAACGTTTCGTTAC CCGGCCAACCTACTTCGAAAAACCTAGTAACACCATGGAAGTTGCGGAGTGTTCGCTCA GCACTACCCAGTGTAGATCAGGTGCGATGAGTCACCGCGTTCCCCACGGGCGACCGTGG CGGTGGCTGCGTTGGCGGCCCTGCCTACGGGGAAACCCGTAGGACGCTCTAATACAGACA TGGTGCGAAGAGTCTATTGAGCTAGTTGGTAATCCTCCGGCCCCCTGAATGCGGCTAATC CTAACTGCGGAGCACATACCTCAAAACCAGGGGGCAGTGTGTCTGAACGGGCAACTCTG CAGCGGAACCGACTACTTTGGGTGTCCGTGTTTCATTTTATTTCTATACCTGGCTGCTTA TGGTGACAATTGACAGGTTGTTACCATATAGTTATTGGATTGGCCATCCGGTGACTAAC AGAGCAATTATATATCTCTTTGTTGGGTTTATACCACTTAGCTTGAAGAGAGGTTAAAAC ACTACATCTCATCATTTAAACTAAATACAACAAA
94	EchoVE1 5	TTAAAACAGCCTGTGGGTTGTACCCACCCACAGGGCCCACTGGGCGCTAGCACTCTGGT ATTACGGTACCTTTGTGCGCCTGTTTATATACCCCGCCCCAAGCAAACGCTAGATGTA ACGCACTTATGATCAATAGCAGGCGTGGCACTCCAGCCACGTTATGATCAAGCACTTCT GTCTCCCCGGACCGAGTATCAATAGACTGCTCACGCGGTCGAAGGAGAAAAACGTTTCGTT ACCCGACCAGCTACTTCGAGAAACCTAGTAACCTCCATGGAGGTTGCAGAGTGTTCGTT CAGCACTTCCCCCGTGTAGATCAGGCTGATGAGTCACCGCGTTCTCTACGGGCGACCGT GGCGGTGGCTGCGTTGGCGGCCCTGCCTATGGGTATACCCATAGGACGCTCTAATACTGA CATGGTGCGAAGAGTCTATTGAGCTAGTTGGTAGTCTCTCCGGCCCCCTGAATGCGGCTAA TCCTAACTGCGGAGCACATACCCCAACCCAGGGGGCAGTGTGTCTGAACGGGTAACCTC TGCAGCGGAACCGACTACTTTGGGTGTCCGTGTTTCTTTTATTCTTACATTGGCTGCT TATGGTGACAATTGAGAGATTGTTACCATATAGCTATTGGATTGGCCATCCAGTGACTA ACAGAGCTATTATTTACCTTTTTGTTGGCTTCGTATCACTTGGTTTAAAGAGGTTAGT ACTTTATATTGCATTATATTACTAAACACGAGAAA
95	EchoV1	TTAAAACAGCCTGTGGGTTGTTCCCAACCACAGGGCCCACTGGGCGTTAGCACACTGGT ATCACGGTACCTTTGTGCGCCTGTTTATACTCCCCCCCCCTAAGGAAACTTTAGAAGCA AAGCAATTGTGATCAATAGTGGGTATGGCACACCAGTCATATCTTGATCAAGCACTTCT GTTCCCCCGGACTTAGTACCAATAGACTGCTCAAGCGGTTGAAGGGGAAAAACGTTTCGTT ATCCGGCCAACCTACTTCGAGAAACCTAGTAGCACCATGAAAGTTGCGGAGTGTTCGCT CAGCACTTCCCCCGTGTAGATCAGGCTGATGAGTCACCGTATTTCCCCACGGGCGACCGT GACGGTGGCTGCGTTGGCGGCCCTGCCATGGGGTAACCCATGGGACGCTCTAAAACAGA CACGGTGCGAAGAGTCTATTGAGCTAGTTGGTAGTCTCTCCGGCCCCCTGAATGCGGCCAA TCCTAACTGCGGAGCACATACCTCCAATCCAGGGAGCAGTGTGTCTGAATGGGTAACCTC TGCAGCGGAACCGACTACTTTGGGTGTCCGTGTTTCTTTTATTCTCACATTGACTGCT TATGGTGACAATTGAAAGATTGTTACCATATAGCTATTGGATTGGTCATCCGGTGAGCA ATAGAGCTATTGTTTATCAATTTGTTGGATTGTACCCTCAACTTTTCTGTTTTGAGA ACACTCAACTACATCTTACTGTAAACACATCAAA
96	EVD94	TTAAAACAGCCTGTGGGTTGTTCCCAACCACAGGGCCCACTGGGCGGCCAGTACTCTGG TATCACGGTACCTTTGTACGCCCTGTTTATATCCCTTCCCCCGCAACTTAGAAGAAAA CAAATCAAGTTCACTAGGAGGGGTACAAACCAGTACCACCAGAACAAAGCACTTCTGT TTCCCCGGTGATGTGCTATAGACTGTAACCACCGGTTGAAAACGATTGATCCGTTATCCG CTCTTGTACTTCGAAAAGCCAGTATCACCTTGGAACTTTCGATGCGTTGCGCTCAGCA CTCAACCCAGAGTGTAGCTTAGGTCGATGAGTCTGGACACTCCTCACGGGCGACGGTG GTCCAGGCTGCGTTGGCGGCCCTACCTGTGGTCCAAAGCCACAGGACGCTAGTTGTGAAC AAGGTGTGAAGAGCCTATTGAGCTACAAGAGAATCCTCCGGCCCCCTGAATGCGGCTAAT CCTAACCCAGGAGCAAGGGTACAAACCAGTGTATATCTTGTCTTAACGCGCAAGTCT GTGGCGGAACCGACTACTTTGGGTGTCCGTGTTTCTTTTGTTCCTATCATGGCTGCTT ATGGTGACAATCTAAGATTGTTATCATATAGCTGTTGGATTGGCCATCCGGTAATTTAT

		TGAGATTGAGCATTTGCTTGTCTTCAACAATTTACCTATTTCATTGCATTTAGCA GTCAAA
97	PV1	TTAAACAGCTCTGGGGTTGTACCCACCCAGAGGCCACGTGGCGGCTAGTACTCCGG TATTGCGGTACCCCTGTACGCCCTGTTTTATACTCCCTTCCCGTAACCTAGACGCACAAA ACCAAGTTCAATAGAAGGGGGTACAAACCAGTACCACCAGAACAAAGCACTTCTGTTTC CCCGGTGATGTCGTATAGACTGCTTGCCTGGTTGAAAGCGACGGATCCGTTATCCGCTT ATGTACTTCGAGAAGCCAGTACCACCTCGGAATCTTCGATGCGTTGCGCTCAGCACTC AACCCAGAGTGTAGCTTAGGCTGATGAGTCTGGACATCCCTCACCGGTGACGGTGGTC CAGGCTGCGTTGGCGGCCTACCTATGGCTAACGCCATGGGACGCTAGTTGTGAACAAGG TGTGAAGAGCCTATTGAGCTACATAAGAATCCTCCGGCCCCCTGAATGCGGCTAATCCCA ACCTCGGAGCAGGTGGTCACAAACCAGTGAATGGCCTGTCTGTAACGCGCAAGTCCGTGG CGGAACCGACTACTTTGGGTGTCCGTGTTTCTTTTATTTTATTGTTGGCTTATGGT GACAATCACAGATTGTTATCATAAAGCGAATTGGATTGGCCATCCGGTGAAAGTGAGAC TCATTATCTATCTGTTTGTCTGGATCCGCTCCATTGAGTGTGTTTACTCTAAGTACAATT TCAACAGTTATTTCAATCAGACAATTGTATCACC
98	PV2	TTAAACAGCTCTGGGGTTGTTCCACCCACCCAGAGGCCACGTGGCGGCCAGTACACTGG TATCGCGGTACCTTTGTACGCCCTGTTTTATACTCCCTTCCCCCGTAACCTAGAAGCACA ACGTCCAAGTTCAATAGAAGGGGGTACAAACCAGTACCACCAGAACAAAGCACTTCTGT TCCCCCGGTGAGGCTGTATAGGCTGTTTCCACGGCTAAAAGCGGCTGATCCGTTATCCG CTCATGTACTTCGAGAAGCCTAGTATCACCTTGAATCTTCGATGCGTTGCGCTCAACA CTCAACCCAGAGTGTAGCTTAGGTCGATGAGTCTGGACGTTTCTCACCGGCGACGGTG GTCCAGGCTGCGTTGGCGGCCTACCTGTGGCCCAAAGCCACAGGACGCTAGTTGTGAAC AAGGTGTGAAGAGCCTATTGAGCTACCTGAGAGTCTCTCCGGCCCCCTGAATGCGGCTAAT CCTAACCGAGCAGGCGAGTGGCAATCCAGCGACCGAGCCTGTCTGTAACGCGCAAGTTC GTGGCGGAACCGACTACTTTGGGTGTCCGTGTTTCTTTTATTTTACAATGGCTGCTT ATGGTGACAATTATTGATAGTTATCATAAAGCAAATTGGATTGGCCATCCGGTGAGAAT TTGATTATTAAATTACTCTCTTGTGGGATTGCTCTTTGAAATCCTGTGCACTCACAC CTATTGGAATTACCTCATTGTTGAGATATTATTACCACT
99	CVA24	ttaaaacagctctggggttggttcccaccccagaggcccacgtggcgccagtaactccgg tattgcggtacccttgtaagcctgttttacactcccttcccggtaacttagaagtaaa caaaccaggttcaatagaaggggtacaaaccagtaaccaccagaaacagcacttctgt tccccgggtgacgttgatagactgtacccacgggtgaaaaaactgatccgttatccg ctcttgtaacttcgagaagcccagtatcaccttggaatcttcgatgctgtgcgctcagca ctcaaccccagagtgtagcttaggtcgatgagctctggacaatcctcacggcgacgggtg gtccaggtgctgtggcgccctacctgtggcccaaagccacaggacgctagttgtgaac aaggtgtgaagagcctattgagctacgagagagtcctccggcccctgaatgcccgtaat cccaaccacggagcaggtgatcacaagccagtgattagcctgtcgtaacgcgcaagtct gtggcggaaccgactactttgggtgtccgtgttctctttattttattatggctgcttat ggtgacaatcacagattgttatcataaagcgacttggttgcccatccagtgaaaataa aacacatcatttacctgattgttggttactccgctgactatctactcaccaggtcctt attaaggttattctgttgattaggtcacaatttcaca
100	CVB6	ttaaaacagcctgtgggttgtaaccaccccagggcccactggcgctagcactctgggt attgcggtacctttgtgcccgtgttttataccaccctcccccaacgtaacttagaagcct gacatatacgggtcacaagccagcccagtaacgccaactgggtaccgaccaagcacttctg ttaccccgactgagtatcaataagctgccaaagcggtgaaggagaaaacgttcgtta cccggccaattacttcgagaaacccagtaacaccatgaaagttgcgtagcggttctattc agcacaatcccagtgtagatcaggtcgatgagtcaccgcattccccacaggcgactgtg gcggtggctgctgtggcgccctgcccattggggcaaacccatgggacgcttcaatactga catggtgccaagagtcattagagctaattggtagtcctccggcccctgaatgcccgttaa tcctaactgaggagcaggcactcgcagaccagcgagcagcttgctgtaattgggcaactc cgagcggaaccgactactttgggtgtccgtgttctctatttctttactggctgct tatggtgacaattgagaaattgttaccatagctattggagtgcccatccgggtgacta acagagcaataataatattgtttgttggtttgtgcccacttaatttgaaaggttggttaag acgctcaatctgatactacttctaaataaagcaaa
101	EV89	ttaaaacagcctctggggttgacccaccccagggcccactggcgctagcacactgggt atcacggtaacctttgtgcccgtgttttatacaccctcccccttgtaacttagaagaaa caaacaaacgatcaatagtaggcttagcactccagctaagtcogataagcacttctg ttaccccgaccgagtatcaataggtgtacccacgggtgaagagaaaaacgttcgtta cccggctatctacttcgggaaacccagtaaccatcatgaagattgcaaagcggttctgctc agcacaccccagtgtagatcaggtcgatgagtcactgcattccccacaggcgactgtg gcagtggtgctgtggcgccctgcccattggggaaacccataggacgctttaaattgac

		atggtgcgaagagtctactgagctaggtagtagtctccggccctgaatgcggctaata cccaaccacggagcaggcgctcacactccagtgagtagcttgctcgtaatgcgaaagtct gtggcggaaccgactactttgggtgtccgtgttctctttatcttatcatggctgctt atggtgacaatctaagattgttatcatatagctattggattggccatccggtgaattct agagatattatcttctgtttgttgggtttaccacattaaacattaccaattacagaac tctcctctacatcctgttattgaataatagaaagagaaa
102	EV91	ttaaaccagctctagggttggttcccaccctagaggccacgtggcggttagcactctggt atcacggtacctttgtgcgcctgttttatatccctttcccatgtaacttagaagttat taaacaagttcaatagaaggggtacaaaccagtgccaccacgaacaagcacttctgt ttcccgggtgaagctgtatagactgtctccacggttgaaaaatggcagatccgttaccg ctttggtagcttcgagaaaacctagatcaccttggaaatcttcgatgcgttgcgctcagca ctcaaccccagagtgtagcttaggtcgatgagctcggacgatcctcactggtgacagt gtccaggctgcgttggcgccctactgtggtgaaaaccacaggagcgtattgtgaaca aggtgtgaagagtctattgagctaccaaagagtctccggccctgaatgcggctaatac ctaaccacggagcaagtgccttcaaaccagttaggtagcttgctcgtaacgcgaagctctg tggcggaaccgactactttgggtgtccgtgttctctttatcttatcatggctgcttat ggtgacaatctaagattgttatcatatagctattggattggccatccggtgactaacag agatctcatatacctttttgttgggtttactaaattaaatattaccacgttcaagacac ttctccttatcatacaactaaatagtgagaagagaaa
103	EVJ108	ttaaaatagcctcagggttggttcccaccctgagggccacgtggcgtagtactctggt ttacggtacctttgtacgcctattttaccccttcccttgtaacttagaagcaaagca aaccagttcaatagtaagcaacacaaccagtgttgtagcaacaagtagcttctgtttc cccgagggtctgacggtgaagctgtaccacggtgaagtagacacacgttaaccg gctacctaacttcgagaagctagtaataaccattgaagtttgttggcggttaacgtc acactaccccggtgttagtttggctgatgagtcacggcatccccacggcgacgctg gccgtggtgcgttgcggccaaccaagggcgcaagctcctggacgtcacttaacaga catggtgtgaagaacctattgagctaggttagtagtctccggccctgaatgcggctaa tcctaactccggagcacatcagtgcacccagcatttgggtgtgttgtaatacgaagtc tggagcggaaccgactactttgggtgtccgtgttctctgtttaccttatttggctgct tatggtgacaatttgatattgttaccatatagctgttggattggccatccggatttttg aaagagacccaaaactttctctacttcagattcaagtgcggaagtttctctttcat atattacttactaatttgaagtaccaaag
104	EV87	ttaaaacagcctgtgggttggttcccaccacagggccactggcgctagcacactggt atcacggtacccttggtgcgcctgttttatacaccctccctactgtgacttagaagca aaacaatgatgaccaatagcaagcgcagcactccagctgtgttccggtcaagcactct gtttcccggactgagtatcaatagggcactcagctggcggaaggagaaaacgtccgt acccggccaattacttcgagaaaaccagtaacaccatagaggttcgcgacgcttccgt ccgcactccccagtgtagatcaggtcgatgagtcacgcgttccccacggcgacgct ggcggtggtgcgttggcgccctgccatggggaacccatgggacgcttcaatactga catggtgcgaagagtctattgagctaattggtagtctccggccctgaatgcggctaa tcctaactgcggagcagataaccacgctccagtgggcagtctgtcgtaacgggtaactc tgcagcggaaccgactactttgggtgtccgtgttctctttatcttatattggctgct tatggtgacaattgaaagattgttaccatatagctattggattggccatccggtgtcta atagagcaattatattctgtttgttgggttttataccactgaatatcacagtttacaag accctcaattacatcctacgtttaaacaagacaaa
105	EchoV31	ttaaaacagcctgtgggttggttcccaccacagggccatggcgctagcacactggt tcgtggtacctttgtgcgcctgttttatacaccctcccatgttaacttagaagtaaa gcaatattgatcaatagcaagcatggcacaccagccgtgttcagatcaagcacttctgt aaccgggactgagtatcaataggctgccaaagcggctgaaggagaaaacgttcggttaa ccggccaactacttcgagaagccagtatcaccatgaaagttgcagagcgtttcgctca gcactcccccggtgtagattaggtcgatgagtcacgcgatccccacaggcgactgtgg cggtggtgcgttggcgccctacctaaggggcaaccgtaggacgctctaatacagaca tggtgagaagagcctattgagctagttggtagtctccggccctgaatgcggataatc ctaactgcggagcatgtgctcgcaatccagcgagaggcatgtcgtaacgggcaactctg cagcggaaccgactactttgggtgtccgtgttctctttatctctgtttggtgctta tggtgacaattgaaagattgttggcatatagctattgggtcgggcacctgggtatctaac agagcaattatcttctgtttgttgggtttgtacccttaaatctcaaggttcttgaaac ccttaactacacattgaaactcaattcattgaa
106	EV68	ttaaaacagctctggggttggttcccactcaagggccacgtggcggttagtactctggt atctcggtacctttgtacgcctgttttaattccctcccaacgtaacttagaagctttt aaaccaagctcaatagggtggagcgcaaacacgcgtcttatgagcaagcacttctgtc

		tccccggtgtggttgtatagactgtccccacggttgaaaacaacttatccgttaaccgc tatagtacttcgagaaacctagattgccttcggagtggtgatgcggttcgctcagcac actaaccggtgtgtagcttgggtcgatgagtcctggacgtacccactggcgacagtggt ccaggctgcggttggcgccctactcatggtgaaaaccatgagacgctagacatgaacaag gtgtgaagagtcctattgagctgctatagagtcctccggccctgaatgcggttaacct aaccatggagcaagtgtcacaaaccagtgagttacttgcgtaacgcgcaagtccgtg gcggaaccgactactttgggtgtccgtgtttcactttttacttttatgactgctaattgg tgacaatttaataattgttaccatttggcttgcgaattgatcacataagatctatagtt ttgttactgatttgcctttgaaataatctcacctcaaacctccagtagacataaacattta aagagtttaaaacttattttataaca
107	EV92	tttaaaacagcctgtgggttgtacccacccacagggccacgcggcgtagtactctgg tatcacggtacctttgtatgcctgtttttatctcccttcccccgtaacgttagaagaatg aacaaaagacaattagcaaatccaccaaccaggtgggttttcgctcaagcacttctgtt tccccggacctagtagcaataagctgtgcccacggtgaaggggaaaaacgttcggttatc cggctacctaacttcgagaagcctagtagcaccaatgattgtctcaagtgttacgctcag cacaaacccagtgtagttcaggtcgatgagtcacccaattccccacgggcgacgctgtc gggtggctgcggttggcgccctgctcgcaggtttcactcgcaggacgcttgattcatgaca tgggtcgaagagcctattgagctaggcagtagtcctccggccctgaatgcggctaatac ctaaccacggagccatgcgtcacaaaccagtgccgtcatggctgtaacgagcaattctg tggcggaaccgactactttgggtgtccgtgtttcctgttttatttacettggetgctta tgggtgacaatctagagttgtcatcatatagcgattggattggccatccggatattactt tgtgtgatttatctcttgcctgtaagcaattttaaatagtttatagcatcatt
108	EVG2	ctaaaacagcctgcgggttgtacccgcgccgacagggccactgggcgctagtagtctggt acgctagtagcactagcagtttagtgccttttgcacgcagtagcctttgtacgcctgtttt attttcccttccccacatgtaactttagaagtttagaaacaaacaatgacaataggag tccagcatccagttggattgcgggtcaagtagtctctgtttccccggacctagtagtgata ggctgtacccacggccgaagatgaaccgcctcgttatccggctacctactcgggaagc ctagttaacattctgaagtctctgaggcggttgcgtcagcagcagcccggtgtagatcgg gctgatgggtctccgcataccccacgggtgaccgtggcggagggccgcttggcgcccg cctatggcgaaaagcctataggacgcctcttagatgacagggtgtgaagagcctactgagc tgggtagtagtctccggccctgaatgcggctaatacctaaccagggagcgtccaccgg caaaccggtcgggagggcgctcgtaacgggttaactctgtggcggaaccgactactttggg tgtccgtgtttccttttgatcctattttggctgcttatggtgacaacgataagttgtta tcataaagcttttgggttggccacctggaaaaagttatcagtggttgatattgttcggc tctcacgccttccaacaaaacaggccttacttcttttaattgcattttatccatcggag gttattacagatttccatctgaagcttcttctcacagtagcagtaaa
109	CVA2	ttaaaacagcctgtgggttgcacccacccacagggccactgggcgctagcactctggt atcacgataacctttgtgcgcctgttttatatccccaccccgagtaaacgttagaagtta cgcaaccccgatcaatagtaggtgtagcactccagctgcatcgagatcaagcacttctg tctccccggaccgagtagtcaatagactgctaaccgcggttgaaggagaaaaacgttcgtta cccgccaattacttcgagaagccagtagtgccgtgaaagtgcggagtggttcgctc agcacttcccccgtagatcaggctgatgagtcacgcgcgacccccacagtgactgtg gcgggtggctgcggttggcgccctgctatggggcaaccataggacgctctaatacagac atggtgcgaagagcctattgagctaattggtagtctccggccctgaatgcggctaata cctaactgcggagcacatgccctcaaaccaggggggtggtgtgctgtaacgggttaactct gcagcgggaaccgactactttgggtgtccgtgtttctttttattcttataatggctgctt atggtgacaattaaagaattgttaccatatagctattggattggccatccggtagactaa caaatcgctcatataaccagtttgttgggtttgttcccttatcacatacagctcataaca ccctcttatatttactacaattgaaatagcaagaa
110	CVA6	ttaaaacagcctgtgggttgcacccactcacagggccactgggcgctagcacactgat ttcccgaatccttgtgcgcctgttttatatccccctccccatgcgcaacttagaagca atctacaccttcgatcaatagcaggcggtggcgcgccagccatgtctagatcaagcactt ctgtttccccggactgagtagtcaataaactgctcacgcggttgaaggagaaaatgttcg ttaccgggctaactacttcgagaaacctagtagcaccatgaaaattgcagagcggttcg ctcagcgcttcccccgctagatcaggctgatgagtcactgcattcctcacggggcgacc gtggcagtggtgctgttggcgccctgccatggggtaacccatgggacgctctaataatg gacatggtgtgaagagtcattgagctagttagtagtcctccggccctgaatgcggct aatcctaactgcggagcacatacccccaaacagggggcggtgtgtcgtaacgggcaac tctgcagcgggaaccgactactttgggtgtccgtgtttccttttattcttatattggctg cttatggtgacaattgaaagattgttaccatatagctattggattggccatccggtagaa

		taacagagccttgatatacctttttgtagggtttataccacttactcttcgcgttggtg agactctaaagtacattctaatcttgaacactagaaa
145	CVB3	UUAAAAACAGCCUGUGGGUUGAUCCACCCACAGGCCCAUUGGGCGCUAGCACUCUGGUA UCACGGUACCUUUGUGCGCCUGUUUUAUACCCCUCCCCAACUGUAACUUAGAAGUAA CACACACCGAUCAACAGUCAGCGUGGCACACCAGCCACGUUUUGAUCAAGCACUUCUGU UACCCCGGACUGAGUAUCAUAGACUGCUCACGCGGUUGAAGGAGAAAGCGUUCGUUAU CCGGCCAACUACUUCGAAAAACCUAGUAACACCGUGGAAGUUGCAGAGUGUUUCGCUCA GCACUACCCAGUGUAGAUACAGGUCGAUGAGUACCGCAUUCCCCACGGGCGACCGUGG CGGUGGUGCGUUGGCGGCCUGCCAUUGGGGAAACCAUGGGACGCUCUAAUACAGACA UGGUGCGAAGAGUCUAUUGAGCUAGUUGGUAGUCCUCCGGCCCCUGAAUGCGGCUAUUC CUAACUGCGGAGCACACCCCUAAGCCAGAGGGCAGUGUGUCGUAACGGGCAACUCUG CAGCGGAACCGACUAUUUGGGUGUCCGUGUUUCAUUUUUAUCCUAUACGGGCGCUUA UGGUGACAAUUGAGAGAUUGUUACCAUAUAGCUAUUGGAUUGGCCAUCCGGUGACUAAU AGAGCUAUUAUAUACCCUUGUUGGGUUUAUACCACUAGCUUGAAAGAGGUUAAAAAC AUUACAAUUCAUUGUUAAGUUGAAUACAGCAAA
146	EV71	UUAAAAACAGCUGUGGGUUGUCACCCACCCACAGGGUCCACUGGGCGCUAGUACACUGGU AUCUCGGUACCUUUGUACGCCUGUUUUAUACCCCUCCCUGAUUUGCAACUUAGAAGCA ACGCAAAACAGAUCAAUAGUAGGUGAGACAUACCAGUCGCAUCUUGAUCAAGCACUUCU GUAUCCCGGACCGAGUAUCAUAGACUGUGCACACGGUUGAAGGAGAAAACGUCCGUU ACCCGGCUAACUACUUCGAGAAGCCUAGUAACGCCAUUGAAGUUGCAGAGUGUUUCGU CAGCACUCCCCCGUGUAGAUACAGUCCGAUGAGUCACCGCAUUCCCCACGGGCGACCGU GGCGGUGGUGCGUUGGCGGCCUGCCUAUGGGGUAACCAUAGGACGCUCUAAUACGGGA CAUGGCGUGAAGAGUCUAUUGAGCUAGUUAUAGUCCUCCGGCCCCUGAAUGCGGCUAU UCCUAACUGCGGAGCACAUACCCUUAUCCAAAGGGCAGUGUGUCGUAACGGGCAACUC UGCAGCGGAACCGACUACUUGGGUGUCCGUGUUUCUUUUUAUUCUUGUAUUGGUGCU UAUGGUGACAAUUAAGAAUUGUUACCAUAUAGCUAUUGGAUUGGCCAUCCAGUGUCAA ACAGAGCUAUUGUAUUCUUCUUGUUGGAUUCACACCUCUCACUCUUGAAACGUAACAC ACCCUCAAUUAUUAUACUGCUGAACACGAAGCG
147	CVB5	UUAAAAACAGCCUGUGGGUUGUCCCCACCCGAGGGGCCACUGGGCGCCAGCACUCUGGU AUCACGGUACCUUUGUGCGCCUGUUUUAUAAACCCUCCCAUUGAAACUUAGAAGC AAUACACCUCGAUCAUAGUAGGCAUGACACGCCAGCCAUUGUCUUGAUCAAGCACUUCU GUUCCCCCGACUGAGUAUCAUAAACUGCUUGCGCGGUCGAAGGAGAAAACGUCCGUU ACCCGACUAACUACUUCGAGAAACCCAGUAACACCAUGGAAUUGCGGAGUGUUACU CAGCACAUUCCAGUGUAGAUACAGGUCGAUGAGUCACCGCAUUCCCCACGGGUGACCGU GGCGGUGGUGCGCUGGCGGCCUGCCAUUGGGGCAACCAUGGGACGCUCUAAUUGGA CAUGGUGUGAAGAGUCUAUUGAGCUAGUUAUAGUCCUCCGGCCCCUGAAUGCGGCUAU UCCUAACUGCGGAGCACGUGCCUCCAUUCCAGGGGGUGGCGUGUCGUAACGGGCAACUC UGCAGCGGAACCGACUACUUGGGUGUCCGUGUUUCUUUUUAUUCUUGGUGCUU AUGGUGACAAUUGAAAGAUUGUUGCCAUUAUAGCUAUUGGAUUGGCCAUCCGGUAUCCAA CAGAGCAAUUGUGUACCUUUUUGUUGGAUUGUACCACUUAACAGAAACAGUUUUAUA CACUGUGUUACAUAUUAUAGACUAAACACAGAAAA
148	EchoV11	UUAAAAACAGCCUGUGGGUUGUACCCACCCACAGGGGCCACUGGGCGCUAGCACACUGGU AUCACGGUACCUUUGUGCGCCUGUUUUAUACCCCUUCCCGCAACCGCAAAUUAGAAG CAAAGCUAACCCGAUCGAUAGCGGAUGCGCAUGCCAGCCGAUUUUGAUCAAGUACUUC UGUUUCCCCGACCGAGUAUCAUAGACUGCUCACGCGGUUGAAGGAGAAAACGUCCGU UACCCGACCAACUACUUCGAGAAACCUAGUAACAUAUGAAUGUUGCAGGGCGUUUCGA UCAGCACGACCCUGGUGUAGAUACAGGUCGAUGAGUACCGCAUUCCCCACGGGUGACCG UGGCGGUGGUGCGUUGGCGGCCUGCCUAUGGGGUGACCCAUAGGACGCUCUAAUACGG ACAUGGUGCGAAGAGUCUAUUGAGCUAGUUGGUAGUCCUCCGGCCCCUGAAUGCGGUUA AUCCUAACUGCGGACGACAUACCCCUAAUCCAAGGGGCAGUGUGUCGUAACGGGCAACU CUGCAGCGGAACCGACUACUUGGGUGUCCGUGUUUCUUUUUAUUCUUGGUGC UUAUGGUGACAAUCUCAGAGUUGUUAACCAUAUAGCUAUUGGAUUGGCCAUCCGGUGAGC AACAGAGCUGUCAUUUAUCAGUUUGUUGGCUUUAUACCUCUAAUACACGGUUUUUUU UUUUUGGAACGCUUGUAUUCUUAACCUUAAGGCAAA
149	EVB107	UUAAAAACAGCCUGUGGGUUGUCCCCACCCGAGGGGCCACUGGGCGCUAGCACACUGGU AUCCCGGUACCUUUGUGCGCCUGUUUUAUUAUACCCUCCCCCUAUGUAACUUAGAAGUA UGAUUCAAAACGGUCGACAGGCGGCUAGUGCACCAACUGAGUCAUGACCAAGCACUUCU GUUACCCCGGACUGAGUAUCAUAAAGCUGUUCACACGGCUGAAGGAGAAAACGUUCGUU ACCCGGCCAAUUAUUCGAGAAACCUAGUACCACCAUGAAGGUUGCGCAGUGUUUCGU CCACACAACCCAGUGUAGAUACAGGUCGAUGAGUACCGCAUUCCCCACGGGCGACCGU GGCGGUGGUGCGUUGGCGGCCUGCCAUUGGGGCAACCAUGGGACGCUCUAAUACUGA

		CAUGGUGUGAAGAGUCUAUUGAGCUAAUUGGUAGUCCUCCGGCCCCUGAAUGCGGCUAA UCCCAACUGUGGAGCAGAUACUCACAAACCAGUGAGCGGUCUGUCGUAACGGGCAACUC CGCAGCGGAACCGACUACUUUGGGUGUCCGUGUUUCUUUUUAUUCUUACAUUGGCUGCU UAUGGUGACAAUUGACAAAUUGUUAACCAUAUAGCUAUUGGAUUGGCCAUCCGGUGACAA ACAGAGCUAUUGUUUACUUGUUUGUUGGUUUCAUACCAUAAAAUACAAGGUCUUAAGAA ACUCUCAACUUUAUUUUGACACUCAAUACAGCAAA
150	PV3	UUAAAAACAGCUCUGGGGUUGCUCCACCCACAGGCCCCACGUGGCGGCCAGUACUCUGG UAUUGCGGUACCUUUGUACGCCUGUUUAUACUCCCUCCCCCGUGCAACUUAGAAGA AAUCCACAAAGUUCAAUAGAGGGGGUGCAAACCAGCACCACACGAACAAGCACUUCUG UUUCCCCGGUGAUGUCGUAUAAGCUGUACCCACGGCUGAAGGCGACGGAUCCGUUAUCC GCUUGAGUACUUCGAGAAGCCUAGUAUUAACCUUGGAAUCUUCGACGCGUUGCGCUCAGC ACUCUACCCCGAGUGUAGCUUAGGUCGAUGAGUCUGGGCACGCCCCACCGCGACGGUG GCCCAGGCUGCGUUGGCGGCCUACCCAUGGCUAUCACCAUGGGACGCUAGUUGUGAACA AGGUGUGAAGAGCCUAUUGAGCUACCCAAGAGUCCUCCGGCCCCUGAAUGCGGCUAAUC CUAACACGGAGCAAGUGUCCUCAACCCAGGGGAUGGCUUGUCGUAACGCGAAAGUCUG UGGCGGAACCGACUACUUUGGGUGUCCGUGUUUCCUUUUUAUUUUUAUGUAUGGCUGCUU AUGGUGACAAUCAAGGUUGUUAACCAUAAAGCAAUUUGGAUUGGCCAUCCGGUGAGAAU CAAACAUAUUAUCUACCUUUUGUUGGGUUUUUUUUUUUAUUUAUUGAACAAUACCUCUAA UAAUAACUGCUAUUAUGUCAUAAGACAUAUUAUCAUCACA
151	CVB1	UUAAAAACAGCCUGUGGGGUUGUCCACCCACAGGCCCCAUUGGGCGCUAGCACUCUGGUA UCACGGUACCUUUGUGCGCCUGUUUAUACUCCCUCCCCAAUUGUAUUUAGAAGUUU CACACACCGAUCAUUAAGCAAGCGUGGCACACCAGCCAUGUUUUGAUCAAGCACUUCUGU UACCCCGGACUGAGUAUCAAUAGACCGCUAACCGCGGUAGAAGGAGAAAACGUUCGUUAC CCGGCCAACUACUUCGAAAAACCUAGUAACACCAUGGAAGUUGCGGAGUUUUCGCUCA GCACUACCCAGUGUAGAUCAGGUCGAUGAGUCACCGCGUUCGCCACGGGCGACCGUGG CGGUGGCUGCGUUGGCGGCCUGCCUACGGGGAAACCCGUAAGGACGCUCUAAUACAGACA UGGUGCGAAGAGUCUAUUGAGCUAGUUGGUAUCCUCCGGCCCCUGAAUGCGGCUAAUC CUAACUGCGGAGCAUAUACCCUCAAACAGGGGGCAGUGUGUCGUAACGGGCAACUCUG CAGCGGAACCGACUACUUUGGGUGUCCGUGUUUUAUUUUUAUUCUUAUACUGGCUGCUUA UGGUGACAAUUGACAGGUUGUUACCAUAUAGUUUAUUGGAUUGGCCAUCCGGUGACUAAAC AGAGCAAUUAUAUACUCUUUGUUGGGUUUAUACCAUUAAGCUUGAAAGAGGUUAAAAAC ACUACAUCUCAUCAUUAUAAACUAAUAACAACAAA
152	EchoVE1 5	UUAAAAACAGCCUGUGGGGUUGUACCCACCCACAGGCCCCACUGGGCGCUAGCACUCUGGU AUUACGGUACCCUUGUGCGCCUGUUUUUAUUAUACCCCGCCCCAAGCAAACGCUAGAUGUA ACGCACUUAUGAUCAAUAGCAGGCGUGGCACUCCAGCCACGUUAUGAUCAAGCACUUCU GUCUCCCGGACCGAGUAUCAAUAGACUGCUCACGCGGUUGAAGGAGAAAACGUUCGUU ACCCGACCAGCUACUUCGAGAAAACCUAGUAACUCCAUGGAGGUUGCAGAGUGUUUCGUU CAGCACUUCCCCCGUGUAGAUCAGGCUGAUGAGUCACCGCGUUCUCCACGGGCGACCGU GGCGGUGGCUGCGUUGGCGGCCUGCCUUAUGGGUAUACCCAUAGGACGCUCUAAUACUGA CAUGGUGCGAAGAGUCUAUUGAGCUAGUUGGUAGUCCUCCGGCCCCUGAAUGCGGCUAA UCCUAAUCUGCGGAGCAUAUACCCCAACCCAGGGGGCAGUGUGUCGUAACGGGUAAUC UGCAGCGGAACCGACUACUUUGGGUGUCCGUGUUUCCUUUUUAUUCUUAUUGGCUGCU UAUGGUGACAAUUGAGAGAUUGUUACCAUAUAGCUAUUGGAUUGGCCAUCCAGUGACUA ACAGAGCUAUUAUUUACCUUUUGUUGGCUUCGUAUCACUUGGUUUAAAAGAGGUUAGU ACUUUAUAUUGCAUUAUAUUAUACUAAACACGAGAAA
153	EchoV1	UUAAAAACAGCCUGUGGGGUUGUCCACCCACAGGCCCCACUGGGCGUUAAGCACACUGGU AUCACGGUACCUUUGUGCGCCUGUUUUUAUACUCCCCCCCUAAGGAAACUUUAGAAGCA AAGCAAUUGUGAUCAAUAGUGGGUAUGGCACACCAAGUCAUAUUCUUGAUCAAGCACUUCU GUUCCCCCGGACUUAAGUACCAAUAGACUGCUCAAAGCGGUUGAAGGGGAAAACGUUCGUU AUCCGGCCAAACUACUUCGAGAAAACCUAGUAGCACCAUGAAAGUUGCGGAGUGUUUCGU CAGCACUUCCCCCGUGUAGAUCAGGCUGAUGAGUCACCGUAUUCCCACGGGCGACCGU GACGGUGGCUGCGUUGGCGGCCUGCCUUAUGGGGUAACCCAUGGGACGCUCUAAAAACAGA CACGGUGCGAAGAGUCUAUUGAGCUAGUUGGUAGUCCUCCGGCCCCUGAAUGCGGCCAA UCCUAAUCUGCGGAGCAUAUACUCCCAAUCCAGGGAGCAGUGUGUCGUAUUGGGUAAUC UGCAGCGGAACCGACUACUUUGGGUGUCCGUGUUUCCUUUUUAUUCUACAUAUGACUGCU UAUGGUGACAAUUGAAAGAUUGUUACCAUAUAGCUAUUGGAUUGGUCAUCCGGUGAGCA AUAGAGCUAUUGUUUAUCAAUUUGUUGGAUUUGUACCACUACUUUUCUGUUUUGAGA ACACUCAACUACAUCUUAUCUGCUAAACACAUCAAA
154	EVD94	UUAAAAACAGCCUGUGGGGUUGUCCACCCACAGGCCCCACGUGGCGGCCAGUACUCUGG UAUCACGGUACCUUUGUACGCCUGUUUUUAUUAUACCCCUCCCCGCAACUUAGAAGAAAA CAAUACAAGUUCACUAGGAGGGGGUACAAACCAGUACCACCACGAACAAGCACUUCUGU

[illegible]

		acagagcaauaauaauuuguuuuguggcuuuuggccacuuauuugaaggguuuaag acgcucaauucugauacuucuaauaaaagcaaa
159	EV89	uuaaaaacagccucugggugcaccacccacagggcccacugggcgcuagcacacuggu aucacggguaccuuugugcgccuguuuuauacacccucccccuuguaacuuagaagaaa caaaacaaacgaucaauaguaggcuuagcacuccagcuaagucugaucaagcacuucug uuacccccggaccgaguaucuuaggcuguaaccacggcguaagagaaaaacguucguua cccgguauucuaucuuccgggaaccacaguaaccuauaggaauugcuaagcguuucgcu agcacacccccaguguaagaucaggucgaugagucacugcauuccccacaggcgacugug gcagugggcugcguuugcgccugccuauuggggaaacccauaggacgcuuauauugac augguugcgaagagucuaucugagcuagguaguaguccuccggccccugaaugcgguauu cccaaccacggagcaggcgucacacuccagugagugcuugucguauugcgaaaguc guggcggaaccgacuacuuugggugccguguuuccuuuuauuuuaucauuggcugcuu auggugacaauuagaaguuuaucauauagcuauuggaauugcaccuccgguguauc agagauuuuuuuacuuuguuuugguuuuacacauuaaacauuaccaaauacagaac ucuccucuacauccuguuauugaauaauagaaagagaaa
160	EV91	uuaaacagcucuaagggugucccaccuagaggcccacguggcggcuagcacucuggu aucacggguaccuuugugcgccuguuuuauucccuuccccauguaacuuagaaguuau uaaacaagaugcaauagaaggggguacaaaccagugccaccacgaacgaacuuucugu uuccccggguagcugugauagacugucuccacgguguaaaaugcgagucuuuacccg cuuugguacuucgagaaaccuaguaucaccuuggaauucucgaugcgugcgucagca cucaaccccagaguguaagcuuaggucgaugagucuggacgauccucacuggugacagug guccaaggcugcguuugcgccuaccuguggugaaaaccacaggacgcuauguuguaaca aggugugaagagucuaauagagcuaccaaagaguccuccggccccugaaugcgguauuc cuacccacggagcaagugccucaaaccaguaagguagcuugucguaacgcgcaagucug uggcggaaccgacuacuuuggguguccguguuuccuuuuauuuuaucauugcgcuuau ggugacaauucuaagauuguuaucauauagcuauuggaauuggccaucgggugacuaacag agaucucuauauaccuuuuuuguuuguuuuacuaauuuuuuuuauuacacgcuuacagacac uucuccuuaucauacaacuaauuagugagaagagaaa
161	EVJ108	uuaaaaauagccucagggguugucccaccucagaggcccacguggcgcuagcacucuggu uaccgguaccuuuuguaacgcguuuuuuacccccuuccccuuguaacuuagaagcaaaagca aaccaugucaauagaaagcaacacacacccagugugagcaaacagaucuuucuguuuc ccccgggagggucugacgguaagcuguaaccacggcguaagauagcacuaccguuaaccg gcuaccuacuucgagaagucuaauuaccuauugaaguuuugugggcuuacgcucacac acacuacccccgugugaguuuuggcugaugagucacggcauuccccacgggcgacccgug gccgugggcugcguuugcgcccaaccaaggggcgcaagcuccuuggacgucacuuuacaga cauggugugaagaaaccuauugagcuagguaguaguccuccggccccuugaauugcgguua uccuaacuccggagcaacauagugcacaccagcauauugguuguguguaauacgcaaguc uggagcggaaccgacuacuuuggguguccguguuuccuguuuuacccuauuuuggcugcu uauuggugacaauuugauauuguuaccuauuagcuguuuggaauuggccaucgggaauuuug aaagagaccccaaaacuuucuuucucuacuuacagauuacagugcgaaguuuuccuuuuc auauuacuuacuaauuugaaguaccaaag
162	EV87	uuaaaaacagccugugggugucccaccacagggcccacugggcgcuagcacacuggu aucacggguaccuuugugcgccuguuuuauacacccuuccccuacugugacuuaagaagca aaacaaugaugaccaauagcaagcgagcacuccagcuguguuuccggucaagcacuuc guuuccccggacugaguaucuuaggccacucacgugggccgaaggagaaaacgucgcu accgggccaauuacuuugagaaaaccaguaacaccuauagagguugcgagcgguuucgcu ccgcacucccccaguguaagaucaggucgaugagucacccgcuuuccccacgggcgacccgu ggcgugggcugcguuugcgccugcccaugggggaacccauaggacgcuucaauacuga caugugcggaagagucuaauagagcuauuugguagucuccuccggccccuugaauugcgcu uccuaacugcgagcagauacccacgcuccagugggcagucugucguaacggguaacuc ugcagcggaaccgacuacuuuggguguccguguuuccuuuuauuucuuauuuggcugcu uauuggugacaauugaaagauuguuaccuauuagcuauuggaauuggccaucgggugucua auagagcaauuuauuuccuguuuuguguuuuuauaccacugaaauacaguuuacaaag accucaaauacauccuacguuuuaacaaagacaaa
163	EchoV31	uuaaaaacagccugugggugucccaccacagggcccacugggcgcuagcacacuggu ucgugguaccuuuugugcgccuguuuuauacacccuuccccauuuguaacuuagaaguaaa gcaauauugaucaauagcaagcauggcacaccagccgugucagaucaagcacuucugu aaccgccggacugaguaucuuaggcugccaaagcggcugaaggagaaaacguucguuaa ccggccaacuacuucgagaagcccaguaucaccaugaaguuugcagagcgguuucgcu gcacucccccuguguaaguuaggucgaugagucacggcaaguccccacagggcgacuggg cggugcgucguuugcgcccuacccuacqgggcaaccggaagcgcucuaauuacagaca c

		uggugagaagagccuauugagcuaugguaguccuccggccccugaaugcggaauaauccuaacugcgagcaugugcgcgaauccagcgagaggaugugcguaacgggcaacucugcagcggaaccgacuacuuuggguguccguguuuccuuuuauuccuuguuuggcugcuuauggugaccaauugaaagauuguugccauauagcuaauugggucgggccaccugguauccuaacagagcaauuauuacuuguuuuguuuguuuuguaacccuuaaaucucaagguuccuugaaaccuuaacucacacauugaaacucaauucauugaa
164	EV68	uuaaaaacagcucugggguuguucccacucaaggggccacgugggcgguagucucugguaucucgguaccuuuguagcccguguuuuauuccucccccaacguaacuuagaagcuuuu aaaccaaagcucaauagguggagcgcaaacacagcgucuaugagcaagcacuucuguc uccccggugugguuauagacuguccccacgguugaaaacaacuuauccguuaaccgc uauaguacuucgagaaaccuaguauugccuucggagugugaugcgugcgucagcac acuaacccgugugagcuugggucgaugagucuggacguacccacugggcgacaguggu ccaggcugcgguuggcgccuacucaugggugaaaaccaugagacguagacaugaaccaag gugugaagagucuaauagagcugcuauagaguccuccggccccugaaugcggucaauccu aaccauggagcaagugcucacaaaaccagugaguuacuugucguaacgcgcgaugccgug gcggaaccgacuacuuuggguguccguguuucacuuuuuacuuuuuagacugcuauugg ugacaauuuauuuguuaccuuuugcgugcgaauugaucacauaagaucuauaguu uuguuacacugauuugcuuugaaaauaauccucaccucaaaaaccuccaguacauaacuuua aagaguuaaaacuuuuuauaaca
165	EV92	uuuuuuacagccugugggguuguacccacccacagggccccacggcgguuaguacucuggu uaucacgguaccuuuguauugccuguuuuauccuucccccguaacguuagaagaauu aacaaaagacaaauagcaauuccaccaaccacagggugguuuucgucaagcacuucuguu uccccggaccuaguaccaauaagcugugccccacggcugaaggggaaaaacguucguuau cggcuaccuacucgagaagccuaguaaccaccaugauugucucaaguguuacgcucag cacaaccccaguguauguacagguugcgaugagucacccgaauuccccacggcgaccgugc gguggcugcgguuggcgcccgucgaguuuacucgcaggacgcugauuaucaugaca uggugcgaagagccuauugagcuaggcaguguccuccggccccugaaugcggucaau cuaaccacggagccauugcucacaaaaccagugggcugcaugugcguaacgagcaauucug uggcggaaccgacuacuuuggguguccguguuuccuguuuuuuuacuuugcgugcuua uggugacaauucagagugugucacauauagcgaauuggauuggccauccggauauuacuu ugugugauuuauucucugcugcuaggaauuuuuauuaguuuauagcauau
166	EVG2	cuaaaacagccugcgguuguacccgcgccagggccccacugggcgcuagucucuggu acgcuaaguacacuaagcacguuagugcuuuuguacggccaguaacuuuguacggccguuuu auuuucccuucccccauguaacuuuagaaguuuagaaacaacaauagaccauaggag uccagcauccaguuggauugcggucaaguacuuucuguuuccccggaccuaguagugaua ggcuuacccacggccgaagaugaaccgcugcuuauccggcuaccuacuuccgggaagc cuaguaaacauucugaaagucucugaggcguuucgucacagcacgcccgguuagucgg gcugauugggucuccgcuaacccacgggugaccgugggcgagggcgcuuggcgccccg ccuauuggcgaaagccauaggacgcucuuagauagacagggugugaagagccuacugagc uggguaguaguccuccggccccugaaugcggucaauccuaaccacggagcguccaccgg caaaccgguccgggagggcgucguaacggguaacucugugggcggaaccgacuacuuuggg uguccgguuuuccuuuugauccuauuuuggcugcuuauugugacaacgaaaguuuuu ucauaaagcuuuuuggguuggccaccuggaaaaaaguuauacaguguuuugauuugucggc ucucacgcccuccaacaacacagggccuuacuucuuuuuauuugcauuuuauccaucggag guuauuacagauuuuccaucugaagcuucucucacaguagcaguaaa
167	CVA2	uuaaaaacagccugugggugcaccacccacagggccccacugggcgcuagcacucuggu aucacgauaccuuugugcgccuguuuuauaucccccaccccgaguaaacguuagaaguua cgcaacccccgaucaaauaguagguguaagcacuccagcugcaucgagaucaagcacuucug ucuccccggaccgaguaucuaauagacugcuaacgcggguugaaggagaaaaacguucguua cccggccaauuacuuucgagaagccagugaguccgugaaaguuugcgaguguuucgcuc agcacuucccccgugauagacagggcugaugagucacccgcgaucacccacaggugacugug gcggugggcugcgguuggcgccugccuauuggggcaaccacauaggacgcucuaauacagac auggugcgaagagccuauugagcuauuugguaguccuccggccccugaaugcggucaau ccuaacugcgagcacauugcccuaaaaccagggggugugugcguaacggguaacucuc gcagcggaaccgacuacuuuggguguccguguuucuuuuuuuauccuauauauggcugcuu auggugacaauuaaagaauuguuaccuauagcuauuggauuggccauccggugacuuaa caaauccgcucauauaccaguuuuguuuuuuucccuaucauacagcucuaaaca ccucuuauuuuacuacaauugaaugcaagaa
168	CVA6	uuaaaaacagccugugggugcaccacucacagggccccacugggcgcuagcacacugau uucccggaauccuugugcgccguuuuuauaucccccucccccaugcgcaacuuagaagca aucuacaccuucgaucuuuagcagggcgugggcgccagccauugcuagaucaagcacuu

		cuguuuccccggacugagaucaauaaacugcucacgcggguagaaggagaaauguuucg uuaccccggaacuaucucgagaaaccuaguagcaccuagaaaauugcagagcgguuucg cucagcgcuucccccgcuagaucaggcugaugagucacugcauuccucacgggcgacc guggcaguggcugcguuggcgccugcccaugggguaacccaugggacgcucuaauaug gacauggugugaagagucuaugagcuaguuauguaguccuccggccccugaaugcggu aauccuaacugcggagcacaauacccccaaaccagggggcgguugugucguaacgggcaac ucugcagcgggaaccgacuacuuuggguguccguguuuuccuuuuauucuuauauuggcug cuuauggugacaaugaaagauuguuaccuauauagcuauugggaauggccauccggugaa uaacagagccuugauauaccuuuuuguaggguuuauaccacuuacucucgcguuguug agacucuaaaguacauucuaaucuugaacacuaagaaa
--	--	--

[0224] In some aspects, inclusion of an IRES permits the translation of one or more open reading frames from a circular RNA (e.g., open reading frames comprised in the exonic sequences). The IRES element attracts a eukaryotic ribosomal translation initiation complex and promotes translation initiation. See, e.g., Kaufman et al., *Nuc. Acids Res.* (1991) 19:4485-4490; Gurtu et al., *Biochem. Biophys. Res. Comm.* (1996) 229:295-298; Rees et al., *BioTechniques* (1996) 20: 102-110; Kobayashi et al., *BioTechniques* (1996) 21:399-402; and Mosser et al., *BioTechniques* 199722 150-161). A multitude of IRES sequences are available and include sequences derived from a wide variety of viruses, such as from leader sequences of picornaviruses such as the encephalomyocarditis virus (EMCV) UTR (Jang et al., *J. Virol.* (1989) 63: 1651-1660), the polio leader sequence, the hepatitis A virus leader, the hepatitis C virus IRES, human rhinovirus type 2 IRES (Dobrikova et al., *Proc. Natl. Acad. Sci.* (2003) 100(25): 15125- 15130), an IRES element from the foot and mouth disease virus (Ramesh et al., *Nucl. Acid Res.* (1996) 24:2697-2700), a giardiavirus IRES (Garlapati et al., *J. Biol. Chem.* (2004) 279(5):3389-3397), and the like.

[0225] In some aspects, the IRES is an IRES sequence of Taura syndrome virus, Triatoma virus, Theiler's encephalomyelitis virus, Simian Virus 40, Solenopsis invicta virus 1, Rhopalosiphum padi virus, Reticuloendotheliosis virus, Human poliovirus 1, Plautia stall intestine virus, Kashmir bee virus, Human rhinovirus 2, Homalodisca coagulata virus- 1, Human Immunodeficiency Virus type 1, , Himetobi P virus, Hepatitis C virus, Hepatitis A virus, Hepatitis GB virus , Foot and mouth disease virus, Human enterovirus 71, Equine rhinitis virus, Ectropis obliqua picorna-like virus, Encephalomyocarditis virus, Drosophila C Virus, Human coxsackievirus B3, Crucifer tobamovirus, Cricket paralysis virus, Bovine viral diarrhea virus 1, Black Queen Cell Virus, Aphid lethal paralysis virus, Avian encephalomyelitis virus, Acute bee paralysis virus, Hibiscus chlorotic ringspot virus, Classical swine fever virus, Human FGF2, Human SFTPA1, Human AML1/RUNX1, Drosophila antennapedia, Human AQP4, Human AT1R, Human BAG-1, Human BCL2, Human BiP, Human c-IAP1, Human c-myc, Human eIF4G, Mouse NDST4L, Human LEF1, Mouse HIF1 alpha, Human n.myc, Mouse Gtx, Human p27kipl, Human PDGF2/c-sis, Human p53, Human Pim-1, Mouse Rbm3, Drosophila reaper, Canine

Scamper, Drosophila Ubx, Human UNR, Mouse UtrA, Human VEGF-A, Human XIAP, Drosophila hairless, S. cerevisiae TFIID, S. cerevisiae YAP1, tobacco etch virus, turnip crinkle virus, EMCV-A, EMCV-B, EMCV-Bf, EMCV-Cf, EMCV pEC9, Picobirnavirus, HCV QC64, Human Cosavirus E/D, Human Cosavirus F, Human Cosavirus JMY, Rhinovirus NAT001, HRV14, HRV89, HRVC-02, HRV-A21, Salivirus A SHI, Salivirus FHB, Salivirus NG-J1, Human Parechovirus 1, Crohivirus B, Yc-3, Rosavirus M-7, Shanbavirus A, Pasivirus A, Pasivirus A2, Echovirus E14, Human Parechovirus 5, Aichi Virus, Hepatitis A Virus HA16, Phopivirus, CVA10, Enterovirus C, Enterovirus D, Enterovirus J, Human Pegivirus 2, GBV-C GT110, GBV-C K1737, GBV-C Iowa, Pegivirus A 1220, Pasivirus A 3, Sapelovirus, Rosavirus B, Bakunsa Virus, Tremovirus A, Swine Pasivirus 1, PLV-CHN, Pasivirus A, Sicinivirus, Hepacivirus K, Hepacivirus A, BVDV1, Border Disease Virus, BVDV2, CSFV-PK15C, SF573 Dicistrovirus, Hubei Picorna-like Virus, CRPV, Salivirus A BN5, Salivirus A BN2, Salivirus A 02394, Salivirus A GUT, Salivirus A CH, Salivirus A SZ1, Salivirus FHB, CVB3, CVB1, Echovirus 7, CVB5, EVA71, CVA3, CVA12, EV24 or an aptamer to eIF4G.

II.B.8. UTRs and Kozak Sequence

[0226] In some aspects, genetic constructs disclosed herein further comprises a UTR. For example, in some aspects, a pre-circRNA or circRNA produced as disclosed herein comprises a payload sequence, wherein the payload sequence comprises a UTR. In some aspects, the UTR is a 5'-UTR. In some aspects, the UTR is a 3'-UTR. In some aspects, the UTR comprises both a 5'-UTR and a 3'-UTR.

[0227] Untranslated regions (UTRs) of a gene are transcribed but not translated. The 5'-UTR starts at the transcription start site and continues to the start codon but does not include the start codon; whereas, the 3'-UTR starts immediately following the stop codon and continues until the transcriptional termination signal. There is growing body of evidence about the regulatory roles played by the UTRs in terms of stability of the nucleic acid molecule and translation. Accordingly, where a pre-circRNA or circRNA produced as disclosed herein comprises a UTR, the stability of the payload sequence is increased, *e.g.*, as compared to a sequence without the UTR. As described herein, in some aspects, increased stability results in increased expression of the encoded protein. Natural 5'-UTRs bear features which play roles in translation initiation. They harbor signatures like Kozak sequences which are commonly known to be involved in the process by which the ribosome initiates translation of many genes. Kozak sequences have the consensus CCR(A/G)CCAUGG, where R is a purine (adenine or guanine) three bases upstream of the start

codon (AUG), which is followed by another 'G'. 5'-UTR also have been known to form secondary structures which are involved in elongation factor binding.

[0228] 5'-UTR secondary structures involved in elongation factor binding can interact with other RNA binding molecules in the 5'-UTR or 3'-UTR to regulate gene expression. For example, the elongation factor EIF4A2 binding to a secondarily structured element in the 5'-UTR is necessary for microRNA mediated repression (Meijer H A et al., Science, 2013, 340, 82-85, herein incorporated by reference in its entirety). The different secondary structures in the 5'-UTR can be incorporated into the flanking region to either stabilize or selectively destabilize mRNAs in specific tissues or cells.

[0229] By engineering the features typically found in abundantly expressed genes of specific target organs, one can enhance the stability and protein production of a nucleic acid sequence (*e.g.*, payload sequence of a synthetic circuit provided herein). For example, introduction of 5'-UTR of liver-expressed mRNA, such as albumin, serum amyloid A, Apolipoprotein A/B/E, transferrin, alpha fetoprotein, erythropoietin, or Factor VIII, could be used to enhance expression of a nucleic acid molecule, such as a mRNA, in hepatic cell lines or liver. Likewise, use of 5'-UTR from other tissue-specific mRNA to improve expression in that tissue is possible—for muscle (MyoD, Myosin, Myoglobin, Myogenin, Herculin), for endothelial cells (Tie-1, CD36), for myeloid cells (C/EBP, AML1, G-CSF, GM-CSF, CD11b, MSR, Fr-1, i-NOS), for leukocytes (CD45, CD18), for adipose tissue (CD36, GLUT4, ACRP30, adiponectin) and for lung epithelial cells (SP-A/B/C/D).

[0230] Other non-UTR sequences can also be incorporated into the UTRs (*e.g.*, 5'-UTR and/or 3'-UTR). For example, introns or portions of introns sequences can be incorporated into the flanking regions of a nucleic acid sequence (*e.g.*, payload sequence of a synthetic circuit provided herein).

[0231] In some aspects, one or more nucleotides within a UTR (*e.g.*, 5'-UTR and/or 3'-UTR) can be mutated, replaced and/or removed. For example, one or more nucleotides upstream of the start codon can be replaced with another nucleotide. The nucleotide or nucleotides to be replaced can be 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 30, 35, 40, 45, 50, 55, 60 or more than 60 nucleotides upstream of the start codon. As another example, one or more nucleotides upstream of the start codon can be removed from the UTR. In some aspects, the genetic constructs disclosed herein comprise a Kozak sequence.

[0232] 3'-UTRs are known to have stretches of adenosines and uridines embedded in them. These AU rich signatures are particularly prevalent in genes with high rates of turnover. Based on

their sequence features and functional properties, the AU rich elements (AREs) can be separated into three classes (Chen et al, 1995): Class I AREs contain several dispersed copies of an AUUUA motif within U-rich regions. C-Myc and MyoD contain class I AREs. Class II AREs possess two or more overlapping UUAUUUA(U/A)(U/A) nonamers. Molecules containing this type of AREs include GM-CSF and TNF- α . Class III AREs are less well defined. These U rich regions do not contain an AUUUA motif. c-Jun and Myogenin are two well-studied examples of this class. Most proteins binding to the AREs are known to destabilize the messenger, whereas members of the ELAV family, most notably HuR, have been documented to increase the stability of mRNA. HuR binds to AREs of all the three classes. Engineering the HuR specific binding sites into the 3'-UTR of nucleic acid molecules can lead to HuR binding and thus, stabilization of the message *in vivo*.

[0233] In some aspects, introduction, removal, or modification of 3'-UTR AU rich elements (AREs) can be used to modulate the stability of a nucleic acid sequence. When engineering specific genetic constructs disclosed herein, one or more copies of an ARE can be introduced to make the nucleic acid sequence less stable and thereby curtail translation and decrease production of the resultant protein. Likewise, AREs can be identified and removed or mutated to increase the intracellular stability and thus increase translation and production of the resultant protein.

[0234] In some aspects, the genetic constructs disclosed herein comprise a 3' UTR. In some embodiments, the 3' UTR is from human beta globin, human alpha globin *Xenopus* beta globin, *Xenopus* alpha globin, human prolactin, human GAP-43, human eEF1 α , human Tau, human TNF α , dengue virus, hantavirus small mRNA, bunyavirus small mRNA, turnip yellow mosaic virus, hepatitis C virus, rubella virus, tobacco mosaic virus, human IL-8, human actin, human GAPDH, human tubulin, hibiscus chlorotic ringspot virus, woodchuck hepatitis virus post translationally regulated element, sindbis virus, turnip crinkle virus, tobacco etch virus, or Venezuelan equine encephalitis virus.

[0235] In some aspects, the genetic constructs disclosed herein comprise a 5' UTR. In some embodiments, the 5' UTR is from human beta globin, *Xenopus laevis* beta globin, human alpha globin, *Xenopus laevis* alpha globin, rubella virus, tobacco mosaic virus, mouse Gtx, dengue virus, heat shock protein 70kDa protein 1 A, tobacco alcohol dehydrogenase, tobacco etch virus, turnip crinkle virus, or the adenovirus tripartite leader.

II.B.9. Spacers

[0236] In some aspects, the genetic constructs disclosed herein comprise a spacer sequence. In some aspects, the genetic constructs disclosed herein comprise a spacer sequence having at least 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or 100% identity with a nucleotide sequence of SEQ ID NOs: 74-75 or 78-79. In some aspects, the genetic constructs disclosed herein comprise a spacer consisting of a nucleotide sequence of SEQ ID NOs: 74-75 or 78-79.

[0237] Non-limiting examples of spacer sequences that are useful for the present genetic constructs are shown in **Table 9**.

Table 9. Exemplary spacers

SEQ ID NO	DESCRIPTION	SEQUENCE
74	5' spacer (all) DNA	AAAATCCGTTGACCTTAAACGGTCGTGTGGGTTCAGTCCCTCCACCC CCACGCCGGAACGCAATAGCCG
75	3' spacer (all) DNA	GGCTATTATGCGTTACCGCGAGACGCTACGGA
78	5' spacer (all) RNA	AAAUCCGUUGACCUUAAACGGUCGUGUGGGUUAAGUCCCUCCACCC CCACGCCGGAACGCAAUAGCCG
79	3' spacer (all) RNA	GGCUAUUAUGCGUUACCGCGAGACGCUACGGA

[0238] In some aspects, the genetic constructs disclosed herein comprise a first (5') and/or a second (3') spacer. In some aspects, including a spacer between the 3' intron fragment and the IRES may conserve secondary structures in those regions by preventing them from interacting, thus increasing splicing efficiency. In some aspects, the first (between 3' group fragment and IRES) and second (between the expression sequence and 5' intron fragment) spacers comprise additional base pairing regions that are predicted to base pair with each other and not to the first and second duplex forming regions. In some aspects, such spacer base pairing brings the intron fragments in close proximity to each other, further increasing splicing efficiency. In some aspects, the combination of base pairing between the first and second duplex forming regions, and separately, base pairing between the first and second spacers, promotes the formation of a splicing bubble containing the intron fragments flanked by adjacent regions of base pairing. In some aspects, a spacer is a contiguous sequence with one or more of the following qualities: 1) predicted to avoid interfering with proximal structures, for example, the IRES, expression sequence, or intron; 2) is at least 7 nt long and no longer than 100 nt; 3) is located after and adjacent to the 3' intron fragment and/or before and adjacent to the 5' intron fragment; and 4) contains one or more of the following: a) an unstructured region at least 5 nt long, b) a region of base pairing at least 5 nt long to a distal sequence, including another spacer, and c) a structured region at least 7 nt long limited in scope to

the sequence of the spacer. Spacers may have several regions, including an unstructured region, a base pairing region, a hairpin/structured region, and combinations thereof. In some aspects, the spacer has a structured region with high GC content. In some aspects, a region within a spacer base pairs with another region within the same spacer. In some aspects, a region within a spacer base pairs with a region within another spacer. In some aspects, a spacer comprises one or more hairpin structures. In some aspects, a spacer comprises one or more hairpin structures with a stem of 4 to 12 nucleotides and a loop of 2 to 10 nucleotides. In some aspects, there is an additional spacer between the 3' intron fragment and the IRES. In some aspects, this additional spacer prevents the structured regions of the IRES from interfering with the folding of the 3' intron fragment or reduces the extent to which this occurs. In some aspects, the 5' spacer sequence is at least 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 25 or 30 nucleotides in length. In some aspects, the 5' spacer sequence is no more than 100, 90, 80, 70, 60, 50, 45, 40, 35 or 30 nucleotides in length. In some aspects the 5' spacer sequence is between 5 and 50, 10 and 50, 20 and 50, 20 and 40, and/or 25 and 35 nucleotides in length. In some aspects, the 5' spacer sequence is 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49 or 50 nucleotides in length. In some aspects, the 5' spacer sequence is a polyA sequence. In some aspects, the 5' spacer sequence is a poly AC sequence. In some aspects, a spacer comprises about 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, or 100% poly AC content. In some aspects, a spacer comprises about 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, or 100% polypyrimidine (C/T or C/U) content.

II.B.10. Poly(A)

[0239] In some aspects, the genetic constructs disclosed herein further comprise a polyA region. In some aspects, the polyA region is at least 15, 30, or 60 nucleotides long. In some aspects, the polyA sequence is removed upon circularization. Thus, an oligonucleotide hybridizing with the polyA sequence, such as a deoxythymine oligonucleotide (oligo(dT)) conjugated to a solid surface (e.g., a resin), can be used to separate circular RNA from its precursor RNA. Other sequences can also be disposed 5' to the 3' intron fragment or 3' to the 5' intron fragment and a complementary sequence can similarly be used for circular RNA purification. In some aspects, the polyA region comprises a nucleotide sequence having at least 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or 100% identity with SEQ ID NOs: 85 or 86. In some aspects, the polyA region consists of SEQ ID NOs: 85 or 86.

[0240] Non-limiting examples of polyA region that are useful for the present genetic constructs are shown in **Table 10**.

[0241] **Table 10.** Exemplary polyA regions

SEQ ID NO	DESCRIPTION	SEQUENCE
85	5' PolyAC	AAAAACAAAAACAAAAAAACAAAAAAACCAAAAAACAAACACA
86	3' PolyAC	AAAAACAAAAACAAAC

II.B.11. Excision Scars

[0242] In some aspects, the RNA or DNA genetic constructs disclosed herein further comprise one or more nucleotide sequences derived from the organism from which the intronic or the exonic sequences are derived. In some aspects, the one or more nucleotide sequences derived from the organism from which the intronic or the exonic sequences are derived are intronic sequences, exonic sequences, or any combination thereof. In some aspects, the one or more nucleotide sequences derived from the organism from which the intronic or the exonic sequences are derived are located in proximity of the splice sites. In some aspects, the one or more nucleotide sequences derived from the organism from which the intronic or the exonic sequences are derived are adjacent to the splice sites. In some aspects, the one or more nucleotide sequences derived from the organism from which the intronic or the exonic sequences are derived are not excised from the pre-circRNA during splicing. In these aspects, the one or more nucleotide sequences derived from the organism from which the intronic or the exonic sequences are derived are also herein referred to as pre-excision scars, when referring to those sequences as comprised in the pre-circRNAs or in the DNA constructs from which the pre-circRNAs can be transcribed, and as excision scars when referring to those sequences as comprised in the circRNAs.

[0243] In some aspects, the excision scar comprises a nucleotide sequence having at least 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or 100% identity with SEQ ID NOs: 80-81 or 82-83. In some aspects, the excision scar consists of SEQ ID NOs: 80-81 or 82-83.

[0244] Non-limiting examples of excision scars that are useful for the present genetic constructs are shown in **Table 11**.

Table 11. Exemplary excision scars

SEQ ID NO	DESCRIPTION	SEQUENCE
80	Excision scar 1 (P3450) DNA	CTACCGTTTAATATTGCGTCAGGTAGTAAACTACTAACTACAA CCTGCTGAAGCA
81	Excision scar 2 (P3450) DNA	CAGAGATGTTTCTTGGGT

82	Excision scar 1 (P3450) RNA	CUACCGUUUAAUAUUGCGUCAGGUAGUAAACUACUACUACAA CCUGCUGAAGCA
83	Excision scar 2 (P3450) RNA	CAGAGAUGUUUUCUUGGGU

II.B.12. Promoter

[0245] In some aspects, the RNA or DNA genetic constructs disclosed herein further comprises a promoter. In some aspects, the promoter is removed upon circularization. In some aspects the promoter is T7, T3, or SP6. In some aspects the promoter is a T7 promoter. In some aspects, the promoter is a T3 promoter. In some aspects, the promoter is a SP6 promoter. In some aspects, the promoter comprises a nucleotide sequence having at least 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or 100% identity with any one of SEQ ID NOs: 133-138. In some aspects, the promoter consists of a nucleotide sequence of any one of SEQ ID NOs: 133-138.

[0246] Non-limiting examples of promoters that are useful for the present genetic constructs are shown in **Table 12**.

Table 12. Exemplary promoters

SEQ ID NO	DESCRIPTION	SEQUENCE
133	T7 RNAP (DNA)	TAATACGACTCACTATAG
134	SP6 RNAP (DNA)	ATTTAGGTGACACTATAG
135	SP3 RNAP (DNA)	AATTAACCCCTCACTAAAG

II.C. Exemplary Constructs

[0247] Exemplary embodiments of the genetic constructs disclosed herein are described in detail below. For simplicity, the exemplary genetic constructs are preferentially described with reference to the RNA genetic constructs (*i.e.*, pre-circRNAs). It is to be understood that the DNA genetic constructs, from which such pre-circRNAs can be transcribed, are also to be considered as encompassed by the following description. As such, for example, any RNA nucleotide sequence reciting uracil (U) it is to be understood as encompassing a corresponding DNA nucleotide sequence reciting thymine (T).

II.C.1. Exemplary Genetic Construct A

[0248] In some aspects, the genetic constructs capable of forming circular RNAs disclosed herein (*i.e.*, pre-circRNAs or DNA genetic constructs from which pre-circRNAs can be transcribed) comprise a formula of: 5'-P-H-A-3S-E-5S-B-H'-3', wherein: P comprises a promoter (e.g., T7, T3, SP6, etc.), H comprises a 5' homology arm, A comprises a 3' intronic sequence, 3S

comprises a 3' splice site, E comprises an exonic sequence (*i.e.*, one or more exons or fragments thereof), 5S comprises a 5' splice site, B comprises a 5' intronic sequence, H' comprises a 3' homology arm, and wherein the promoter (P) is to be understood to be comprised in the DNA genetic construct, but not in the RNA genetic construct. In some aspects, the 5' and 3' intronic sequences are reverse complement of one another.

[0249] In some aspects, the intronic sequence comprises a nucleotide sequence derived from a group I intron. In some aspects, the group I intron comprises a nucleotide sequence having at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, at least about 99%, or 100% sequence identity with SEQ ID NOs: 1-5 or 6-10. In some aspects, the group I intron consists of SEQ ID NOs: 1-5 or 6-10.

[0250] In some aspects, the 5' intronic sequence comprises a nucleotide sequence derived from a group I intron. In some aspects, the 5' intronic sequence comprises an internal guide site (IGS). In some aspects, the 5' intronic sequence consists of an internal guide site (IGS). In some aspects, the IGS comprises a nucleotide sequence of any one of CTAA (or CUAA), CCCTCT (or CCCUCU), agT (or agU), CTT (or CUU), GGT (or GGU), TTGAG (or UUGAG), GGAGGG, tttGct (or uuuGcu), CTGAG (or CUGAG) or GGCC. In some aspects, the internal guide sequence comprises a nucleotide sequence of agT (or agU) or tttGct (or uuuGcu). In some aspects, the IGS consists of any one of CTAA (or CUAA), CCCTCT (or CCCUCU), agT (or agU), CTT (or CUU), GGT (or GGU), TTGAG (or UUGAG), GGAGGG, tttGct (or uuuGcu), CTGAG (or CUGAG) or GGCC. In some aspects, the internal guide sequence consists of any one of agT (or agU) or tttGct (or uuuGcu). In some aspects, the 5' intronic sequence comprises a nucleotide sequence having at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, at least about 99%, or 100% sequence identity with SEQ ID NOs: 23-26 or 27-30, or a nucleotide of AAA. In some aspects, the 5' intronic sequence consists of SEQ ID NOs: 23-26, or of a nucleotide sequence of AAA.

[0251] In some aspects, the 3' intronic sequence comprises a nucleotide sequence derived from a group I intron. In some aspects, the 3' intronic sequence comprises a catalytic core. In some aspects, the catalytic core comprises a sequence of GAG and CC, wherein the GAG and the CC are separated by (N)_n, wherein N is any one of A, C, G, T, or U, and n is an integer. In some aspects, the catalytic core comprises a sequence of ACG and CT(or U), wherein the ACG and the CT(or U) are separated by (N)_n, wherein N is any one of A, C, G, T, or U, and n is an integer.

[0252] In some aspects, the catalytic core comprises a nucleotide sequence of any one of AGAGA[N]₆₋₄₁A[N]₄₃₋₄₅GUCC, ACAGA[N]₆₋₄₁A[N]₄₃₋₄₅GUCG, or ACGACU[N]₇₋₆₈UAGUCU,

wherein N is any one of A, U, C, or G, and wherein each N in each one of the chains of [N]₆₋₄₁, [N]₄₃₋₄₅, or [N]₇₋₆₈, can be independently any one of A, U, C, or G. In some aspects, the catalytic core consists of AGAGA[N]₆₋₄₁A[N]₄₃₋₄₅GUCC, ACAGA[N]₆₋₄₁A[N]₄₃₋₄₅GUCG, or ACGACU[N]₇₋₆₈UAGUCU, wherein N is any one of A, U, C, or G, and wherein each N in each one of the chains of [N]₆₋₄₁, [N]₄₃₋₄₅, or [N]₇₋₆₈, can be independently any one of A, U, C, or G. In some aspects, the 3' intronic sequence comprises a nucleotide sequence having at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, at least about 99%, or 100% sequence identity with SEQ ID NO: 11-15 or 17-21. In some aspects, the 3' intronic sequence consists of SEQ ID NO: 11-15 or 17-21.

[0253] In some aspects, the 5' intronic sequence and the 3' intronic sequence together are less than a full intron.

[0254] In some aspects, the 5' intronic sequence comprises a nucleotide sequence having at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, at least about 99%, or 100% sequence identity with SEQ ID NOs: 23 or 27. In some aspects, the 3' intronic sequence comprises a nucleotide sequence having at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, at least about 99%, or 100% sequence identity with SEQ ID NOs: 11 or 17. In some aspects, the 5' intronic sequence comprises a nucleotide sequence having at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, at least about 99%, or 100% sequence identity with SEQ ID NOs: 23 or 27, and the 3' intronic sequence comprises a nucleotide sequence having at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, at least about 99%, or 100% sequence identity with SEQ ID NOs: 11 or 17.

[0255] In some aspects, the 5' intronic sequence comprises a nucleotide sequence having at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, at least about 99%, or 100% sequence identity with SEQ ID NOs: 24 or 28. In some aspects, the 3' intronic sequence comprises a nucleotide sequence having at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, at least about 99%, or 100% sequence identity with SEQ ID NOs: 12 or 18. In some aspects, the 5' intronic sequence comprises a nucleotide sequence having at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, at least about 99%, or 100% sequence identity with SEQ ID NOs: 24 or 28, and the 3' intronic sequence comprises a nucleotide sequence having at least 70%, at least 75%, at least 80%, at least

85%, at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, at least about 99%, or 100% sequence identity with SEQ ID NOs: 12 or 18.

[0256] In some aspects, the 5' intronic sequence comprises a nucleotide sequence of AAA. In some aspects, the 3' intronic sequence comprises a nucleotide sequence having at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, at least about 99%, or 100% sequence identity with SEQ ID NOs: 13 or 19. In some aspects, the 5' intronic sequence comprises a nucleotide sequence of AAA, and the 3' intronic sequence comprises a nucleotide sequence having at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, at least about 99%, or 100% sequence identity with SEQ ID NOs: 13 or 19.

[0257] In some aspects, the 5' intronic sequence comprises a nucleotide sequence having at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, at least about 99%, or 100% sequence identity with SEQ ID NOs: 25 or 29. In some aspects, the 3' intronic sequence comprises a nucleotide sequence having at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, at least about 99%, or 100% sequence identity with SEQ ID NOs: 14 or 20. In some aspects, the 5' intronic sequence comprises a nucleotide sequence having at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, at least about 99%, or 100% sequence identity with SEQ ID NOs: 25 or 29, and the 3' intronic sequence comprises a nucleotide sequence having at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, at least about 99%, or 100% sequence identity with SEQ ID NOs: 14 or 20.

[0258] In some aspects, the 5' intronic sequence comprises a nucleotide sequence having at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, at least about 99%, or 100% sequence identity with SEQ ID NOs: 26 or 30. In some aspects, the 3' intronic sequence comprises a nucleotide sequence having at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, at least about 99%, or 100% sequence identity with SEQ ID NOs: 15 or 21. In some aspects, the 5' intronic sequence comprises a nucleotide sequence having at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, at least about 99%, or 100% sequence identity with SEQ ID NOs: 26 or 30, and the 3' intronic sequence comprises a nucleotide sequence having at least 70%, at least 75%, at least 80%, at least

85%, at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, at least about 99%, or 100% sequence identity with SEQ ID NOs: 15 or 21.

[0259] In some aspects, the 3' splice site comprises 2 nucleotides, wherein one nucleotide is upstream of the site of splice and one nucleotide is downstream of the site of splice. In some aspects the nucleotide upstream of the site of splice is a G, and the nucleotide downstream of the site of splice is a N, wherein N is any one of A, C, G, T, or U. In some aspects, N is C. In some aspects, the 3' splice site comprises a sequence of formula 5'-G-N-3', wherein the site of splice is between G and N, and wherein N is any one of A, C, G, T, or U. In some aspects, the 3' splice site comprises a sequence of formula 5'-G-C-3'.

[0260] In some aspects, the 5' splice site comprises 2 nucleotides, wherein one nucleotide is upstream of the site of splice and one nucleotide is downstream of the site of splice. In some aspects the nucleotide upstream of the site of splice is a U or a T, and the nucleotide downstream of the site of splice is a N, wherein N is any one of A, C, G, T, or U. In some aspects, N is A. In some aspects, the 5' splice site comprises a sequence of formula 5'-U(or T)-N-3', wherein the site of splice is between U (or T) and N, and wherein N is any one of A, C, G, T, or U. In some aspects, the 5' splice site comprises a sequence of formula 5'-U(or T)-A-3'.

[0261] In some aspects, the 3' splice site comprises a nucleotide sequence having at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, at least about 99%, or 100% sequence identity with SEQ ID NOs: 31-35. In some aspects, the 5' splice site comprises a nucleotide sequence having at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, at least about 99%, or 100% sequence identity with SEQ ID NOs: 37-40, or a nucleotide sequence of NAGU ^AAA. In some aspects, the 3' splice site comprises a nucleotide sequence having at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, at least about 99%, or 100% sequence identity with SEQ ID NOs: 31-35, and the 5' splice site comprises a nucleotide sequence having at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, at least about 99%, or 100% sequence identity with SEQ ID NOs: 37-40, or a nucleotide sequence of NAGU ^AAA.

[0262] In some aspects, the 3' splice site comprises a nucleotide sequence having at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, at least about 99%, or 100% sequence identity with SEQ ID NOs: 31. In some aspects, the 5' splice site comprises a nucleotide sequence having at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, at least

about 99%, or 100% sequence identity with SEQ ID NOs: 37. In some aspects, the 3' splice site comprises a nucleotide sequence having at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, at least about 99%, or 100% sequence identity with SEQ ID NOs: 31, and the 5' splice site comprises a nucleotide sequence having at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, at least about 99%, or 100% sequence identity with SEQ ID NOs: 37.

[0263] In some aspects, the 3' splice site comprises a nucleotide sequence having at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, at least about 99%, or 100% sequence identity with SEQ ID NOs: 32. In some aspects, the 5' splice site comprises a nucleotide sequence having at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, at least about 99%, or 100% sequence identity with SEQ ID NOs: 38. In some aspects, the 3' splice site comprises a nucleotide sequence having at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, at least about 99%, or 100% sequence identity with SEQ ID NOs: 32, and the 5' splice site comprises a nucleotide sequence having at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, at least about 99%, or 100% sequence identity with SEQ ID NOs: 38.

[0264] In some aspects, the 3' splice site comprises a nucleotide sequence having at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, at least about 99%, or 100% sequence identity with SEQ ID NOs: 33. In some aspects, the 5' splice site comprises a nucleotide sequence of NAGU[^]AAA. In some aspects, the 3' splice site comprises a nucleotide sequence having at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, at least about 99%, or 100% sequence identity with SEQ ID NOs: 33, and the 5' splice site comprises a nucleotide sequence of NAGU[^]AAA.

[0265] In some aspects, the 3' splice site comprises a nucleotide sequence having at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, at least about 99%, or 100% sequence identity with SEQ ID NOs: 34. In some aspects, the 5' splice site comprises a nucleotide sequence having at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, at least about 99%, or 100% sequence identity with SEQ ID NOs: 39. In some aspects, the 3' splice site comprises a nucleotide sequence having at least 70%, at least 75%, at least 80%, at least 85%, at least 90%,

at least 95%, at least 96%, at least 97%, at least 98%, at least about 99%, or 100% sequence identity with SEQ ID NOs: 34, and the 5' splice site comprises a nucleotide sequence having at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, at least about 99%, or 100% sequence identity with SEQ ID NOs: 39.

[0266] In some aspects, the 3' splice site comprises a nucleotide sequence having at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, at least about 99%, or 100% sequence identity with SEQ ID NOs: 35. In some aspects, the 5' splice site comprises a nucleotide sequence having at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, at least about 99%, or 100% sequence identity with SEQ ID NOs: 40. In some aspects, the 3' splice site comprises a nucleotide sequence having at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, at least about 99%, or 100% sequence identity with SEQ ID NOs: 35, and the 5' splice site comprises a nucleotide sequence having at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, at least about 99%, or 100% sequence identity with SEQ ID NOs: 40.

[0267] In some aspects, the 5' homology arm comprises a nucleotide sequence having at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, at least about 99%, or 100% sequence identity with SEQ ID NO: 72 or 76. In some aspects, the 3' homology arm comprises a nucleotide sequence having at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, at least about 99%, or 100% sequence identity with SEQ ID NO: 73 or 77.

[0268] In some aspects, the 5' splice site is at the border between the 5' intronic sequence and the exonic sequence. In some aspects, the 3' splice site is at the border between the 3' intronic sequence and the exonic sequence. In some aspects, the intronic sequences are fully excised from the RNA genetic constructs (*i.e.*, pre-circRNAs). In some aspects, the resulting circRNAs do not comprise any intronic sequence. In some aspects, the resulting circRNAs do not comprise one or more intronic sequence residue. In some aspects, the resulting circRNAs do not comprise one or more nucleotide sequences derived from the organism from which the 5' intronic sequence and the 3' intronic sequence are derived, *i.e.*, the resulting circRNAs does not comprise "excision scars."

[0269] In some aspects, the pre-circRNA is capable of forming a circular RNA. In some aspects, the pre-circRNA is capable of forming a circular RNA in the presence of an exogenous G. In some aspects, the pre-circRNA is capable of forming a circular RNA, wherein the circular RNA

does not comprise any sequence derived from the organism from which the 5' intronic sequence and the 3' intronic sequence are derived ("excision scar").

[0270] In some aspects, the genetic constructs comprises from 5' to 3' a 5' homology arm, a 3' intronic sequence, a 3' splice site, a polyA, an IRES, one or more exons, a polyA, a 5' intronic sequence comprising an internal guide sequence (IGS), a 5' splice site, and H' comprises a 3' homology arm, wherein the 3' intronic sequence comprises a catalytic core, and wherein the RNA construct is capable of forming a circular RNA that does not contain any sequence derived from the genome from which the 5' intronic sequence and the 3' intronic sequence are derived ("excision scar").

II.C.2 Exemplary Genetic Construct B

[0271] In some aspects, the genetic constructs capable of forming circular RNAs disclosed herein (*i.e.*, pre-circRNAs or DNA genetic constructs from which pre-circRNAs can be transcribed) comprise a formula of: 5'-P-H-A-3S-E-5S-B-H'-3', wherein: P comprises a promoter (e.g., T7, T3, SP6, etc.), H comprises a 5' homology arm, A comprises an intronic sequence, 3S comprises a 3' splice site, E comprises an exonic sequence comprising a first excision scar precursor (X) and a second excision scar precursor X', B comprises a 5' intronic sequence, and H' comprises a 3' homology arm; and wherein the promoter (P) is to be understood to be comprised in the DNA genetic construct, but not in the RNA genetic construct. In some aspects, the 5' and 3' intronic sequences are reverse complement of one another.

[0272] In some aspects, the intronic sequence comprises a nucleotide sequence derived from a group I intron. In some aspects, the group I intron comprises a nucleotide sequence having at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, at least about 99%, or 100% sequence identity with SEQ ID NO: 5 or 10. In some aspects, the group I intron consists of SEQ ID NOs: 5 or 10.

[0273] In some aspects, the 5' intronic sequence comprises an internal guide site (IGS). In some aspects, the IGS comprises the nucleotide sequence of GGT (or GGU) or GGCC. In some aspects, the 5' intronic sequence comprises a nucleotide sequence having at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, at least about 99%, or 100% sequence identity with SEQ ID NO: 26 or 30. In some aspects, the 5' intronic sequence consists of SEQ ID NO: 26 or 30.

[0274] In some aspects, the 3' intronic sequence comprises a nucleotide sequence derived from a group I intron. In some aspects, the catalytic core comprises a nucleotide sequence of

ACGACU[N]₇₋₆₈UAGUCU, wherein N is any one of A, U, C, or G, and wherein each N in each one of the chains of [N]₆₋₄₁, [N]₄₃₋₄₅, or [N]₇₋₆₈, can be independently any one of A, U, C, or G. In some aspects, the catalytic core consists of ACGACU[N]₇₋₆₈UAGUCU, wherein N is any one of A, U, C, or G, and wherein each N in each one of the chains of [N]₆₋₄₁, [N]₄₃₋₄₅, or [N]₇₋₆₈, can be independently any one of A, U, C, or G. In some aspects, the 3' intronic sequence comprises a nucleotide sequence having at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, at least about 99%, or 100% sequence identity with SEQ ID NO: 15 or 21.

[0275] In some aspects the 5' intronic sequence and the 3' intronic sequence together are less than a full intron.

[0276] In some aspects, the 5' intronic sequence comprises a nucleotide sequence having at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, at least about 99%, or 100% sequence identity with SEQ ID NOs: 26 or 30. In some aspects, the 3' intronic sequence comprises a nucleotide sequence having at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, at least about 99%, or 100% sequence identity with SEQ ID NOs: 15 or 21. In some aspects, the 5' intronic sequence comprises a nucleotide sequence having at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, at least about 99%, or 100% sequence identity with SEQ ID NOs: 26 or 30, and the 3' intronic sequence comprises a nucleotide sequence having at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, at least about 99%, or 100% sequence identity with SEQ ID NOs: 15 or 21.

[0277] In some aspects, the 3' splice site comprises 2 nucleotides, wherein one nucleotide is upstream of the site of splice and one nucleotide is downstream of the site of splice. In some aspects the nucleotide upstream of the site of splice is a G, and the nucleotide downstream of the site of splice is a N, wherein N is any one of A, C, G, T, or U. In some aspects, N is C. In some aspects, the 3' splice site comprises a sequence of formula 5'-G-N-3', wherein the site of splice is between G and N, and wherein N is any one of A, C, G, T, or U. In some aspects, the 3' splice site comprises a sequence of formula 5'-G-C-3'.

[0278] In some aspects, the 5' splice site comprises 2 nucleotides, wherein one nucleotide is upstream of the site of splice and one nucleotide is downstream of the site of splice. In some aspects the nucleotide upstream of the site of splice is a U or a T, and the nucleotide downstream of the site of splice is a N, wherein N is any one of A, C, G, T, or U. In some aspects, N is A. In

some aspects, the 5' splice site comprises a sequence of formula 5'-U(or T)-N-3', wherein the site of splice is between U (or T) and N, and wherein N is any one of A, C, G, T, or U. In some aspects, the 5' splice site comprises a sequence of formula 5'-U(or T)-A-3'.

[0279] In some aspects, the 3' splice site comprises a nucleotide sequence having at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, at least about 99%, or 100% sequence identity with SEQ ID NOs: 36. In some aspects, the 5' splice site comprises a nucleotide sequence having at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, at least about 99%, or 100% sequence identity with SEQ ID NOs: 41. In some aspects, the 3' splice site comprises a nucleotide sequence having at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, at least about 99%, or 100% sequence identity with SEQ ID NOs: 36, and the 5' splice site comprises a nucleotide sequence having at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, at least about 99%, or 100% sequence identity with SEQ ID NOs: 41.

[0280] In some aspects, the 3' splice site comprises a nucleotide sequence having at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, at least about 99%, or 100% sequence identity with the 3' intronic sequence derived from *Synechococcus* PCC 6301 tRNA (LEU) as set forth in SEQ ID NOs: 36. In some aspects, the 5' splice site comprises a nucleotide sequence having at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, at least about 99%, or 100% sequence identity with the 5' intronic sequence derived from *Synechococcus* PCC 6301 tRNA (LEU) as set forth in SEQ ID NOs: 41. In some aspects, the 3' splice site comprises a nucleotide sequence having at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, at least about 99%, or 100% sequence identity with the 3' intronic sequence derived from *Synechococcus* PCC 6301 tRNA (LEU) as set forth in SEQ ID NOs: 36, and the 5' splice site comprises a nucleotide sequence having at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, at least about 99%, or 100% sequence identity with the 5' intronic sequence derived from *Synechococcus* PCC 6301 tRNA (LEU) as set forth in SEQ ID NOs: 41.

[0281] In some aspects, the 5' homology arm comprises a nucleotide sequence having at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, at least about 99%, or 100% sequence identity with SEQ ID NO: 72 or 76. In some aspects, the 3' homology arm comprises a nucleotide sequence having at least 70%, at least

75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, at least about 99%, or 100% sequence identity with SEQ ID NO: 73 or 77.

[0282] In some aspects, the 5' splice site is not at the border between the 5' intronic sequence and the exonic sequence. In some aspects, the 3' splice site is not at the border between the 3' intronic sequence and the exonic sequence. In some aspects, the intronic sequences are not fully excised from the RNA genetic constructs (*i.e.*, pre-circRNAs). In some aspects, the resulting circRNAs comprise at least a portion of the intronic sequence. In some aspects, the resulting circRNAs comprise one or more intronic sequence residues. In some aspects, the resulting circRNAs comprise one or more nucleotide sequences derived from the organism from which the 5' intronic sequence and the 3' intronic sequence are derived, *i.e.*, the resulting circRNAs comprise at least one "excision scar." In some aspects, the excision scar comprises a nucleotide sequence having at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, at least about 99%, or 100% sequence identity with any one of SEQ ID NOs: 80-81 or 82-83. In some aspects, the excision scar consists of any one of SEQ ID NOs: 80-81 or 82-83.

[0283] In some aspects, the pre-circRNA is capable of forming a circular RNA. In some aspects, the pre-circRNA is capable of forming a circular RNA in the presence of an exogenous G. In some aspects, the pre-circRNA is capable of forming a circular RNA, wherein the circular RNA comprising an excision scar.

II.C.3. Exemplary Genetic Construct C

[0284] In some aspects, the genetic constructs capable of forming circular RNAs disclosed herein (*i.e.*, pre-circRNAs or DNA genetic constructs from which pre-circRNAs can be transcribed) comprises a formula of: **5'-P-A-3S-E-3'**, wherein P comprises a promoter (e.g., T7, T3, SP6, etc.), A comprises a 3' intronic sequence, 3S comprises a 3' splice site, and E comprises an exonic sequence comprising a catalytic nucleotide (CN); and wherein the promoter (P) is to be understood to be comprised in the DNA genetic construct, but not in the RNA genetic construct.

[0285] In some aspects, the intronic sequence comprises a nucleotide sequence derived from a group I intron. In some aspects, the group I intron comprises a nucleotide sequence having at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, at least about 99%, or 100% sequence identity with SEQ ID NOs: 1, 3-5; or 6, 8-10. In some aspects, the group I intron consists of SEQ ID NOs: 1, 3-5; or 6, 8-10.

[0286] In some aspects, the 3' intronic sequence comprises a catalytic core. In some aspects, the catalytic core comprises a sequence of GAG and CC, wherein the GAG and the CC are separated by (N)_n, wherein N is any one of A, C, G, T, or U, and n is an integer. In some aspects, the catalytic core comprises a sequence of ACG and CT(or U), wherein the ACG and the CT(or U) are separated by (N)_n, wherein N is any one of A, C, G, T, or U, and n is an integer. In some aspects, the catalytic core comprises a nucleotide sequence of AGAGA[N]₆₋₄₁A[N]₄₃₋₄₅GUCC, or ACGACU[N]₇₋₆₈UAGUCU, wherein N is any one of A, U, C, or G, and wherein each N in each one of the chains of [N]₆₋₄₁, [N]₄₃₋₄₅, or [N]₇₋₆₈, can be independently any one of A, U, C, or G. In some aspects, the catalytic core consists of AGAGA[N]₆₋₄₁A[N]₄₃₋₄₅GUCC, or ACGACU[N]₇₋₆₈UAGUCU, wherein N is any one of A, U, C, or G, and wherein each N in each one of the chains of [N]₆₋₄₁, [N]₄₃₋₄₅, or [N]₇₋₆₈, can be independently any one of A, U, C, or G. In some aspects, the 3' intronic sequence comprises a nucleotide sequence having at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, at least about 99%, or 100% sequence identity with any one of SEQ ID NOs: 13-16 or 19-22. In some aspects, the 3' intronic sequence consists of any one of SEQ ID NOs: 13-16 or 19-22.

[0287] In some aspects the 3' intronic sequence is less than a full intron.

[0288] In some aspects, the 3' splice site comprises 2 nucleotides, wherein one nucleotide is upstream of the site of splice and one nucleotide is downstream of the site of splice. In some aspects the nucleotide upstream of the site of splice is a G, and the nucleotide downstream of the site of splice is a N, wherein N is any one of A, C, G, T, or U. In some aspects, N is C. In some aspects, the 3' splice site comprises a sequence of formula 5'-G-N-3', wherein the site of splice is between G and N, and wherein N is any one of A, C, G, T, or U. In some aspects, the 3' splice site comprises a sequence of formula 5'-G-C-3'. In some aspects, the 3' splice site comprises a nucleotide sequence having at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, at least about 99%, or 100% sequence identity with SEQ ID NOs: 31 or 33-35. In some aspects, the 3' splice site consists of SEQ ID NOs: 31 or 33-35.

[0289] In some aspects, the 5' homology arm comprises a nucleotide sequence having at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, at least about 99%, or 100% sequence identity with SEQ ID NO: 72 or 76.

[0290] In some aspects, the 3' splice site is at the border between the 3' intronic sequence and the exonic sequence. In some aspects, the intronic sequences are fully excised from the RNA genetic constructs (*i.e.*, pre-circRNAs). In some aspects, the resulting circRNAs do not comprise

any intronic sequence. In some aspects, the resulting circRNAs do not comprise one or more intronic sequence residue. In some aspects, the resulting circRNAs do not comprise one or more nucleotide sequences derived from the organism from which the 3' intronic sequence is derived, *i.e.*, the resulting circRNAs does not comprise "excision scars."

[0291] In some aspects, the pre-circRNA is capable of forming a circular RNA. In some aspects, the pre-circRNA is capable of forming a circular RNA in the presence of an exogenous G. In some aspects, the pre-circRNA is capable of forming a circular RNA, wherein the circular RNA does not comprise an excision scar.

II.C.4. Exemplary Genetic Construct D

[0292] In some aspects, the genetic constructs capable of forming circular RNAs disclosed herein (*i.e.*, pre-circRNAs or DNA genetic constructs from which pre-circRNAs can be transcribed) comprises a formula of: **5'-P-A-3S-E-3'**, wherein P comprises a promoter (e.g., T7, T3, SP6, etc.), A comprises a 3' intronic sequence, E comprises an exonic sequence (E) comprising a first portion of the exonic sequence (E'), a first homology arm (HI), a second portion of the exonic sequence (E''), a second homology arm (HI'), and a third exonic sequence (E''') comprising a catalytic nucleotide (CN); and wherein the promoter (P) is to be understood to be comprised in the DNA genetic construct, but not in the RNA genetic construct.

[0293] In some aspects, the intronic sequence comprises a nucleotide sequence derived from a group I intron. In some aspects, the group I intron comprises a nucleotide sequence having at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, at least about 99%, or 100% sequence identity with SEQ ID NOs: 1 or 6. In some aspects, the group I intron consists of SEQ ID NOs: 1 or 6.

[0294] In some aspects, the 3' intronic sequence comprises a catalytic core. In some aspects, the catalytic core comprises a sequence of AGAGA[N]₆₋₄₁A[N]₄₃₋₄₅GUCC, wherein N is any one of A, U, C, or G, and wherein each N in each one of the chains of [N]₆₋₄₁, or [N]₄₃₋₄₅, can be independently any one of A, U, C, or G. In some aspects, the catalytic core consists of AGAGA[N]₆₋₄₁A[N]₄₃₋₄₅GUCC, wherein N is any one of A, U, C, or G, and wherein each N in each one of the chains of [N]₆₋₄₁, or [N]₄₃₋₄₅, can be independently any one of A, U, C, or G. In some aspects, the 3' intronic sequence comprises a nucleotide sequence having at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, at least about 99%, or 100% sequence identity with any one of SEQ ID NOs: 11 or 17. In some aspects, the 3' intronic sequence consists of any one of SEQ ID NOs: 11 or 17.

[0295] In some aspects the 3' intronic sequence is less than a full intron.

[0296] In some aspects, the 3' splice site comprises 2 nucleotides, wherein one nucleotide is upstream of the site of splice and one nucleotide is downstream of the site of splice. In some aspects the nucleotide upstream of the site of splice is a G, and the nucleotide downstream of the site of splice is a N, wherein N is any one of A, C, G, T, or U. In some aspects, N is A. In some aspects, the 3' splice site comprises a sequence of formula 5'-G-N-3', wherein the site of splice is between G and N, and wherein N is any one of A, C, G, T, or U. In some aspects, the 3' splice site comprises a sequence of formula 5'-G-A-3'. In some aspects, the 3' splice site comprises a nucleotide sequence having at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, at least about 99%, or 100% sequence identity with SEQ ID NOs: 169 or 170. In some aspects, the 3' splice site consists of SEQ ID NOs: 169 or 170.

[0297] In some aspects, the first homology arm comprises a nucleotide sequence having at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, at least about 99%, or 100% sequence identity with SEQ ID NOs: 171 or 172. In some aspects, the second homology arm comprises a nucleotide sequence having at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, at least about 99%, or 100% sequence identity with SEQ ID NOs: 173 or 174.

[0298] In some aspects, the 3' splice site is at the border between the 3' intronic sequence and the exonic sequence. In some aspects, the intronic sequences are fully excised from the RNA genetic constructs (*i.e.*, pre-circRNAs). In some aspects, the resulting circRNAs do not comprise any intronic sequence. In some aspects, the resulting circRNAs do not comprise one or more intronic sequence residue. In some aspects, the resulting circRNAs do not comprise one or more nucleotide sequences derived from the organism from which the 3' intronic sequence is derived, *i.e.*, the resulting circRNAs does not comprise "excision scars."

[0299] In some aspects, the pre-circRNA is capable of forming a circular RNA. In some aspects, the pre-circRNA is capable of forming a circular RNA in the presence of an exogenous G. In some aspects, the pre-circRNA is capable of forming a circular RNA, wherein the circular RNA does not comprise an excision scar.

II.C.5. Exemplary Genetic Construct E

[0300] In some aspects, the genetic constructs capable of forming circular RNAs disclosed herein (*i.e.*, pre-circRNAs or DNA genetic constructs from which pre-circRNAs can be

transcribed) comprise a formula of: **5'-P-B-A-3S-E-3'**, wherein P comprises a promoter (e.g., T7, T3, SP6, etc.), B comprises a 5' intronic sequence comprising an internal guide sequence (IGS), A comprises a 3' intronic sequence, 3S comprises a 3' splice site, and E comprises an exonic sequence comprising a catalytic nucleotide; and wherein the promoter (P) is to be understood to be comprised in the DNA genetic construct, but not in the RNA genetic construct.

[0301] In some aspects, the intronic sequence comprises a nucleotide sequence derived from a group I intron. In some aspects, the group I intron comprises a nucleotide sequence having at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, at least about 99%, or 100% sequence identity with SEQ ID NOs: 1 or 6. In some aspects, the group I intron consists of SEQ ID NOs: 1 or 6.

[0302] In some aspects, the 5' intronic sequence comprises an internal guide site (IGS). In some aspects, the IGS comprises the nucleotide sequence of CTAA (or CUUAA) or TTGAG (or UUGAG). In some aspects, the 5' intronic sequence comprises a nucleotide sequence having at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, at least about 99%, or 100% sequence identity with SEQ ID NOs: 175 or 176. In some aspects, the 5' intronic sequence consists of SEQ ID NOs: 175 or 176.

[0303] In some aspects, the 3' intronic sequence comprises a catalytic core. In some aspects, the catalytic core comprises a sequence of AGAGA[N]₆₋₄₁A[N]₄₃₋₄₅GUCC, wherein N is any one of A, U, C, or G, and wherein each N in each one of the chains of [N]₆₋₄₁, or [N]₄₃₋₄₅, can be independently any one of A, U, C, or G. In some aspects, the catalytic core consists of AGAGA[N]₆₋₄₁A[N]₄₃₋₄₅GUCC, wherein N is any one of A, U, C, or G, and wherein each N in each one of the chains of [N]₆₋₄₁, or [N]₄₃₋₄₅, can be independently any one of A, U, C, or G. In some aspects, the 3' intronic sequence comprises a nucleotide sequence having at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, at least about 99%, or 100% sequence identity with any one of SEQ ID NOs: 177 or 178. In some aspects, the 3' intronic sequence consists of any one SEQ ID NOs: 177 or 178.

[0304] In some aspects, the 5' intronic sequence comprises a nucleotide sequence having at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, at least about 99%, or 100% sequence identity with SEQ ID NOs: 175 or 176, and the 3' intronic sequence comprises a nucleotide sequence having at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, at least about 99%, or 100% sequence identity with any one of SEQ ID NOs: 177 or 178. In some

aspects, the 5' intronic sequence consists of any one of SEQ ID NOs: 175 or 176, and the 3' intronic sequence consists of any one of SEQ ID NOs: 177 or 178.

[0305] In some aspects, the 3' splice site comprises a sequence having at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, at least about 99%, or 100% sequence identity with SEQ ID NOs: 179 or 180. In some aspects, the 3' splice site consists of any one of SEQ ID NOs: 179 or 180.

[0306] In some aspects the 5' intronic sequence and the 3' intronic sequence together are less than a full intron.

[0307] In some aspects, the 3' splice site is at the border between the 3' intronic sequence and the exonic sequence. In some aspects, the intronic sequences are fully excised from the RNA genetic constructs (*i.e.*, pre-circRNAs). In some aspects, the resulting circRNAs do not comprise any intronic sequence. In some aspects, the resulting circRNAs do not comprise one or more intronic sequence residue. In some aspects, the resulting circRNAs do not comprise one or more nucleotide sequences derived from the organism from which the 3' intronic sequence is derived, *i.e.*, the resulting circRNAs does not comprise "excision scars."

[0308] In some aspects, the pre-circRNA is capable of forming a circular RNA. In some aspects, the pre-circRNA is capable of forming a circular RNA in the presence of an exogenous G. In some aspects, the pre-circRNA is capable of forming a circular RNA, wherein the circular RNA does not comprise an excision scar.

[0309] In some aspects, the genetic constructs disclosed herein comprise a nucleotide sequence having at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, at least about 99%, or 100% sequence identity with SEQ ID NOs: 113-122, wherein "N" represents any one of A, T (or U), C, or G; and wherein "n" represents any integer. In some aspects, "(N)n" comprises a nucleotide sequence of an exonic sequences. In some aspects, "(N)n" comprises a nucleotide sequence of an exonic sequences and a nucleotide sequence of an IRES. Non-limiting examples of constructs of the present disclosure are shown in **Table 13**.

SEQ ID NO	DESCRIPTION	SEQUENCE	FORMULA
113	Anabaena PCC7120	CACTAATACGACTCACTATAGGGAGACCCCTCGACCGTCGATTG TCCACTGGTCAACAATAGATGACTTACAATAATCGGAAGGTG CAGAGACTCGACGGGAGCTACCCTAACGTCAAGACGAGGGTAA AGAGAGAGTCCAATTCTCAAAGCCAATAGGCAGTAGCGAAAGC TGCAAGAGAATGAAAATCCGTTGACCTTAAACGGTCGTGTGGG TTCAAGTCCCTCCACCCCAACGCGGAAACGCAATAGCCGAAA AACAAAAACAAAAAAACAAAAAAACCAAAAAACAAAA CACA (N) nAAAAAACAAAAACAAACCGCTATTTATGCGTTAC CGGCGAGACGCTACGGACTTAAATAATTGAGCCTTAAAGAAGA	5'-P-H-A-3S- E-5S-B-H'-3' (construct A)

		AATTCTTTAAGTGGATGCTCTCAAACCTCAGGGAAACCTAAATC TAGTTATAGACAAGGCAATCCTGAGCCAAGCCGAAGTAGTAAT TAGTAAGACCAGTGGACAATCGACGGATAACAGCATATCTAGA TAAGAAGAGCATCCCATGGTCATAGCTGTTTCTTGGCAGCTCT GGCCCGTGTCTCAAAATCTCTGATGTTACATTGCACAAGATAA AAATATATCATCATGAACAATAAAACTGTCTGCTTACATAAAC AGTAATACAAGGGGTGTTATGAGCCATATTCAACGGGAAACGT CGAGGCCGCGATTAAATTTCCAACATGGATGCTGATTTATATGG GTATAAATGGGCTCGCGATAATGTCTGGGCAATCAGGTGCGACA ATCTATCGCTTGTATGGGAAGCCCGATGCGCCAGAGTTGTTTC TGAAACATGGCAAAGGTAGCGTTGCCAATGATGTTACAGATGA GATGGTCAGACTAAACTGGCTGACGGAATTTATGCCTCTTCCG ACCATCAAGCATTTTATCCGTACTCCTGATGATGCATGGTTAC TCACCACTGCGATCCCCGAAAAACAGCATTCCAGGTATTAGA AGAATATCCTGATTCAGGTGAAAAATATTGTTGATGCGCTGGCA GTGTTCTCGCGCCGGTTGCATTTCGATTCTGTTTGTAAATTGTC CTTTTAACAGCGATCGCGTATTTTCGTCTCGCTCAGGCGCAATC ACGAATGAATAACGGTTTGGTTGATGCGAGTGATTTTGATGAC GAGCGTAATGGCTGGCCTGTTGAACAAGTCTGGAAAGAAATGC ATAAACTTTTGCCATTCTCACCGGATTCAGTCGTCACTCATGG TGATTTCTCACTTGATAACCTTATTTTGACGAGGGGAAATTA ATAGGTTGTATTGATGTTGGACGAGTCGGAATCGCAGACCGAT ACCAGGATCTTGCCATCCTATGGAACGCTCGGTGAGTTTTTC TCCTTCATTACAGAAACGGCTTTTTTCAAAAATATGGTATTGAT AATCCTGATATGAATAAATTGCAGTTTCATTTGATGCTCGATG AGTTTTTCTAATCAGAAATTGGTTAATTGGTTGTAACTGGCA GAGCATTACGCTGACTTGACGGGACGGCGCAAGCTCATGACCA AAATCCCTTAACGTGAGTTACGCGTCGTTCCACTGAGCGTCAG ACCCCGTAGAAAAGATCAAAGGATCTTCTTGAGATCCTTTTTT TCTGCGCGTAATCTGCTGCTTGCAAACAAAAAACACCGCTA CCAGCGGTGGTTTGTGTTGCCGGATCAAGAGCTACCAACTCTTT TTCCGAAGGTAACCTGGCTTCAGCAGAGCGCAGATACCAAATAC TGTTCTTCTAGTGTAGCCGTAGTTAGGCCACCACCTCAAGAAC TCTGTAGCACCGCTACATACCTCGCTCTGCTAATCCTGTTAC CAGTGGCTGCTGCCAGTGGCGATAAGTCTGTCTTACCGGGTT GGACTCAAGACGATAGTTACCGGATAAGGCGCAGCGGTGCGGC TGAACGGGGGGTTTCGTGCACACAGCCCAGCTTGGAGCGAACGA CCTACACCGAACTGAGATACCTACAGCGTGAGCTATGAGAAAG CGCCACGCTTCCCGAAGGGAGAAAGGCGGACAGGTATCCGGTA AGCGGCAGGGTCGGAACAGGAGAGCGCACGAGGGAGCTTCCAG GGGGAACCGCTTGGTATCTTTATAGTCTGTGCGGTTTCGCCA CCTCTGACTTGAGCGTCGATTTTTGTGATGCTCGTCAGGGGGG CGGAGCCTATGGAAAAACGCCAGCAACGCGGCTTTTTTACGGT TCCTGGCCTTTTGCTGGCCTTTTGCTCACATGTTCTTTCCTGC GTTATCCCTGATTCTGTGGATAACCGTATTACCGCCTTTGAG TGAGCTGATACCGCTCGCCGACCGGAACGACCGAGCGCAGCG AGTCAGTGAGCGAGGAAGCGGAGGAGCGCCCAATACGCAAAAC GCCTCTCCCCGCGCTTGGCCGATTCATTAATGCAGCTGGCAC GACAGGTTTCCCGACTGGAAAGCGGGCAGTGAGCGCAACGCAA TTAATACGCGTACCGCTAGCCAGGAAGAGTTTGTAGAAACGCA AAAAGGCCATCCGTGAGGATGGCCTTCTGCTTAGTTTGTATGCC TGGCAGTTTATGGCGGGCGTCTGCCCCGCCACCCTCCGGGCCG TTGCTTCAACAACGTTCAAATCCGCTCCCGGCGGATTTGTCTTA CTCAGGAGAGCGTTACCCGACAAACAACAGATAAAACGAAAGG CCCAGTCTTCCGACTGAGCCTTTTCGTTTTATTTGATGCCTGGC AGTTCCCTACTCTCGCGTTAACGCTAGCATGGATGTTTTCCCA GTCACGACGTTGTAAACGACGGCCAGTCTTAAGCTCGGCTCT TCG	
114	Tetrahymena	CCCGTAGAAAAGATCAAAGGATCTTCTTGAGATCCTTTTTTTC TGCGCGTAATCTGCTGCTTGCAAACAAAAAACACCGCTACC AGCGGTGGTTTGTGTTGCCGGATCAAGAGCTACCAACTCTTTTT	5'-P-H-A-3S- E-5S-B-H'-3' (Construct A)

thermoph ila	CCGAAGGTAACTGGCTTCAGCAGAGCGCAGATACCAAATACTG TTCTTCTAGTGTAGCCGTAGTTAGGCCACCACTTCAAGAAGCTC TG TAGCACCGCCTACATACCTCGCTCTGCTAATCCTGTTACCA GTGGCTGCTGCCAGTGGCGATAAGTCGTGTCTTACCGGGTTGG ACTCAAGACGATAGTTACCGGATAAGGCGCAGCGGTCTGGGCTG AACGGGGGGTTTCGTGCACACAGCCAGCTTGGAGCGAACGACC TACACCGAACTGAGATACCTACAGCGTGAGCTATGAGAAAGCG CCACGCTTCCCGAAGGGAGAAAGGCGGACAGGTATCCGGTAAG CGGCAGGGTCGGAACAGGAGAGCGCACGAGGGAGCTTCCAGGG GGAAACGCCTGGTATCTTTATAGTCTGTCTGGGTTTCGCCACC TCTGACTTGAGCGTCGATTTTTGTGATGCTCGTCAGGGGGGCG GAGCCTATGGAAAAACGCCAGCAACGCGGCCTTTTTACGGTTT CTGGCCTTTTGTCTGGCCTTTTGTCTACATGTTCTTTCCTGCGT TATCCCTGATTCTGTGGATAACCGTATTACCGCCTTTGAGTG AGCTGATACCGCTCGCCGACGCCAACGACCGAGCGCAGCGAG TCAGTGAGCGAGGAAGCGGAGGAGCGCCCAATACGCAAAACCGC CTCTCCCCGCGCGTTGGCCGATTCAATTAATGCAGCTGGCACGA CAGGTTTTCCCGACTGGAAAGCGGGCAGTGAGCGCAACGCAATT AATACGCGTACCGCTAGCCAGGAAGAGTTTGTAGAAACGCAAA AAGGCCATCCGTCAGGATGGCCTTCTGCTTAGTTTGATGCCTG GCAGTTTTATGGCGGGCGTCTGCCCCGCCACCTCCGGGCCGTT GCTTACAAACGTTCAAATCCGCTCCCGGCGGATTTGTCTTACT CAGGAGAGCGTTTACCGACAAACAACAGATAAAACGAAAGGCC CAGTCTTCCGACTGAGCCTTTTCGTTTTATTTGATGCTTGGCAG TTCCCTACTCTCGCGTTAACGCTAGCATGGATGTTTTCCAGT CACGACGTTGTAAACGACGGCCAGTCTTAAGCTCGGCTCTTC GCACTAATACGACTCACTATAGGGAGTCCCTCGACCGTCGATT GTCCACTGGTCATTTACCTTTGGAGGGAAAAGTTATCAGGCAT GCACCTGGTAGCTAGTCTTTAAACCAATAGATTGCATCGGTTTT AAAAGGCAAGACCGTCAAATTGCGGGAAAGGGGTCAACAGCCG TTCAGTACCAAGTCTCAGGGGAAACTTTGAGATGGCCTTGCAA AGGGTATGGTAATAAGCTGACGGACATGGTCCTAACACGCGAG CCAAGTCCTAAGTCAACAGATCTTCTGTTGATATGGATGCAGT TCACAGACTAAATGTCGGTTCGGGGAAGATGTATTTCTCTCATA AGATATAGTCGGACCTCTCCTTAATGGGAGCTAGCGGATGAAG TGATGCAACACTGGAGCCGCTGGGAACTAATTTGTATGCGAAA GTATATTTGATTAGTTTTGGAGTACTCGTAAGGTAAAAATCCGT TGACCTTAAACGGTCGTGTGGGTTCAAGTCCCTCCACCCCCAC GCCGGAACGCAATAGCCGAAAAACAAAAACAAAAAACAA AAAAAACCAAAAAACAAAAACACA (N) nAAAAACAAAAAA CAAAACGGCTATTATGCGTTACCGCGAGACGCTACGGACCT CTAAATAGCAATGACCAGTGGACAATCGACGGATAACAGCATA TCTAGATAAGAAGAGCATCCCATGGTCATAGCTGTTTCTTGGC AGCTCTGGCCCGTGTCTCAAAATCTCTGATGTTACATTGCACA AGATAAAAAATATATCATCATGAACAATAAACTGTCTGCTTAC ATAAACAGTAATAACAAGGGGTGTTATGAGCCATATTCACGGG AAACGTCGAGGCCGCGATTAAATTCCAACATGGATGCTGATTT ATATGGGTATAAATGGGCTCGCGATAATGTGGGCAATCAGGT GCGACAATCTATCGCTTGTATGGGAAGCCCGATGCGCCAGAGT TGTTTTCTGAAACATGGCAAAGGTAGCGTTGCCAATGATGTTAC AGATGAGATGGTCAGACTAAACTGGCTGACGGAATTTATGCCT CTTCCGACCATCAAGCATTTTATCCGTACTCCTGATGATGCAT GGTTACTCACCCTGCGATCCCCGAAAAACAGCATTCAGGT ATTAGAAGAATATCCTGATTCAGGTGAAAAATTTGTTGATGCG CTGGCAGTGTTCCTGCGCCGGTTGCATTCGATTCCTGTTTGTA ATTGTCTTTTAAACAGCGATCGCGTATTTTCGTCTCGCTCAGGC GCAATCACGAATGAATAACGGTTTGGTTGATGCGAGTGATTTT GATGACGAGCGTAATGGCTGGCCTGTTGAACAAGTCTGGAAG AAATGCATAAACTTTTGCCATTCTACCGGATTTCAGTCGTAC TCATGGTGATTTCTCACTTGATAACCTTATTTTTGACGAGGGG AAATTAATAGGTTGTATGATGTTGGACGAGTCGGAATCGCAG
-----------------	---

		ACCGATACCAGGATCTTGCCATCCTATGGAACGCCTCGGTGA GTTTTCTCCTTCATTACAGAAACGGCTTTTTCAAAAATATGGT ATTGATAATCCTGATATGAATAAATTGCAGTTTTCATTTGATGC TCGATGAGTTTCTTAATCAGAAATTGGTTAATTGGTTGTAACA CTGGCAGAGCATTACGCTGACTTGACGGGACGGCGCAAGCTCA TGACCAAAATCCCTTAACGTGAGTTACGCGTCGTTCCACTGAG CGTCAGAC	
115	Azoarcus	CCCGTAGAAAAGATCAAAGGATCTTCTTGAGATCCTTTTTTTTC TGCGCGTAATCTGCTGCTTGCAAACAAAAAACACCGCTACC AGCGGTGGTTTGTGTTGCCGGATCAAGAGCTACCAACTCTTTTT CCGAAGGTAACCTGGCTTCAGCAGAGCGCAGATACCAAATACTG TTCTTCTAGTGTAGCCGTAGTTAGGCCACCACCTCAAGAACTC TGTAGCACCGCCTACATACCTCGCTCTGCTAATCCTGTTACCA GTGGCTGCTGCCAGTGGCGATAAGTCGTGTCTTACCGGGTTGG ACTCAAGACGATAGTTACCGGATAAGGCGCAGCGGTCGGGCTG AACGGGGGGTTCTGTGCACACAGCCCAGCTTGGAGCGAACGACC TACACCGAACTGAGATACCTACAGCGTGAGCTATGAGAAAGCG CCACGCTTCCCGAAGGGAGAAAGGCGGACAGGTATCCGGTAAG CGGCAGGGTCGGAACAGGAGAGCGCACGAGGGAGCTTCCAGGG GGAAACGCCTGGTATCTTTATAGTCTGTGCGGTTTCGCCACC TCTGACTTGAGCGTCGATTTTTGTGATGCTCGTCAGGGGGGCG GAGCCTATGGAAAAACGCCAGCAACGCGGCCTTTTACGGTTC CTGGCCTTTTGTGCTGGCCTTTTGTCTACATGTTCTTCTGCGT TATCCCCGTGATTCTGTGGATAACCGTATTACCGCCTTTGAGTG AGCTGATACCGCTCGCCGAGCCGAACGACCGAGCGCAGCGAG TCAGTGAGCGAGGAAGCGGAGGAGCGCCAATACGCAAACCGC CTCTCCCCGCGCGTTGGCCGATTCTAATGACAGCTGGCACGA CAGGTTTCCCGACTGGAAGCGGGCAGTGAGCGCAACGCAATT AATACGCGTACCGCTAGCCAGGAAGAGTTTGTAGAAACGCAAA AAGGCCATCCGTACGGATGGCCTTCTGCTTAGTTTGATGCGCT GCAGTTTATGGCGGGCGTCCTGCCCGCCACCCTCCGGGCGGT GCTTCACAACGTTCAAATCCGCTCCCGGCGGATTTGTCTACT CAGGAGAGCGTTACCGACAAACAACAGATAAAACGAAAGGCC CAGTCTTCCGACTGAGCCTTTCTGTTTATTTGATGCCTGGCAG TTCCCTACTCTCGCGTTAACGCTAGCATGGATGTTTCCAGT CACGACGTTGTAAAACGACGGCCAGTCTTAAGCTCGGCTCTTC GACTAATACGACTCACTATAGGGAGACCCTCGACCGTCGATT GTCCACTGGTCTtGctCCTTGCGCCGGGAAACCACGCAAGGG ATGGTGTCAAATTCGGCGAAACCTAAGCGCCCGCCGGGCGTA TGGCAACGCCGAGCCAAGCTTCGGCGCCTGCGCCGATGAAGGT GTAGAGACTAGACGGCACCCACCTAAGGCAAACGCTATGGTGA AGGCATAGTCCAGGGAGTGGCGAAAGTCACACAAACCGGAAAA TCCGTTGACCTTAAACGGTCGTGTGGGTTCAAGTCCCTCCACC CCCACGCCGAAACGCAATAGCCGAAAAACAAAAACAAAAAA AACAAAAAAAACCAAAAAACAAAAACACA (N) nAAAAAACCA AAAAACAAACCGCTATTATGCGTTACCGGCGAGACGCTACGG AagTaaaGACCACTGGACAATCGACGATAACAGCATATCTAG ATAAGAAGAGCATCCCATGGTCATAGCTGTTTCTGGCAGCTC TGGCCCGTGTCTCAAAATCTCTGATGTTACATTGCACAAGATA AAAATATATCATCATGAACAATAAACTGTCTGCTTACATAAAA CAGTAATACAAGGGGTGTTATGAGCCATATTCAACGGGAAACG TCGAGGCCGCGATTAAATTCCAACATGGATGCTGATTTATATG GGTATAAATGGGCTCGCGATAATGTCGGGCAATCAGGTGCGAC AATCTATCGCTTGTATGGGAAGCCCGATGCGCCAGAGTTGTTT CTGAAACATGGCAAAGGTAGCGTTGCCAATGATGTTACAGATG AGATGGTCAGACTAACTGGCTGACGGAATTTATGCCTCTTCC GACCATCAAGCATTTTATCCGTACTCCTGATGATGCATGGTTA CTCACCCTGCGATCCCGGAAAAACAGCATTCAGGTATTAG AAGAATATCCTGATTACGGTGAAAAATATTGTTGATGCGCTGGC AGTGTTCCTGCGCCGGTTGCATTCGATTCTGTGTTGTAATTGT CCTTTTAAACAGCGATCGCGTATTTCTGCTCTGCTCAGGCGCAAT	5'-P-H-A-3S- E-5S-B-H'-3' (Construct A)

		CACGAATGAATAACGGTTTGGTTGATGCGAGTGATTTTGATGACGAGCGTAATGGCTGGCCTGTTGAACAAGTCTGGAAAAGAAATGCATAAACCTTTTGCCATTCTCACCGGATTCAGTCGTCACCTCATGTGATTTCTCACTTGATAACCTTATTTTGGACGAGGGGAAATTAATAGGTTGTATTGATGTTGGACGAGTCGGAATCGCAGACCGATACCAGGATCTTGCCATCCTATGGAACCTGCCTCGGTGAGTTTTCTCCTTCATTACAGAAACGGCTTTTCAAAAATATGGTATTGATAATCCTGATATGAATAAATTGCAGTTTCATTTGATGCTCGATGAGTTTTTCTAATCAGAATTGGTTAATTGGTTGTAACACTGGCAGAGCATTACGCTGACTTGACGGGACGGCGCAAGCTCATGACCAAAATCCCTTAACGTGAGTTACGCGTCGTTCCACTGAGCGTGACGAC	
116	Anabaena azollae	CCCGTAGAAAAGATCAAAGGATCTTCTTGAGATCCTTTTTTTCTGCGCGTAATCTGCTGCTTGCAAAACAAAAAACCCACCGCTACCAGCGGTGGTTTGTGTTGCGCGATCAAGAGCTACCAACTCTTTTTCCGAAGGTAACCTGGCTTCAGCAGAGCGCAGATACCAAATACTGTTCTTCTAGTGTAGCCGTAGTTAGGCCACCCTTCAAGAACCTCTGTAGCACCGCCTACATACCTCGCTCTGCTAATCCTGTTACCATGTGGCTGCTGCCAGTGGCGATAAGTCGTGTCTTACCGGGTTGGACTCAAGACGATAGTTACCGGATAAGGCGCAGCGGTCTGGGCTGAACGGGGGGTTCTGTGCACACAGCCCAGCTTGGAGCGAACGACCTACACCGAACTGAGATACCTACAGCGTGAGCTATGAGAAAGCGCCACGCTTCCCGAAGGGAGAAAGGCGGACAGGTATCCGGTAAGCGGCAGGGTTCGGAACAGGAGAGCGCACGAGGGAGCTTCCAGGGGAAACGCTGGTATCTTTATAGTCTGTCTGGGTTTCGCCACCCTGTGACTTGAGCGTCGATTTTTGTGATGCTCGTCAGGGGGGCGGAGCCTATGGAACACGCCAGCAACGCGGCCTTTTTACGGTTCTGCGCCTTTTGCTGGCCTTTTGCTCACATGTTCTTTCCTGCGTTATCCCCTGATTCTGTGGATAACCGTATTACCGCCTTTGAGTGAAGTGTATACCGCTCGCCGACGCCAAGACCGAGCGCAGCGAGTCAGTGAGCGAGGAAGCGGAGGAGCGCCCAATACGCAACCGCCTCTCCCCGCGCGTTGGCCGATTCAATTAATGCAGCTGGCACGACAGGTTTCCCGACTGGAAGCGGGCAGTGAGCGCAACGCAATTAAATACGCGTACCGCTAGCCAGGAAGAGTTTGTAGAAACGCAAAAGGCCATCCGTGAGGATGGCCTTCTGCTTAGTTTGATGCTGTGCAGTTTATGGCGGGCGTCTGCGCCGCCACCTCCGGGCGGTTGCTTCACAACGTTCAAATCCGCTCCCGGCGGATTTGTCTACTCAGGAGAGCGTTACCGACAAACAACAGATAAAACGAAAGGCCAGTCTTCCGACTGAGCCTTTCGTTTTATTTGATGCTTGGCAGTTCCCTACTCTCGCGTTAACGCTAGCATGGATGTTTTCCAGTCACGACGTTGTAAAACGACGGCCAGTCTTAAGCTCGGCTCTTCGCACTAATACGACTCACTATAGGGAGACCTCGACCGTCGATTGTCCACTGGTCAACTCAAAATTTAAAATCCCAAATTCAAAATTCGGGAAGGTGCAGAGACTCGACGGGAGCTACCCTAACGTAAAGCCGAGGGTAAAGAGAGAGTCCAATTCTCAAAGCCTGAAGTTGCTGAAGCAACAAGGCAGTAGTGAAAGCTGCGAGAGAATGAAAAATCCGTTGACCTTAAACGTCGTGTGGGTTCAAGTCCCTCCACCACCGCGCGGAAACGCAATAGCCGAAAAACAAAAACAAAAAACAAAAACAAAAACCAAAAAACAAAAACACA (N) nAAAAACAAAAACAAACGGCTATTATGCGTTACCGGCGAGACGCTACGGAAGTCTTAGAAAACTGAGCCTTGATAGAGAAATCTTTTAAGTGGAGCTCTCAAATTCAGGGAAACCTAAATCTGTTTACAGATATGGCAATCCTGAGCCAAGCCAGAAAAATTAGACTTGAGATTTGATTTGGAGTTAGACAGTGGACAATCGACGGATAACAGCATATCTAGATAAGAAGAGCATCCCATGGTCATAGCTGTTTCTGGCAGCTCTGGCCCCGTCTCAAAATCTCTGATGTTACATTGCACAAGATAAAAAATATATCATCATGAACAATAAACTGTCTGCTTACATAACAGTAATACAAGGGGTGTTATGAGCCATATTCAACGGGAAACGTGAGGCGCGGATTAAATTCCAACATGGATGCTGATTATATGGGTATAAATGGGCTCGCGATAATGTCCGGCAATCAGGTGCGCAATCTATCGCTTGTATGGGAAGCCCGATGCGCCAGAGTTGTT	5'-P-H-A-3S-E-5S-B-H'-3' (Construct A)

		TCTGAAACATGGCAAAGGTAGCGTTGCCAATGATGTTACAGAT GAGATGGTCAGACTAAACTGGCTGACGGAATTTATGCCTCTTC CGACCATCAAGCATTTTATCCGTACTCCTGATGATGCATGGTT ACTCACCCTGCGATCCCCGGAACACAGCATTCAGGTATTA GAAGAATATCCTGATTCAGGTGAAAAATTTGTTGATGCGCTGG CAGTGTTCTGCGCCGGTTGCATTCGATTCCTGTTTGTAAATTG TCCTTTTAAACAGCGATCGCGTATTTCTGCTCGCTCAGGCGCAA TCACGAATGAATAACGGTTTGGTTGATGCGAGTGATTTTGATG ACGAGCGTAATGGCTGGCTGTTGAACAAGTCTGGAAAGAAAT GCATAAACTTTTGCCATTCTCACCAGGATTCAGTCGTCATCAT GGTGATTTCTCACTTGATAACCTTATTTTGACGAGGGGAAAT TAATAGGTGTATTTGATGTTGGACGAGTCGGAATCGCAGACCG ATACCAGGATCTTGCCATCCTATGGAACGCTCGGTGAGTTT TCTCCTTCATTACAGAAACGGCTTTTCAAAAAATATGGTATTG ATAATCCTGATATGAATAAATTGCAGTTTCATTTGATGCTCGA TGAGTTTTTCTAATCAGAATTGGTTAATTGGTTGTAAACACTGG CAGAGCATTACGCTGACTTGACGGGACGGCGCAAGCTCATGAC CAAAATCCCTTAACGTGAGTTACGCGTCGTTCCACTGAGCGTC AGAC	
117	Bacterio phage T4 Td1	CCCGTAGAAAAGATCAAAGGATCTTCTTGAGATCCTTTTTTTC TGCGCGTAATCTGCTGCTTGCAAACAAAAAACACCGCTACC AGCGGTGGTTTGTGTGCGCGATCAAGAGCTACCAACTCTTTT CCGAAGGTAACTGGCTTCAGCAGAGCGCAGATACCAAATACTG TTCTTCTAGTGTAGCCGTAGTTAGGCCACCACTTCAAGAATC TGTAGCACCGCCTACATACCTCGCTCTGCTAATCCTGTTACCA GTGGCTGCTGCCAGTGGCGATAAGTCGTGTCTTACCGGGTTGG ACTCAAGACGATAGTTACCGGATAAGGCGCAGCGGTCTGGGCTG AACGGGGGGTTCTGTCACACAGCCAGCTTGAGAGCGAACGACC TACACCGAACTGAGATACCTACAGCGTGAGCTATGAGAAAGCG CCACGCTTCCCGAAGGGAGAAAGGCGGACAGGTATCCGGTAAG CGGCAGGGTCGGAACAGGAGAGCGCACGAGGGAGCTTCCAGGG GGAAACGCCTGGTATCTTTATAGTCCTGTCTGGGTTTCGCCACC TCTGACTTGAGCGTCGATTTTTGTGATGCTCGTCAGGGGGGCG GAGCCTATGGAAAAACGCCAGCAACGCGGCCTTTTACGGTTC CTGGCCTTTTGTCTGGCCTTTTGTCTACATGTTCTTCTGCGT TATCCCCGTGATTCTGTGGATAACCGTATTACCGCCTTTGAGTG AGCTGATACCGCTCGCCGACGCCAACGACCGAGCGCAGCGAG TCAGTGAGCGAGGAAGCGGAGGAGCGCCAATACGCAAACCGC CTCTCCCCGCGCGTTGGCCGATTCAATTAATGCAGCTGGCACGA CAGGTTTTCCCGACTGGAAGCGGGCAGTGAGCGCAACGCAATT AATACGCGTACCGCTAGCCAGGAAGAGTTTGTAGAAACGCAAA AAGCCATCCGTCAGGATGGCCTTCTGCTTAGTTTGATGCCTG GCAGTTTTATGGCGGGCGTCTGCCCCGCCACCTCCGGGCGTT GCTTCACAACGTTCAAAATCCGCTCCCGGCGGATTTGTCTACT CAGGAGAGCGTTACCGACAAACAACAGATAAAACGAAAGGCC CAGTCTTCCGACTGAGCCTTTCTGTTTATTTGATGCCTGGCAG TTCCCTACTCTCGCGTTAACGCTAGCATGGATGTTTCCAGT CACGACGTTGTAAAACGACGGCCAGTCTTAAGCTCGGCTCTTC GACTAATACGACTCACTATAGGGAGACCTCGACCGTCGATT GTCCACTGGTCGGGAGACCTCGAATGGAATTGGTTCTACATA AATGCCTAACGACTATCCCTTTGGGGAGTAGGGTCAAGTGACT CGAAACGATAGACAACCTTGCTTTAACAAGTTGGAGATATAGTC TGCTCTGCATGGTGACATGCAGCTGGATATAAATCCGGGGTAA GATTAACGACCTTATCTGAACATAATGCTAAAAATCCGTTGACC TTAAACGGTCGTGTGGGTTCAAGTCCCTCCACCCACGCGG AAACGCAATAGCCGAAAAACAAAAACAAAAACAAAAACAA AAACCAAAAAACAAAAACACA (N) nAAAAACAAAAACAAA CGGCTATTATGCGTTACCGGCGAGACGCTACGGAGGTTAATTG AGGCCTGAGTATAAGGTGACTTATACTTGTAATCTATCTAAAC GGGGAACCTCTCTAGTAGACAATCCCGTGCTAAATTGTAGGAC TAATTCATTTATCAGATTTCTAGGACCAGTGGACAATCGACG	5'-P-H-A-3S- E-5S-B-H'-3' (Construct A)

		GATAACAGCATATCTAGATAAGAAGAGCATCCCATGGTCATAG CTGTTTCCCTGGCAGCTCTGGCCCGTGTCTCAAAATCTCTGATG TTACATTGCACAAGATAAAAAATATATCATCATGAACAATAAAA CTGTCTGCTTACATAAACAGTAATACAAGGGGTGTTATGAGCC ATATTTCAACGGGAAACGTCGAGGCCGCGATTAAATTTCCAACAT GGATGCTGATTTATATGGGTATAAAATGGGCTCGCGATAATGTC GGGCAATCAGGTGCGACAATCTATCGCTTGTATGGGAAGCCCCG ATGCGCCAGAGTTGTTTCTGAAACATGGCAAAGGTAGCGTTGC CAATGATGTTACAGATGAGATGGTCAGACTAAACTGGCTGACG GAATTTATGCCTCTTCCGACCATCAAGCATTTTATCCGTACTC CTGATGATGCATGGTTACTCACCCTGCGATCCCCGGAAGAAC AGCATTTCCAGGTATTAGAAGAATATCCTGATTCAAGTGAAAAAT ATTGTTGATGCGCTGGCAGTGTTCCTGCGCCGGTTGCATTCTGA TTCTGTGTTGTAATTGTCCTTTTAACAGCGATCGCGTATTTTCG TCTCGCTCAGGCGCAATCACGAATGAATAACGGTTTGGTTGAT GCGAGTGATTTTGATGACGAGCGTAATGGCTGGCCTGTTGAAC AAGTCTGGAAAGAAATGCATAAACTTTTGCCATTCTCACCGBA TTCAGTCTGCTCACTCATGGTGATTTCTCACTTGATAACCTTATT TTTGACGAGGGGAAATTAATAGGTGTATTGATGTTGGACGAG TCGGAATCGCAGACCGATACCAGGATCTTGCCATCCTATGGAA CTGCCTCGGTGAGTTTTCTCCTTCATTACAGAAACGGCTTTTTT CAAAAATATGGTATTGATAATCCTGATATGAATAAATTGCAGT TTCATTTGATGCTCGATGAGTTTTTCTAATCAGAATTGGTTAA TTGGTTGTAAACACTGGCAGAGCATTACGCTGACTTGACGGGAC GGCGCAAGCTCATGACCAAAATCCCTTAACGTGAGTTACGCGT CGTTCCACTGAGCGTCAGAC	
118	Bacterio phage T4 Td2	CCCGTAGAAAAGATCAAAGGATCTTCTTGAGATCCCTTTTTTTC TGCGCGTAATCTGCTGCTTGCAAACAAAAAACACCCTGCTACC AGCGGTGGTTTGTGTTGCCGGATCAAGAGCTACCAACTCTTTTT CCGAAGGTAACTGGCTTCAGCAGAGCGCAGATACCAATACTG TTCTTCTAGTGTAGCCGTAGTTAGGCCACCACTTCAAGAACTC TGTAGCACCGCCTACATACCTCGCTCTGCTAATCCTGTTACCA GTGGCTGCTGCCAGTGGCGATAAGTCGTGTCTTACCGGGTTGG ACTCAAGACGATAGTTACCAGGATAAGGCGCAGCGGTGCGGGCTG AACGGGGGGTTTCGTGCACACAGCCAGCTTGAGCGAACGACG TACACCGAACTGAGATACCTACAGCGTGAGCTATGAGAAAGCG CCACGCTTCCCGAAGGGAGAAAGGCGGACAGGTATCCGGTAAG CGGCAGGGTCGGAACAGGAGAGCGCACGAGGGAGCTTCCAGGG GGAAACGCCTGGTATCTTTATAGTCTGTGCGGGTTTCGCCACC TCTGACTTGAGCGTTCGATTTTTGTGATGCTCGTCAGGGGGGCG GAGCCTATGGAACAAACGCCAGCAACGCGGCCCTTTTTACGGTTT CTGGCCCTTTTGCTGGCCCTTTTGCTCACATGTTCTTTCTGCGT TATCCCTGATTCTGTGGATAACCGTATTACCGCCTTTGAGTG AGCTGATACCGCTCGCCGAGCCGAACGACCGAGCGCAGCGAG TCAGTGAGCGAGGAAGCGGAGGAGCGCCCAATACGCAAAACCGC CTCTCCCCGCGCGTTGGCCGATTCAATTAATGCAGCTGGCACGA CAGGTTTCCCGACTGGAAGCGGGCAGTGAGCGCAACGCAATT AATACGCGTACCGCTAGCCAGGAAGAGTTTGTAGAAACGCAAA AAGGCCATCCGTGAGGATGGCCTTCTGCTTAGTTTGTATGCTG GCAGTTTATGGCGGGCGTCTGCCCCGCCACCTCCGGGCCGTT GCTTACAAACGTTCAAATCCGCTCCCGGCGGATTTGTCTTACT CAGGAGAGCGTTTACCGACAAACAACAGATAAAACGAAAGGCC CAGTCTTCCGACTGAGCCTTTTCGTTTTATTTGATGCTTGGCAG TTCCCTACTCTCGCGTTAACGCTAGCATGGATGTTTTCCAGT CACGACGTTGTAAAACGACGGCCAGTCTTAAGCTCGGCTCTTC GCACTAATACGACTCACTATAGGGAGACCCTCGACCGTCGATT GTCCACTGGTCGGGAGACCCTCGAATGGAATTGGTTCTACATA AATGCCTAACGACTATCCCTTTGGGGAGTAGGGTCAAGTGACT CGAAACGATAGACAACCTTGCTTTAAACAAGTTGGAGATATAGTC TGCTCTGCATGGTGACATGCAGCTGGATATAATTCCGGGGTAA GATTAACGACCTTATCTGAACATAATGCTACCGTTTAATATTG	5'-P-H-A-3S- X-E-X'-5S-B- H'-3' (Construct B)

		CGTCAGGTAGTAAACTACTAACTACAACCTGCTGAAGCA (N) n CAGAGATGTTTTCTTGGGTAAATTGAGGCCTGAGTATAAGGTG ACTTATACTTGTAATCTATCTAAACGGGGAACCTCTCTAGTAG ACAATCCCGTGCTAAATTTGTAGGACTAATTCATTTATCAGAT TTCTAGGACCAGTGGACAATCGACGGATAACAGCATATCTAGA TAAGAAGAGCATCCCATGGTCATAGCTGTTTCTTGGCAGCTCT GGCCCGTGCTCAAAATCTCTGATGTTACATTGCACAAGATAA AAATATATCATCATGAACAATAAACTGTCTGCTTACATAAAC AGTAATACAAGGGGTGTTATGAGCCATATTCAACGGGAAACGT CGAGGCCGCGATTAAATTCACACATGGATGCTGATTTATATGG GTATAAATGGGCTCGCGATAATGTCGGGCAATCAGGTGCGACA ATCTATCGCTTGATATGGGAAGCCCGATGCGCCAGAGTTGTTTT TGAAACATGGCAAAGGTAGCGTTGCCAATGATGTTACAGATGA GATGGTCAGACTAAACTGGCTGACGGAATTTATGCCTCTTCCG ACCATCAAGCATTTTATCCGTACTCTGATGATGCATGGTTAC TCACCACTGCGATCCCCGGAACAGCATTCCAGGTATTAGA AGAATATCCTGATTGAGGTGAAAAATATTGTTGATGCGCTGGCA GTGTTCTGCGCCGGTTGCATTGATTCCTGTTTTGTAATTGTC CTTTTAACAGCGATCGCGTATTTCTGCTCGCTCAGGCGCAATC ACGAATGAATAACGGTTTGGTTGATGCGAGTGATTTGATGAC GAGCGTAATGGCTGGCTGTTGAACAAGTCTGGAAGAAATGC ATAAACTTTTGCCATTTCTACCGGATTGAGTCGTCACCTCATGG TGATTTCTCACTTGATAACCTTATTTTGACGAGGGGAAATTA ATAGGTTGTATTGATGTTGGACGAGTCGGAATCGCAGACCGAT ACCAGGATCTTGCCATCCTATGGAACGCTCGGTGAGTTTTTC TCCTTCATTACAGAAACGGCTTTTTCAAAAATATGGTATTGAT AATCCTGATATGAATAAATTGCAGTTTCATTTGATGCTCGATG AGTTTTTCTAATCAGAAATGGTTAATTGGTTGTAACACTGGCA GAGCATTACGCTGACTTGACGGGACGGCGCAAGCTCATGACCA AAATCCCTTAACGTGAGTTACGCGTCGTTCCACTGAGCGTCAG AC	
119	Anabaena PCC7120 2	TAATACGACTCACTATAGGGAGACCTCGACCGTCGATTGTCC ACTGGTCAACAATAGATGACTTACAATAATCGGAAGGTGCAG AGACTCGACGGGAGCTACCCTAACGTCAAGACGAGGGTAAAGA GAGAGTCCAATTCTCAAAGCCAATAGGCAGTAGCGAAAGCTGC AAGAGAATGAAAATCCGTTGACCTTAAACGGTCGTGTGGGTTT AAGTCCCTCCACCCCCACGCCGGAACGCAATAGCCGAAAAAC AAAAAACAAAAAAACAAAAAAACCAAAAAACAAAACAC A (N) nAAAAAACAAAAAACAAACGGCTATTATGCGTTACCGG CGAGACGCTACGGACTT	5'-P-A-3S-E- 3' (Construct C)
120	Azoarcus 2	TAATACGACTCACTATAGGGAGACCTCGACCGTCGATTGTCC ACTGGTCtttGctCCTTGCGCCGGGAAACACGCAAGGGATGG TGTCAAATTCGGCGAAACCTAAGCGCCCGCCGGGCGTATGGC AACGCCGAGCCAAGCTTCGGCGCCTGCGCCGATGAAGGTGTAG AGACTAGACGGCACCCACCTAAGGCAAACGCTATGGTGAAGGC ATAGTCCAGGGAGTGGCGAAAGTCACACAAACCGGAAAAATCCG TTGACCTTAAACGGTCGTGTGGGTTCAAGTCCCTCCACCCCA CGCCGGAACGCAATAGCCGAAAAACAAAAACAAAAAAACA AAAAAAACCAAAAAACAAAACACA (N) nAAAAAACAAAA ACAAAACGGCTATTATGCGTTACCGGCGAGACGCTACGGAagT	5'-P-A-3S-E- 3' (Construct C)
121	Anabaena azollae 2	TAATACGACTCACTATAGGGAGACCTCGACCGTCGATTGTCC ACTGGTCAACTCAAAATTTAAATCCCAATTCAAAATTCGGG GAAGGTGCAGAGACTCGACGGGAGCTACCCTAACGTAAAGCCG AGGGTAAAGAGAGAGTCCAATTCTCAAAGCCTGAAGTTGCTGA AGCAACAAGGCAGTAGTGAAAGCTGCGAGAGAATGAAAATCCG TTGACCTTAAACGGTCGTGTGGGTTCAAGTCCCTCCACCCCA CGCCGGAACGCAATAGCCGAAAAACAAAAACAAAAAAACA AAAAAAACCAAAAAACAAAACACA (N) nAAAAAACAAAA ACAAAACGGCTATTATGCGTTACCGGCGAGACGCTACGGACTT	5'-P-A-3S-E- 3' (Construct C)

122	Bacterio phage T4 Td3	TAATACGACTCACTATAGGGAGACCCTCGACCGTCGATTGTCC ACTGGTCGGGAGACCCTCGAATGGAATTGGTTCTACATAAATG CCTAACGACTATCCCTTTGGGGAGTAGGGTCAAGTGACTCGAA ACGATAGACAACTTGCTTTAACAAGTTGGAGATATAGTCTGCT CTGCATGGTGACATGCAGCTGGATATAATTCCGGGGTAAGATT AACGACCTTATCTGAACATAATGCTAAAATCCGTTGACCTTAA ACGGTCGTGTGGGTTCAAGTCCCTCCACCCCCACGCCGAAAC GCAATAGCCGAAAAACAAAAACAAAAACAAAAACAAAAAC CAAAAAACAAAAACACA (N) nAAAAACAAAAACAAAAACGGC TATTATGCGTTACCGGCGAGACGCTACGGAGGT	5'-P-A-3S-E- 3' (Construct C)
-----	-----------------------------	---	-------------------------------------

[0310] In some aspects, the genetic constructs disclosed herein comprise a nucleotide sequence having at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, at least about 99%, or 100% sequence identity with SEQ ID NOs: 123-132, wherein "N" represents any one of A, T (or U), C, or G; and wherein "n" represents any interger. In some aspects, "(N)n" comprises a nucleotide sequence of an exonic sequences. In some aspects, "(N)n" comprises a nucleotide sequence of an exonic sequences and a nucleotide sequence of an IRES.

[0311] In some aspects, an RNA transcript comprising a nucleotide sequence having at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, at least about 99%, or 100% sequence identity with SEQ ID NOs: 123-132, wherein "N" represents any one of A, T (or U), C, or G; and wherein "n" represents any interger, is transcribed from the DNA genetic constructs disclosed herein. In some aspects, "(N)n" comprises a nucleotide sequence of an exonic sequences. In some aspects, "(N)n" comprises a nucleotide sequence of an exonic sequences and a nucleotide sequence of an IRES.

[0312] Non-limiting examples of constructs of the present and of RNA trasncripts transcribed from the DNA genetic constructs disclosed herein are shown in **Table 14**.

Table 14. Exemplary genetic constructs and exemplary RNA trasncripts transcribed from the DNA genetic constructs disclosed herein ((N)n is a location for an exonic sequence)

SEQ ID NO	DESCRIPT ION	SEQUENCE	FORMULA
123	Anabaena PCC7120	GGAGACCCUCGACCGUCGAUUGUCCACUGGUCAACAAUAGAUGAC UUACAACUAAUCGGAAGGUGCAGAGACUCGACGGGAGCUACCCUA ACGUCAAGACGAGGGUAAAGAGAGAGUCCAAUUCUCAAAGCCAAU AGGCAGUAGCGAAAGCUGCAAGAGAAUGAAAAUCCGUUGACCUUA AACGGUCGUGUGGUUCAAGUCCUCCACCCCCACGCCGAAACG CAAUAGCCGAAAAACAAAAACAAAAACAAAAACAAAAACCAA AAAAACAAAAACACA (N) nAAAAACAAAAACAAAACGGCUAUUA UGCGUUAACGGCGAGACGCUACGGACUAAAUAUUGAGCCUUA AGAAGAAAUUCUUUAAGUGGAUGCUCUCAAACUCAGGGAAACCUA AAUCUAGUUAUAGACAAGGCAAUCCUGAGCCAAGCCGAAGUAGUA AUUAGUAAGACCAGUGGACAAUCGACGGAUAACAGCAUAUCUAGA UA	5'-P-H-A- 3S-E-5S-B- H'-3' (Construct A)

124	Tetrahymena thermophila	GGAGuCCCUCGACCGUCGAUUGUCCACUGGUCAUUUACCUUUGGA GGGAAAAGUUUAUCAGGCAUGCACCGUGUAGCUAGUCUUUAAACCA AUAGAUUGCAUCGGUUUAAAAGGCAAGACCGUCAAAUUGCGGGAA AGGGGUCACAGCCGUUCAGUACCAAGUCUCAGGGGAAACUUUGA GAUGGCCUUGCAAAGGGUAUGGUAAUAGCUGACGGACAUGGUCC UAACCACGCAGCCAAGUCCUAAGUCAACAGAUUCUUGUUGAUU GGAUGCAGUUCACAGACUAAAUGUCGGUCGGGGAAGAUUAUUCU UCUCAUAAGAUUAGUCGGACCUCUCCUUAUUGGGAGCUAGCGGA UGAAGUGAUGCAACACUGGAGCCGCUGGGAACUAAUUGUAUGCG AAAGUAUAUUGAUUAGUUUUGGAGUACUCGUAAGGUAAAAUCCG UUGACCUUAAACGGUCGUGUGGGUUAAGUCCUCCACCCCCACG CCGGAACCGCAAUAGCCGAAAAACAAAAACAAAAACAAAAA AAAAACAAAAACAAAAACACA (N) nAAAAACAAAAACAAAA CGGCUAUUAUGCGUUAACCGGCGAGACGCUACGGACCCUCUAAUA GCAAUGACCAGUGGACAAUCGACGGAUAACAGCAUAUCUAGAU	5'-P-H-A- 3S-E-5S-B- H'-3' (Construct A)
125	Azoarcus	GGAGACCCUCGACCGUCGAUUGUCCACUGGUUuuuGcuCCUUGCG CCGGGAAACCACGCAAGGGAUGGUGUCAAAUUCGGCGAAACCUAA GCGCCCGCCGGGCGUAUGGCAACGCCGAGCCAAGCUUCGGCGCC UGCGCCGAUGAAGGUGUAGAGACUAGACGGCACCCACCUAAGGCA AACGCUAUGGUGAAGGCAUAGUCCAGGGAGUGGCGAAAGUCACAC AAACCGGAAAAUCCGUUGACCUUAAACGGUCGUGUGGGUUAAGU CCCUCACCCCCACGCCGAAACGCAAUAGCCGAAAAACAAAAA CAAAAAACAAAAACAAAAACAAAAACAAAAACACA (N) nAAA AAACAAAAACAAAAACGGCUAUUAUGCGUUAACCGGCGAGACGCUA CGGAagUaaaGACCAGUGGACAAUCGACGGAUAACAGCAUAUCUA GAUA	5'-P-H-A- 3S-E-5S-B- H'-3' (Construct A)
126	Anabaena azollae	GGAGACCCUCGACCGUCGAUUGUCCACUGGUCAACUAAAAUUUA AAAUCCCAAUUCAAAAUCCGGGAAGGUGCAGAGACUCGACGGG AGCUACCCUAACGUAAAGCCGAGGGUAAAGAGAGAGUCCAAUUCU CAAAGCCUGAAGUUGCUGAAGCAACAAAGGCAUAGUGAAAGCUGC GAGAGAAUGAAAAUCCGUUGACCUUAAACGGUCGUGUGGGUUCAA GUCCCUCACCCCCACGCCGAAACGCAAUAGCCGAAAAACAAAA AACAAAAACAAAAACAAAAACAAAAACAAAAACACA (N) nA AAAAACAAAAACAAAAACGGCUAUUAUGCGUUAACCGGCGAGACGC UACGGACUUAAGAAACUGAGCCUUGAUAGAGAAUUCUUUAAGUG GAAGCUCUCAAUUCAGGGAACCUAAUUCUGUUUACAGAUUUGG CAAUCCUGAGCCAAGCCAGAAAAUAGACUUGAGAUUUGAUUUU GGAGUUAGACCAGUGGACAAUCGACGGAUAACAGCAUAUCUAGAU A	5'-P-H-A- 3S-E-5S-B- H'-3' (Construct A)
127	Bacteriophage T4 Td1	GGAGACCCUCGACCGUCGAUUGUCCACUGGUCGGGAGACCCUCGA AUGGAAUUGGUUCUACAUAUAAUGCCUAACGACUAUCCCUUUGGG AGUAGGGUCAAGUGACUCGAAACGAUAGACAACUUGCUUUAACAA GUUGGAGAUUAGUCUGCUCUGCAUGGUGACAUGCAGCUGGAUUA AAUUCGGGGUAAGAUUAACGACCUUAUCUGAACAUAAUGCUAAA AUCCGUUGACCUUAAACGGUCGUGUGGGUUAAGUCCUCCACCC CCACGCCGGAAACGCAAUAGCCGAAAAACAAAAACAAAAAAC AAAAACAAAAACAAAAACAAAAACACA (N) nAAAAACAAAAA CAAAACGGCUAUUAUGCGUUAACCGGCGAGACGCUACGGAGGUUA UUGAGGCCUGAGUAUAAGGUGACUUAUACUUGUAAUCUAUCUAAA CGGGGAACCUCUCUAGUAGACAACUCCGUGCUAAUUGUAGGACU AAUUCUUAUUAUCAGAUUUUCUAGGACCAGUGGACAAUCGACGGAU AACAGCAUAUCUAGAU	5'-P-H-A- 3S-E-5S-B- H'-3' (Construct A)
128	Bacteriophage T4 Td2	GGAGACCCUCGACCGUCGAUUGUCCACUGGUCGGGAGACCCUCGA AUGGAAUUGGUUCUACAUAUAAUGCCUAACGACUAUCCCUUUGGG AGUAGGGUCAAGUGACUCGAAACGAUAGACAACUUGCUUUAACAA GUUGGAGAUUAGUCUGCUCUGCAUGGUGACAUGCAGCUGGAUUA AAUUCGGGGUAAGAUUAACGACCUUAUCUGAACAUAAUGCUACC GUUUAUAUUGCGUCAGGUAGUAAACUACUAAACUACAACCGUCUG AAGCA (N) nCAGAGAUGUUUUCUUGGGUUAUUGAGGCCUGAGUA UAAGGUGACUUAUACUUGUAAUUCUAUCUAAACGGGGAACCUCUCU	5'-P-H-A- 3S-X-E-X'- 5S-B-H'-3' (Construct B)

		AGUAGACAAUCCCGUGCUAAAUUGUAGGACUAAUUCUUAUUAUCA GAUUUCUAGGACCAGUGGACAAUCGACGGAUAAACAGCAUAUCUAG	
129	Anabaena PCC7120 2	GGAGACCCUCGACCGUCGAUUGUCCACUGGUCAACAAUAGAUGAC UUACAACUAAUCGGAAGGUGCAGAGACUCGACGGGAGCUACCCUA ACGUCAAGACGAGGGUAAAAGAGAGAGUCCAAUUCUCAAAGCCAAU AGGCAGUAGCGAAAGCUGCAAGAGAAUGAAAAUCCGUUGACCUUA AACGGUCGUGUGGGUUCAGUCCUCCACCCACGCCGGAACG CAAUAGCCGAAAAACAAAAACAAAAACAAAAACAAAAACCAA AAAAACAAACACA (N) nAAAAAACAAAAACAAACGGCUAUUA UGCGUUAACGGCGAGACGCUACGGACUU	5'-P-A-3S- E-3' (Construct C)
130	Azoarcus 2	GGAGACCCUCGACCGUCGAUUGUCCACUGGUUuuuGcuCCUUGCG CCGGGAAACACGCAAGGGAUGGUGUCAAUUCGGCGAAACCUAA GCGCCCGCCGGGCGUAUGGCAACGCGGAGCCAAGCUUCGGCGCC UGCGCCGAUGAAGGUGUAGAGACUAGACGGCACCCACCUAAGGCA AACGCUAUGGUGAAGGCAUAGUCCAGGGAGUGGCGAAAGUCACAC AAACCGGAAAAUCCGUUGACCUUAAACGGUCGUGUGGGUUAAGU CCCUCCACCCCCACGCCGGAACGCAUAGCCGAAAAACAAAAA CAAAAAACAAAAACAAAAACAAAAACAAAAACACA (N) nAAA AAACAAAAACAAAAACGGCUAUUAUGCGUUAACGGCGAGACGCUA CGGAagU	5'-P-A-3S- E-3' (Construct C)
131	Anabaena azollae 2	GGAGACCCUCGACCGUCGAUUGUCCACUGGUCAACUAAAAUUUA AAAUCCCAAUUCAAUUUCCGGGAAGGUGCAGAGACUCGACGGG AGCUACCCUAACGUAAAGCCGAGGGUAAAGAGAGAGUCCAAUUCU CAAAGCCUGAAGUUGCUAAGCAACCAAGGCAGUAGUGAAAGCUGC GAGAGAAUGAAAAUCCGUUGACCUUAAACGGUCGUGUGGGUUCAA GUCCCUCCACCCCCACGCCGGAACGCAUAGCCGAAAAACAAAA AACAAAAACAAAAACAAAAACAAAAACAAAAACACA (N) nA AAAAACAAAAACAAAAACGGCUAUUAUGCGUUAACGGCGAGACGC UACGGACUU	5'-P-A-3S- E-3' (Construct C)
132	Bacterio phage T4 Td3	GGAGACCCUCGACCGUCGAUUGUCCACUGGUCCGGAGACCCUCGA AUGGAAUUGGUUCUACAUAUAAUGCCUACGACUUAUCCCUUUGGGG AGUAGGGUCAAGUGACUCGAAACGAUAGACAACUUGCUUUAACAA GUUGGAGAUUAGUCUGCUCUGCAUGGUGACAUGCAGCUGGAUUA AAUUCGCGGUAAGAUAACGACCUUAUCUGAACUAUUAUGCUAAA AUCCGUUGACCUUAAACGGUCGUGUGGGUUAAGUCCCUCCACCC CCACGCCGGAACGCAUAGCCGAAAAACAAAAACAAAAACAAAAAC AAAAACAAAAACAAAAACAAAAACACA (N) nAAAAACAAAAAC CAAAACGGCUAUUAUGCGUUAACGGCGAGACGCUACGGAGGU	5'-P-A-3S- E-3' (Construct C)
181	Genetic Construct D	TAATACGACTCACTATAGAACAAATAGATGACTTACAATAATCGG AAGGTGCAGAGACTCGACGGGAGCTACCCCTAACGTCAGACGAGG GTAAAGAGAGAGTCCAATTCTCAAAGCCAATAGGCAGTAGCGAAA GCTGCAAGAGAAATGAAAATCCGTTGACCTTAAACGGTCTGTGGG TTCAAGTCCCTCCACCCCANNNNNNNNNNNNNNNNNNNCGCCGG AAACGCAATAGCCGAAAAACAAAAACAAAAACAAAAACAAAAA ACCAAAAAACAAAACACATTAAACAGCCTGTGGGTTGATCCCA CCCACAGGCCCATTTGGGCGCTAGCACTCTGGTATCACGGTACCTT TGTGCGCCTGTTTATACCCCTCCCCAACTGTAACCTAGAAAT AACACACACCGATCAACAGTCAGCGTGGCACACCAGCCACGTTTTT GATCAAGCACTTCTGTTACCCCGACTGAGTATCAATAGACTGCT CACGCGGTTGAAGGAGAAAGCGTTTCGTTATCCGGCCAACTACTTC GAAAAACCTAGTAACACCGTGGAAGTTGCAGAGTGTTCGCTCAG CACTACCCAGTGTAGATCAGGTCGATGAGTCACCGCATTCCCA CGGGCGACCGTGGCGGTGGCTGCGTTGGCGGCTGCCATGGGGA AACCCATGGGACGCTCTAATACAGACATGGTGCGAAGAGTCTATT GAGCTAGTTGGTAGTCCCTCCGGCCCCGTAATGCGGCTAATCCTAA CTGCGGAGCACACACCTCAAGCCAGAGGGCAGTGTGTGTAACG GGCAACTCTGCAGCGGAACCGACTACTTTGGGTGTCCGTGTTTCA TTTTATTCTATACTGGCTGCTTATGGTGACAATTGAGAGATCGT TACCATATAGCTATTGGATTGGCCATCCGGTGACTAATAGAGCTA TTATATATCCCTTTGTTGGGTTTATACCACTTAGCTTGAAAGAGG TTAAACATTACAATTTCATGTTAAGTTGAATACAGCAAAatggt	

		gagcaagggcgaggagctgttcaccgggggtggtgcccacctcctggtcgagctggacggcgacgtaaacggccacaagttcagcgtgtccggcgagggcgagggcgatgccacctacggcaagctgacctgaagttcatctgcaccaccggcaagctgcccgtgcccctggcccaccctcgtgaccacctgacctacggcgtgcagtgttcagccgctaccccgaacacatgaagcagcacgacttcttcaagtcggccatgcccgaaggctacgtccaggagcgacccatcttcttcaaggacgacggcaactacaagaccgcgcgaggtgaagttcgagggcgacacctgggtgaacgcgcatcgagctgaaggcgatcgacttcaaggaggacggcaacatcctggggcacaagctggagtacaactacaacagccacaacgtctatatcatggccgacaagcagaagaacggcatcaaggtgaacttcaagatccggccacaacatcgaggacggcagcgtgcagctcgccgaccactaccagcagaacaccccccatcggcgacggccccgtgctgctgccgacaaccactacctgagcaccagtcggccctgagcaaagaccccaacgagaagcgcgatcacatggtcctgctggagttcgtgacggcgccgggatcactctcggcgatggacgagctgtacaagtaaAAAAACAAAAACAAAACNNNNNNNNNNNNNNNNNNNGGCTATTATGCGTTACCGGCGAGACGCTACGGACTT	
182	Genetic Construct E DNA	TAATACGACTCACTATAGGAGCCTTAAAGAAGAAATCTTTAAGTGGATGCTCTCAAACTCAGGGAACCTAAATCTAGTTATAGACAAGGCAATCCTGAGCCAAGCCGAAGTAGTAATTAGTAAAACAATAGATGACTTACAATAATCGGAAGGTGCAGAGACTCGACGGGAGCTACCTAACGTCGAAGACGAGGGTAAAGAGAGAGTCCAATTCTCAAAGCCAATAGGCAGTAGCGAAAAGCTGCAAGAGAATGAAAATCCGTTGACCTTAAACGGTTCGTGTGGGTTCAAGTCCCTCCACCCCCACGCCGGAAACGCAATAGCCGAAAAACAAAAACAAAAACAAAAACAAAAACAAAAACAAAAACACATTAATAACAGCCTGTGGGTTGATCCCACCACAGGCCCATTTGGGCGCTAGCACTCTGGTATCACGGTACCTTTGTGCGCCTGTTTTATACCCCTCCCCAACTGTAACCTAGAAGTAAACACACCGATCAACAGTCAGCGTGGCACACCAGCCACGTTTTGTATCAAGCACTTCTGTTACCCCGACTGAGTATCAATAGACTGCTCA CGCGTTGAAGGAGAAAGCGTTCGTTATCCGGCCAACTACTTTCGA AAAACCTAGTAACACCGTGAAGTTGCAGAGTGTTCGCTCAGCACTACCCAGTGTAGATCAGGTCGATGAGTCACCGCATTCCCCACGGCGACCGTGGCGGTGGCTGCGTTGGCGGCCTGCCATGGGGAAA CCCATGGGACGCTCTAATACAGACATGGTGCGAAGAGTCTATTGAGCTAGTTGGTAGTCCTCCGGCCCCGAATGCGGCTAATCCTAACTGCGGAGCACACACCTCAAGCCAGAGGGCAGTGTGTCGTAACGGGCAACTCTGCAGCGGAACCGACTACTTTGGGTGTCCGTGTTTCATT TTATTCCTATACTGGCTGCTTATGGTGACAATTGAGAGATCGTTA CCATATAGACTATTGGATTGGCCATCCGGTGACTAATAGAGCTATT ATATATCCCTTTGTTGGGTTTATACCACTTAGCTTGAAAGAGGTT AAAACATTACAATTCATTGTTAAGTTGAATACAGCAAAatggtgagcaagggcgaggagctgttcaccgggggtggtgcccacctcctggtcgagctggacggcgacgtaaacggccacaagttcagcgtgtccggcgagggcgagggcgatgccacctacggcaagctgacctgaagttcatctgcaccaccggcaagctgcccgtgcccctggcccaccctcgtgacacctgacctacggcgtgcagtgttcagccgctaccccgaccacatgaagcagcacgacttcttcaagtcggccatgcccgaaggctacgtccaggagcgacccatcttcttcaaggacgacggcaactacaagaccgcgcgaggtgaagttcgagggcgacacctgggtgaaccgcacgagctgaaggcgatcgacttcaaggaggacggcaacatctggggcacaagctggagtacaactacaacagccacaacgtctata tcatggccgacaagcagaagaacggcatcaaggtgaacttcaagatccggccacaacatcgaggacggcagcgtgcagctcgccgaccactaccagcagaacaccccccatcggcgacggccccgtgctgctgcccgaacaccactacctgagcaccagtcggccctgagcaaagaccccaacgagaagcgcgatcacatggtcctgctggagttcgtgacggccgcgggatcactctcggcgatggacgagctgtacaagtaaAAAAACAAAAACAAAACGGCTATTATGCGTTACCGGCGAGACGCTACGGACTT	

183	Genetic Construc t E RNA	UAAUACGACUCACUAUAGGAGCCUUAAGAAGAAAUCUUUAAGU GGAUGCUCUCAAAACUCAGGGAAACCUAAAUCUAGUUUAUAGACAAG GCAAUCCUGAGCCAAGCCGAAGUAGUAAUAGUAAAACAAUAGAU GACUUACAACUAUUCGGAAGGUGCAGAGACUCGACGGGAGCUACC CUAACGUCAAAGACGAGGGUAAAGAGAGAGUCCAAUUCUCAAGGCC AAUAGGCAGUAGCGAAAAGCUGCAAGAGAAUGAAAAUCCGUUGACC UUAAACGGUCGUGUGGGUUCAAGUCCCUCCACCCCCACGCCGGAA ACGCAAUAGCCGAAAAACAAAAACAAAAACAAAAACAAAAAC CAAAAAACAAAAACACAUAUAAACAGCCUGUGGGUUGAUCCACC CACAGGCCCAUUGGGCGCUAGCACUCUGGUUAUCACGGUACCUUUG UGCGCCUGUUUUUAUACCCCUCCCCAACUGUAACUUAGAAGUAA CACACACCGAUCAACAGUCAGCGUGGCACACCAGCCACGUUUUGA UCAAGCACUUCUGUUACCCCGGACUGAGUAUCAAUAGACUGCUCA CGCGGUUGAAGGAGAAAGCGUUCGUUAUCCGGCCAACUACUUCGA AAAACCUAGUAAACACCGUGGAAGUUGCAGAGUGUUUCGUCAGCA CUACCCACAGUUGAUAUAGGUCGAUGAGUACCCGCAUUCCCCACG GGCGACCUGGGCGGUGGCGUGGUGGCGGCCUGCCAUUGGGGAAA CCCAUGGGACGCUCUAAUACAGACAUGGUGCGAAGAGUUAUUGA GCUAGUUGGUAGUCCUCCGGCCCCUGAAUGCGGCUAAUCCUAACU GCGGAGCACACCCCUAAGCCAGAGGGCAGUGUGUCGUAACGGG CAACUCUGCAGCGGAACCGACUACUUUGGGUGUCCGUGUUUCAUU UUAUUCUUAUACUGGCUGCUUAUGGUGACAAUUGAGAGAUUGUUA CCAUAUAGCUAUUGGAUUGGCCAUCCGGUGACUAAUAGAGCUAUU AUAUAUCCCUUUGUUGGGUUUAUACCACUAGCUUGAAAGAGUU AAAACAUUAUCAAUUAUUGUUAAGUUGAAUACAGCAAAaugguga gcaagggcgaggagcguuacccgggugggugcccauccuggucg agcuggacggcgacguaaacggccacaaguucagcguguccggcg agggcgagggcgauGCCaccuacggcaagcugaccuagaaguca ucugcaccaccggcaagcugcccugcccugggcccaccucguga ccaccugaccuacggcgugcagugcuucagccgcuaccccgacc acaugaagcagcagcguucuucaaguccgccaugcccgaaggcu acguccaggagcgacccaucuuucaaaggacgacggcaacuaca agaccccgcccgaggugaaguucgagggcgacaccucggugaacc gcaucgagcugaagggcaucgacucaaaggaggacggcaacaucc uggggcacaagcuggaguacaacuacaacagccacaacgucuaa ucaugggcgacaagcagaagaacggcaucaaggugaacucaaaga uccggccacaacaucgaggacggcagcgugcagcucgcccgaccacu accagcagaacacccccaucggcgacggccccgugcugcugcccg acaaccacuaccugagcaccaguccgcccugagcaaaagacccca acgagaagcgcgaucaaugguccugcuggaguucgugacccgccc ccgggaucacucucggcauggacgagcuguaacaaguuaAAAAAC AAAAAACAAAACGGCUAUUAUGCGUUAACGGCGAGACGCUACGGA CUU	
-----	--------------------------------	---	--

III. Vectors

[0313] In some aspects, the genetic DNA or RNA genetic constructs are comprised in a vector. In some aspects, the vector is a DNA vector. In some aspects, the vector is a RNA vector. In some aspects, the vector is a DNA/RNA hybrid vector. In some aspects, the vector is a plasmid.

[0314] In some aspects, the vector is produced by any of the method known in the art for the production of vectors comprising genetic constructs. In some aspects, the vector is produced by chemical synthesis. In some aspects, the vector is produced by molecular cloning. In some aspects, the vector comprises naturally occurring nucleobases, naturally occurring sugar moieties, naturally occurring phosphates groups, naturally occurring nucleic acids backbones, naturally

occurring nucleosides, naturally occurring internucleoside linkages, naturally occurring nucleotides, modified nucleobases, modified sugar moieties, modified phosphates groups, modified nucleic acids backbones, modified nucleosides, modified internucleoside linkages, modified nucleotides, or any combination thereof. In some aspects, the genetic construct or the vector is suitable for *in vivo* transcription. In some aspects, the vector is suitable for *in vitro* transcription. In some aspects, the vector is suitable for *in vivo* and *in vitro* transcription.

[0315] In some aspects, the genetic construct (e.g., comprised in a vector) described herein comprises one or more additional components that aid in the function of the DNA genetic construct, of the pre-circRNA, or of the circRNA produced as disclosed herein.

[0316] In some aspects, the genetic constructs of the present disclosure comprise one or more additional components, wherein the one or more additional components, for example, promotes the transcription of the pre-circRNAs from the DNA genetic constructs, promotes circularization of the pre-circRNAs, enhances the expression of an encoded payload, increases the stability of the payload sequence, or any combination thereof.

[0317] Non-limiting examples of additional components that are useful for the present disclosure include: (1) an internal ribosome entry cite (IRES), (2) an untranslated region (UTR), (3) a sequence encoding a signal peptide, (4) a translation initiation sequence, (5) a sequence encoding a RNA binding protein, (6) a sequence encoding a 2A peptide, (7) a translation enhancer element, or (8) any combination of (1) to (7).

[0318] In some aspects, the vectors disclosed herein further comprise additional genetic elements. In some aspects, the additional genetic element is: a restriction enzymes recognition site, a selectable marker, an antibiotic resistance conferring sequences (e.g., an ampicillin or a kanamycin resistance element), a self-cleaving sites coding sequences (e.g., self-cleaving 2A peptides: P2A, T2A, and F2A either with or without furin cleavage site), comprises a translational enhancer element (TEE), origin of replication sites, and a promoter sequence (e.g., a T7, T3, or Sp6 phage polymerase promoter).

[0319] In some aspects, the vectors disclosed herein comprise a translational enhancer element (TEE). As used herein, the term "**translational enhancer element**" refers to cis-acting sequences that increase the expression of a protein encoded by a nucleotide sequence. Non-limiting examples of TEEs that can be used with the present disclosure are known in the art, *see, e.g.*, US20130177581A, which is incorporated herein by reference in its entirety. When such a pre-circRNA or circRNA produced as disclosed herein is introduced into a target cell, the expression

of the payload is increased, *e.g.*, as compared to a corresponding pre-circRNA or circRNA where the payload sequence does not comprise the TEE.

[0320] In some aspects, the TEE is positioned between the transcription promoter and the start codon of a sequence (*e.g.*, payload sequence). In some aspects, a TEE useful for the present disclosure has at least about 50%, at least about 55%, at least about 60%, at least about 65%, at least about 70%, at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 95%, or at least about 99% sequence identity with any of the TEEs provided in U.S. Publication Number US 20140147454, US20090226470, US20070048776, US20130177581, US20110124100, WO1999024595, WO2012009644, WO2009075886, WO2007025008, U.S. Pat. No. 6,310,197, U.S. Pat. No. 6,849,405, U.S. Pat. No. 7,456,273, U.S. Pat. No. 7,183,395, each of which is herein incorporated by reference in its entirety.

[0321] In some aspects, the vectors disclosed herein comprises multiple TEEs. For example, in some aspects, a pre-circRNA or circRNA produced as disclosed herein comprises a payload sequence, wherein the payload sequence comprises at least 1, at least 2, at least 3, at least 4, at least 5, at least 6, at least 7, at least 8, at least 9, at least 10, at least 11, at least 12, at least 13, at least 14, at least 15, at least 16, at least 17, at least 18 at least 19, at least 20, at least 21, at least 22, at least 23, at least 24, at least 25, at least 30, at least 35, at least 40, at least 45, at least 50, at least 55, or more than 60 TEE sequences. In some aspects, the TEE sequences in the 5'UTR of the RNA (*e.g.*, modified RNA) are the same or different TEE sequences. In some aspects, the TEE sequences are in a pattern such as ABABAB or AABBAABBAABB or ABCABCABC or variants thereof repeated once, twice, or more than three times. In these patterns, each letter, A, B, or C represent a different TEE sequence at the nucleotide level.

[0322] In some aspects, the vectors disclosed herein comprise an antibiotic resistance conferring sequence. In some aspects, the genetic constructs disclosed herein comprise an antibiotic resistance conferring sequence at least 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% identical to an antibiotic resistance conferring sequence of SEQ ID NO: 111. In some aspects, the genetic constructs (*e.g.*, comprised in a vector), precursor RNAs, or circular RNAs disclosed herein comprises an antibiotic resistance conferring sequence of SEQ ID NO: 111.

[0323] In some aspects, the DNA genetic constructs (*e.g.*, comprised in a vector), pre-circRNAs, and circRNA disclosed herein are between 300 and 10000, 400 and 9000, 500 and 8000, 600 and 7000, 700 and 6000, 800 and 5000, 900 and 5000, 1000 and 5000, 1100 and 5000, 1200 and 5000, 1300 and 5000, 1400 and 5000, and/or 1500 and 5000 nucleotides in length. In some

aspects, the DNA genetic constructs (e.g., comprised in a vector), pre-circRNAs, and circRNA disclosed herein are at least 300 nt, 400 nt, 500 nt, 600 nt, 700 nt, 800 nt, 900 nt, 1000 nt, 1100 nt, 1200 nt, 1300 nt, 1400 nt, 1500 nt, 2000 nt, 2500 nt, 3000 nt, 3500 nt, 4000 nt, 4500 nt, or 5000 nt in length. In some aspects, the DNA genetic constructs (e.g., comprised in a vector), pre-circRNAs, and circRNA disclosed herein are no more than 3000 nt, 3500 nt, 4000 nt, 4500 nt, 5000 nt, 6000 nt, 7000 nt, 8000 nt, 9000 nt, or 10000 nt in length. In some embodiments, the length of a DNA, linear RNA, and/or circular RNA polynucleotide provided herein is about 300 nt, 400 nt, 500 nt, 600 nt, 700 nt, 800 nt, 900 nt, 1000 nt, 1100 nt, 1200 nt, 1300 nt, 1400 nt, 1500 nt, 2000 nt, 2500 nt, 3000 nt, 3500 nt, 4000 nt, 4500 nt, 5000 nt, 6000 nt, 7000 nt, 8000 nt, 9000 nt, or 10000 nt.

IV. Modified Nucleotides

[0324] In some aspects, the genetic constructs disclosed herein comprises one or more modified nucleosides. In some aspects, the one or more modified nucleosides comprises 6-azacytidine, 2-thio-cytidine, α -thio-cytidine, pseudo-iso-cytidine, 5-aminoallyl-uridine, 5-iodo-uridine, N1-methyl-pseudouridine, 5,6-dihydrouridine, α -thio-uridine, 4-thio-uridine, 6-azauridine, 5-hydroxy-uridine, deoxy-thymidine, pseudo-uridine, inosine, α -thio-guanosine, 8-oxoguanosine, O6-methyl-guanosine, 7-deaza-guanosine, N1-methyl adenosine, 2-amino-6-chloropurine, N6-methyl-2-amino-purine, 6-chloro-purine, N6-methyl-adenosine, α -thio-adenosine, 8-azido-adenosine, 7-deaza-adenosine, pyrrolo-cytidine, 5-methyl-cytidine, N4-acetyl-cytidine, 5-methyl-uridine, 5-iodo-cytidine, and combinations thereof.

[0325] In some aspects, the genetic constructs disclosed herein comprises one or more uridines which have been replaced by a modified nucleoside. In some aspects, the modified nucleoside replacing uridine is pseudouridine (ψ), N1-methyl-pseudouridine (m1 ψ) or 5-methyl-uridine (m5U).

V. Circular RNAs

[0326] In some aspects, the circRNAs prepared by the pre-circRNA constructs or methods described herein have greater stability than a linear RNA (e.g., an mRNA) comprising a same sequence. In some aspects, circRNAs prepared by the pre-circRNA constructs or methods described herein have a stability at least 0.5, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95, 100, 150, 200, 300, 400, 500, 600, 700, 800, 900, 1000 folds greater than that of a linear RNA (e.g., an mRNA) comprising a same sequence. In some aspects, the circRNAs prepared by the pre-circRNA constructs or methods described herein have greater

stability than a linear RNA (e.g., an mRNA) comprising a same sequence when administered to a cell. In some aspects, the cell is comprised in an organism. In some aspects, the organism is a human. In some aspects, the circRNAs prepared by the pre-circRNA constructs or methods described herein have greater stability than a linear RNA (e.g., an mRNA) comprising a same sequence, a same modification, a same optimized UTR, a cap, and/or a polyA tail, and encoding a same payload. In some aspects, the circRNAs prepared by the pre-circRNA constructs or methods described herein have a stability at least 0.5, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95, 100, 150, 200, 300, 400, 500, 600, 700, 800, 900, 1000 folds greater than that of a linear RNA (e.g., an mRNA) comprising a same sequence, a same modification, a same optimized UTR, a cap, and/or a polyA tail, and encoding a same payload. In some aspects, the circRNAs prepared by the pre-circRNA constructs or methods described herein has greater stability than a linear RNA (e.g., an mRNA) comprising a same sequence, a same modification, a same optimized UTR, a cap, and/or a polyA tail, and encoding a same protein when administered to a cell. In some aspects, the cell is comprised in an organism. In some aspects, the organism is a human. In some aspects, the circRNAs prepared by the pre-circRNA constructs or methods described herein have a stability greater than or equal to that of a pre-determined threshold value. In some aspects, the pre-determined threshold value is a stability of a reference linear RNA (e.g., an mRNA) comprising a same sequence as the circular RNA. In some aspects, the pre-determined threshold value is the stability of a reference linear RNA (e.g., an mRNA) comprising a same sequence, a same modification, a same optimized UTR, a cap, and/or a polyA tail, and encoding a same payload. In some aspects, the circRNAs have a stability greater than or equal to that of a pre-determined threshold value when administered to a cell. In some aspects, the cell is comprised in an organism. In some aspects, the organism is a human.

[0327] In some aspects, the circRNAs prepared by the pre-circRNA constructs or methods described herein has greater half-life or functional half-life than a linear RNA (e.g., an mRNA) comprising a same sequence. In some aspects, the circRNAs prepared by the pre-circRNA constructs or methods described herein have a half-life or functional half-life at least 0.5, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95, 100, 150, 200, 300, 400, 500, 600, 700, 800, 900, 1000 folds greater than that of a linear RNA (e.g., an mRNA) comprising a same sequence. In some aspects, the circRNAs prepared by the pre-circRNA constructs or methods described herein have greater half-life or functional half-life than a linear RNA (e.g., an mRNA) comprising a same when administered to a cell. In some aspects, the cell is comprised in an organism. In some aspects, the organism is a human. In some aspects, the

circRNAs prepared by the pre-circRNA constructs or methods described herein have greater half-life or functional half-life than a linear RNA (e.g., an mRNA) comprising a same sequence, a same modification, a same optimized UTR, a cap, and/or a polyA tail, and encoding a same payload. In some aspects, the circRNAs prepared by the pre-circRNA constructs or methods described herein have a half-life or functional half-life at least 0.5, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95, 100, 150, 200, 300, 400, 500, 600, 700, 800, 900, 1000 folds greater than that of a linear RNA (e.g., an mRNA) comprising a same sequence, a same modification, a same optimized UTR, a cap, and/or a polyA tail, and encoding a same payload. In some aspects, the circRNAs prepared by the pre-circRNA constructs or methods described herein have greater half-life or functional half-life than a linear RNA (e.g., an mRNA) comprising a same sequence, a same modifications, a same optimized UTR, a cap, and/or a polyA tail, and encoding a same payload when administered to a cell. In some aspects, the cell is comprised in an organism. In some aspects, the organism is a human. In some aspects, the circRNAs have a half-life or functional half-life in a cell greater than or equal to that of a pre-determined threshold value. In some aspects, the pre-determined threshold value is the half-life or functional half-life of a reference linear RNA (e.g., an mRNA) comprising a same sequence as the circRNAs. In some aspects, the pre-determined threshold value is a half-life or functional half-life of a reference linear RNA (e.g., an mRNA) comprising a same sequence, a same modification, a same optimized UTR, a cap, and/or a polyA tail, and encoding a same payload. In some aspects, the circRNAs have a half-life or functional half-life greater than or equal to that of a pre-determined threshold value when administered to a cell. In some aspects, the cell is comprised in an organism. In some aspects, the organism is a human.

[0328] In some aspects, the circRNAs prepared by the pre-circRNA constructs or methods described herein have a half-life or functional half-life of at least 5 hours, 10 hours, 15 hours, 20 hours, 30 hours, 40 hours, 50 hours, 60 hours, 70 hours or 80 hours. In some aspects, the circular RNA provided herein has a half-life of 5-80, 10-70, 15-60, and/or 20-50 hours. In some aspects, the circRNAs prepared by the pre-circRNA constructs or methods described herein have a half-life or functional half-life of at least 1 day, 2 days, 3 days, 4 days, 5 days, 6 days, 7 days, 8 days, 9 days, or 10 days. In some aspects, the half-life or functional half-life is determined by a nucleic acid assay. In some aspects the functional half-life is determined by a functional protein assay. In some aspects, the circular RNA comprises a nucleotide sequence encoding a reporter gene. In some aspects, the functional half-life is determined by an *in vitro* luciferase assay measuring the expression of the reporter gene. In some aspects, the functional half-life is determined by an *in*

vitro luciferase assay, wherein the activity of Gaussia luciferase (GLuc) is measured in cells (e.g. HepG2) expressing the circular RNA polynucleotide every 1, 2, 6, 12, or 24 hours over 1, 2, 3, 4, 5, 6, 7, or 14 days. In some aspects, the functional half-life is determined by an *in vitro* fluorescence assay, wherein fluorescence (e.g., from a GFP protein) is measured in cells (e.g. HepG2) expressing the circular RNA every 1, 2, 6, 12, or 24 hours over 1, 2, 3, 4, 5, 6, 7, or 14 days. In other aspects, the functional half-life is determined by an *in vivo* assay, wherein levels of a protein encoded by the expression sequence of the circular RNA polynucleotide are measured in patient serum or tissue samples every 1, 2, 6, 12, or 24 hours over 1, 2, 3, 4, 5, 6, 7, or 14 days.

[0329] In some aspects, the circRNAs prepared by the pre-circRNA constructs or methods described herein have greater magnitude of expression than a linear RNA (e.g., an mRNA) comprising a same sequence. In some aspects, the circRNAs prepared by the pre-circRNA constructs or methods described herein have a magnitude of expression at least 0.5, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95, 100, 150, 200, 300, 400, 500, 600, 700, 800, 900, 1000 folds greater than that of a linear RNA (e.g., an mRNA) comprising a same sequence. In some aspects, the circRNAs prepared by the pre-circRNA constructs or methods described herein have greater magnitude of expression than a linear RNA (e.g., an mRNA) comprising a same sequence when administered to a cell. In some aspects, the cell is comprised in an organism. In some aspects, the organism is a human. In some aspects, the circRNAs prepared by the pre-circRNA constructs or methods described herein have greater magnitude of expression than a linear RNA (e.g., an mRNA) comprising a same sequence, a same modification, a same optimized UTR, a cap, and/or a polyA tail, and encoding a same payload. In some aspects, the circRNAs prepared by the pre-circRNA constructs or methods described herein have a magnitude of expression at least 0.5, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95, 100, 150, 200, 300, 400, 500, 600, 700, 800, 900, 1000 folds greater than that of a linear RNA (e.g., an mRNA) comprising a same sequence, a same modification, a same optimized UTR, a cap, and/or a polyA tail, and encoding a same payload. In some aspects, the circRNAs prepared by the pre-circRNA constructs or methods described herein have greater magnitude of expression than a linear RNA (e.g., an mRNA) comprising a same sequence, a same modification, a same optimized UTR, a cap, and/or a polyA tail, and encoding a same payload when administered to a cell. In some aspects, the cell is comprised in an organism. In some aspects, the organism is a human. In some aspects, the circRNAs have a magnitude of expression greater than or equal to that of a pre-determined threshold value. In some aspects, the pre-determined threshold value is a magnitude of expression of a reference linear RNA (e.g., an mRNA)

comprising a same sequence as the circRNAs. In some aspects, the pre-determined threshold value is a magnitude of expression of a reference linear RNA (e.g., an mRNA) comprising a same sequence, a same modification, a same optimized UTR, a cap, and/or a polyA tail, and encoding a same payload. In some aspects, the circRNAs have a magnitude of expression greater than or equal to that of a pre-determined threshold value when administered to a cell. In some aspects, the cell is comprised in an organism. In some aspects, the organism is a human.

[0330] In some aspects, the circRNAs prepared by the pre-circRNA constructs or methods described herein has less immunogenicity than a linear RNA (e.g., an mRNA) comprising a same sequence. In some aspects, the circRNAs described herein have an immunogenicity at least 0.5, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95, 100, 150, 200, 300, 400, 500, 600, 700, 800, 900, 1000 folds lower than that of a linear RNA (e.g., an mRNA) comprising a same sequence. In some aspects, the circular RNA described herein have less immunogenicity than a linear RNA (e.g., an mRNA) comprising a same sequence when administered to a cell. In some aspects, the cell is comprised in an organism. In some aspects, the organism is a human. In some aspects, the circRNAs provided herein have less immunogenicity than a linear RNA (e.g., an mRNA) comprising a same sequence, a same modifications, a same optimized UTR, a cap, and/or a polyA tail, and encoding a same payload. In some aspects, the circRNAs provided herein have an immunogenicity at least 0.5, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95, 100, 150, 200, 300, 400, 500, 600, 700, 800, 900, 1000 folds lower than that of a linear RNA (e.g., an mRNA) comprising a same sequence, a same modification, a same optimized UTR, a cap, and/or a polyA tail, and encoding a same payload. In some aspects, the circRNAs provided herein have less immunogenicity than a linear RNA (e.g., an mRNA) comprising a same sequence, a same modification, a same optimized UTR, a cap, and/or a polyA tail, and encoding a same payload when administered to a cell. In some aspects, the cell is comprised in an organism. In some aspects, the organism is a human. In some aspects, the circRNAs have an immunogenicity lower than or equal to that of a pre-determined threshold value. In some aspects, the pre-determined threshold value is an immunogenicity of a reference linear RNA (e.g., an mRNA) comprising a same sequence as the circRNAs. In some aspects, the pre-determined threshold value is an immunogenicity of a reference linear RNA (e.g., an mRNA) comprising a same sequence, a same modification, a same optimized UTR, a cap, and/or a polyA tail, and encoding a same payload. In some aspects, the circRNAs have an immunogenicity lower than or equal to that of a pre-determined threshold value when administered

to a cell. In some aspects, the cell is comprised in an organism. In some aspects, the organism is a human.

[0331] In some aspects, the circular RNA described herein is less immunogenic than an equivalent linear RNA (*i.e.*, a linear RNA comprising a same sequence as the circular RNA, or a linear RNA comprising a same sequence, a same modification, a same optimized UTR, a cap, and/or a polyA tail, and encoding the same protein as the circular RNA (e.g., mRNA)) when exposed to an immune system of an organism or a certain type of immune cell.

[0332] In some aspects, the circRNAs described herein do not comprise an excision scar and have less immunogenicity than a same circRNA comprising an excision scar and comprising a same sequence. In some aspects, the circRNAs provided herein have an immunogenicity at least 0.5, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95, 100, 150, 200, 300, 400, 500, 600, 700, 800, 900, 1000 folds lower than that of a same circRNA comprising an excision scar and comprising a same sequence. In some aspects, the circRNAs provided herein have less immunogenicity than a same circRNA comprising an excision scar and comprising a same sequence when administered to a cell. In some aspects, the cell is comprised in an organism. In some aspects, the organism is a human. In some aspects, the circRNAs provided herein have less immunogenicity than a same circRNA comprising an excision scar and comprising a same sequence, a same modification, a same optimized UTR, and encoding a same payload. In some aspects, the circRNAs provided herein have an immunogenicity at least 0.5, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95, 100, 150, 200, 300, 400, 500, 600, 700, 800, 900, 1000 folds lower than that of a same circRNA comprising an excision scar and comprising a same sequence, a same modification, a same optimized UTR, and encoding a same payload. In some aspects, the circRNAs provided herein have less immunogenicity than a same circRNA comprising an excision scar and comprising a same sequence, a same modification, a same optimized UTR, and encoding a same payload when administered to a cell. In some aspects, the cell is comprised in an organism. In some aspects, the organism is a human. In some aspects, the circRNAs have an immunogenicity lower than or equal to that of a pre-determined threshold value. In some aspects, the pre-determined threshold value is an immunogenicity of a reference circRNA comprising an excision scar and comprising a same sequence as the circRNAs. In some aspects, the pre-determined threshold value is an immunogenicity of a reference circRNA comprising an excision scar and comprising a same sequence, a same modification, a same optimized UTR, and encoding a same payload. In some aspects, the circRNAs have an immunogenicity lower than or equal to that of a pre-determined threshold value when administered

to a cell. In some aspects, the cell is comprised in an organism. In some aspects, the organism is a human.

[0333] In some aspects, the circular RNA described herein is less immunogenic than an equivalent circRNA comprising an excision scar (*i.e.*, a circRNA comprising an excision scar and comprising a same sequence, a same modifications, an same optimized UTR, and encoding the same payload) when exposed to an immune system of an organism or a certain type of immune cell.

[0334] In some aspects, the circRNAs provided herein are associated with modulated production of cytokines when exposed to an immune system of an organism or a certain type of immune cell. In some aspects, the circRNAs provided herein are associated with reduced production of IFN- β 1, RIG-I, IL-2, IL-6, IFN γ , and/or TNF α when exposed to an immune system of an organism or a certain type of immune cell as compared to an equivalent linear RNA (*i.e.*, a linear RNA comprising a same sequence as the circRNAs, or a linear RNA comprising a same sequence, a same modification, a same optimized UTR, a cap, and/or a polyA tail, and encoding a same payload as the circular RNA (e.g., mRNA)).

[0335] In some aspects, the circRNAs provided herein does not comprise an excision scar. In some aspects, the circRNAs provided herein are associated with modulated production of cytokines when exposed to an immune system of an organism or a certain type of immune cell. In some aspects, the circRNAs provided herein are associated with reduced production of IFN- β 1, RIG-I, IL-2, IL-6, IFN γ , and/or TNF α when exposed to an immune system of an organism or a certain type of immune cell as compared to an equivalent circRNA comprising an excision scar (*i.e.*, a circRNA comprising a same sequence as the circRNA, and further comprising an excision scar; or a circRNA comprising a same sequence, a same modification, a same optimized UTR, and encoding a same payload as the circular RNA (e.g., mRNA), and further comprising an excision scar).

[0336] In some aspects, the circular RNA provided herein is associated with less IFN- β 1, RIG-I, IL-2, IL-6, IFN γ , and/or TNF α transcript induction when exposed to an immune system of an organism or a certain type of immune cell as compared to an equivalent linear RNA (*i.e.*, a linear RNA comprising the same expression sequence as the circular RNA, or a linear RNA comprising the same expression sequence, the same modifications, an optimized UTR, a cap, and/or a polyA tail, and encoding the same protein as the circular RNA (e.g., mRNA)).

[0337] In some aspects, the circular RNA provided herein is associated with less IFN- β 1, RIG-I, IL-2, IL-6, IFN γ , and/or TNF α transcript induction when exposed to an immune system of

an organism or a certain type of immune cell as compared to an equivalent circRNA comprising an excision scar (*i.e.*, a circRNA comprising a same sequence as the circRNA, and further comprising an excision scar; or a circRNA comprising a same sequence, a same modification, a same optimized UTR, and encoding a same payload as the circular RNA (e.g., mRNA), and further comprising an excision scar).

[0338] In some aspects, transcription of the pre-circRNAs from the DNA genetic constructs (e.g., comprised in a vector), is *in vitro*. In some aspects, *in vitro* transcription of the pre-circRNAs from the DNA genetic constructs (e.g., comprised in a vector), is in a cell. In some aspects the cell is a prokaryotic cell. In some aspects the cell is a eukaryotic cell. *In vitro* transcription of the precursor RNA from the genetic construct (e.g., comprised in a vector) in a cell can be via added polymerases or polymerases encoded by nucleic acids transfected into the cell, or via endogenous (e.g., RNA polymerase II). In some aspects, the pre-circRNAs are transcribed in the cytoplasm of the cell. In some aspects, the pre-circRNAs are transcribed in the nucleus of the cell. In some aspects, *in vitro* transcription of the pre-circRNAs from the DNA genetic construct (e.g., comprised in a vector), is in a cell-free reaction. *In vitro* transcription of the precursor RNA from the genetic construct (e.g., comprised in a vector) in a cell free reaction can be by any RNA polymerase. In some aspects, the RNA polymerase is a bacteriophage RNA polymerase (e.g., T7, T3, or Sp6).

[0339] In some aspects, circularization of the pre-circRNAs to produce the circRNAs, is *in vitro*. In some aspects, *in vitro* circularization of the pre-circRNAs to produce the circRNAs, is in a cell. In some aspects the cell is a prokaryotic cell. In some aspects the cell is a eukaryotic cell. In some aspects, the circRNAs are produced in the cytoplasm of the cell. In some aspects, the circRNAs are produced in the nucleus of the cell. In some aspects, circularization of the pre-circRNAs to produce the circRNAs is in a cell-free reaction.

[0340] In some aspects, transcription of the pre-circRNAs from the DNA genetic constructs (e.g., comprised in a vector), is *in vivo*. In some aspects, DNA genetic constructs (e.g., comprised in a vector) is administered to a cell comprised in a multicellular organism and the pre-circRNAs are transcribed from the DNA genetic constructs (e.g., comprised in a vector) in the cell comprised in the multicellular organism.

[0341] In some aspects, circularization of the pre-circRNAs to produce the circRNAs is *in vivo*. In some aspects, DNA genetic constructs (e.g., comprised in a vector) are administered to a cell comprised in a multicellular organism, the pre-circRNAs are transcribed from the DNA genetic constructs (e.g., comprised in a vector) and the pre-circRNAs circularize in the cell comprised in

the multicellular organism. In some aspects, pre-circRNAs are administered to a cell comprised in a multicellular organism and the pre-circRNAs circularize in the cell comprised in the multicellular organism.

[0342] In certain aspects, the circRNA provided herein is administered to a cell. In some aspects the sequence encoding the payload comprised in the circular RNAs described herein is expressed in a cell. In some aspects, the cell is comprised in an organism. In some aspects, the organism is an animal. In some aspects, the animal is a human.

VI. Pre-circRNA Circularization

[0343] In some aspects, the pre-circRNA comprises a 5' splice site and a 3' splice site. In the aspects, circularization of the pre-circRNA takes place by a two-step transesterification reaction with an exogenous GTP (α G) with its 3'-OH acting as an initiating nucleophile. Binding of the α G in the G-binding site in P7 positions the 3'-OH of GTP to attack the 5' splice site. During the first transesterification step the α G is attached to the 5'-end of the intron RNA by a 3'-5' phosphodiester bond. This step is followed by conformational changes allowing the upstream exon's terminal 3' guanosine (ω G) to trade position with the α G and occupy the G-binding site to initiate the second transesterification reaction. The 3'-OH of the upstream exon attacks the 3' splice site (an interaction facilitated by the formation of P10) promoting the ligation of upstream and downstream exons and the release of the intron RNA.

[0344] In some aspects, the pre-circRNA comprises a 3' splice site and a catalytic nucleotide. In the aspects, circularization of the pre-circRNA takes place by a one-step transesterification reaction with the upstream exon's terminal 3' guanosine (ω G) trading position with the α G and occupy the G-binding site to initiate the transesterification reaction. The 3'-OH of the upstream exon attacks the 3' splice site (an interaction facilitated by the formation of P10) promoting the ligation of upstream and downstream exons and the release of the intron RNA.

VII. Compositions Comprising the Genetic Constructs and Methods of Making circRNAs

[0345] In some aspects, provided herein are compositions comprising the genetic constructs capable of forming circular RNAs disclosed herein (*i.e.*, pre-circRNAs or DNA genetic constructs from which pre-circRNAs can be transcribed) and one or more additional component.

[0346] In some aspects, the additional component is one or more metal ion. In some aspects, the one or more metal ion can be used in compositions comprising the pre-circRNAs disclosed herein to produce a circRNA. In some aspects the metal ion is a cation. In some aspects

the cation is a monovalent cation. In some aspects, the monovalent cation is Na⁺, K⁺, NH₄⁺, which can be included in the form NaCl, KCl, NH₄Cl, Ammonium acetate, and Ammonium sulfate. In some aspects, the cation is a bivalent cation. In some aspects, the metal ion is Mg²⁺, Ca²⁺, Cu²⁺, Mn²⁺, Zn²⁺, Pb²⁺, Co²⁺, Cd²⁺, Sn²⁺, Ni²⁺, Fe²⁺, Ba²⁺, or Sr²⁺. In some aspects, the cation is a trivalent cation. In some aspects the metal ion is Ytterbium (Yb³⁺).

[0347] In some aspects, the additional component is a RNA polymerase. In some aspects, the RNA polymerase can be used in compositions comprising the DNA genetic constructs disclosed herein to produce the pre-circRNAs disclosed herein. In some aspects, the RNA polymerase is a T7 RNA polymerase, a T3 RNA polymerase, a SP6 RNA polymerase, or any combination thereof.

[0348] In some aspects, the additional component in a salt, a buffer, or any additional component that may facilitate the transcription of the DNA genetic constructs disclosed herein into the pre-circRNAs disclosed herein, and/or any additional component that may facilitate the circularization of the pre-circRNAs disclosed herein to produce the circRNAs disclosed herein.

[0349] Any suitable method known in the published literature can be used to transcribe the DNA genetic constructs disclosed herein into the pre-circRNAs disclosed herein, and/or to circularize the pre-circRNAs disclosed herein to produce the circRNAs disclosed herein.

[0350] The DNA genetic constructs disclosed herein can be transcribed into the pre-circRNAs disclosed herein in vitro, in vivo, or any combination thereof. Similarly, the pre-circRNAs disclosed herein in vitro or in vivo can be circularized to produce the circRNAs disclosed herein in vitro, in vivo, or any combination thereof.

[0351] The vectors comprising the DNA genetic constructs and the pre-circRNAs provided herein can be made using standard techniques of molecular biology. For example, the various elements of the vectors provided herein can be obtained by chemical synthesis or by recombinant methods. For example, the genetic elements comprised in the genetic constructs can be derived from cDNA derived from cells, or from a vector known to include the genetic element of interest. Alternatively, the genetic elements of the genetic constructs provided herein can also be produced synthetically, rather than cloned, based on the known sequences. The complete sequence can be assembled from overlapping oligonucleotides prepared by standard methods and assembled into the complete sequence. See, e.g., Edge, Nature (1981) 292:756; Nambair et al. , Science (1984) 223:1299; and Jay et al. , J. Biol. Chem. (1984) 259:631 1. Particular nucleotide sequences can be obtained from vectors harboring the desired sequences or synthesized completely or in part using various oligonucleotide synthesis techniques known in the art. Additionally, nucleotides sequences

can be modified or amplified using any technique known in the art, such as site-directed mutagenesis and polymerase chain reaction (PCR) techniques where appropriate. One method of obtaining nucleotide sequences encoding the desired vector elements is by annealing complementary sets of overlapping synthetic oligonucleotides produced in a conventional, automated polynucleotide synthesizer, followed by ligation with an appropriate DNA ligase and amplification of the ligated nucleotide sequence via PCR. See, e.g., Jayaraman et al., Proc. Natl. Acad. Sci. USA (1991) 88:4084-4088. Additionally, oligonucleotide-directed synthesis (Jones et al., Nature (1986) 54:75-82), oligonucleotide directed mutagenesis of preexisting nucleotide regions (Riechmann et al., Nature (1988) 332:323-327 and Verhoeven et al., Science (1988) 239:1534-1536), and enzymatic filling-in of gapped oligonucleotides using T4 DNA polymerase (Queen et al., Proc. Natl. Acad. Sci. USA (1989) 86: 10029-10033) can be used.

[0352] The pre-circRNAs provided herein can be generated by incubating a vector provided herein under conditions permissive of transcription of the precursor RNA encoded by the vector. In some aspects, a precursor RNA is synthesized by incubating a genetic construct (e.g., comprised in a vector) provided herein that comprises an RNA polymerase promoter with a compatible RNA polymerase enzyme (e.g., a T7, T3, or Sp6 RNA polymerase) under conditions permissive of *in vitro* transcription. In some aspects, a genetic construct (e.g., comprised in a vector) is incubated inside of a cell and transcribed by a bacteriophage RNA polymerase or by a RNA polymerase II.

[0353] In some aspects, provided herein is a method of generating pre-circRNA by performing *in vitro* transcription using a DNA genetic construct (e.g., comprised in a vector) provided herein as a template (e.g., a vector provided herein comprising a RNA polymerase promoter). In some aspects, the resulting pre-circRNA can be used to generate circRNAs provided herein by incubating it in the presence of magnesium and ammonium ions at a temperature at which RNA circularization occurs (e.g., between 20 °C and 60 °C).

[0354] In some aspects, a composition comprising circular RNA is purified. Circular RNA may be purified by any known method known in the art, such as column chromatography, gel filtration chromatography, and size exclusion chromatography. In some aspects, purification comprises one or more of the following steps: phosphatase treatment, HPLC size exclusion purification, and RNase R digestion. In some aspects, purification comprises the following steps in order: RNase R digestion, phosphatase treatment, and HPLC size exclusion purification. In some aspects, purification comprises reverse phase HPLC. In some aspects, a purified composition contains less double stranded RNA, DNA splints, triphosphorylated RNA, phosphatase proteins,

protein ligases, capping enzymes and/or nicked RNA than unpurified RNA. In some aspects, a purified composition is less immunogenic than an unpurified composition. In some aspects, immune cells exposed to a purified composition produce less IFN- β 1, RIG-I, IL-2, IL-6, IFN γ , and/or TNF α than immune cells exposed to an unpurified composition.

VIII. Pharmaceutical Compositions

[0355] In some aspects, provided herein are pharmaceutical compositions comprising a DNA genetic construct (e.g., comprised in a vector), a pre-circRNA, or a circRNA disclosed herein. In some aspects, the DNA genetic constructs (e.g., comprised in a vector), the pre-circRNAs, or the circRNAs disclosed her disclosed herein can be used as therapeutic agents. In some aspects, the therapeutic agent is a DNA genetic construct (e.g., comprised in a vector) provided herein. In some aspects, the therapeutic agent is a pre-circRNA provided herein. In some aspects, the therapeutic agent is a circRNA provided herein. In some aspects the therapeutic agent is a vector comprising a genetic construct provided herein. In some aspects, the therapeutic agent is a cell comprising a genetic construct provided herein (e.g., a human cell, such as a human T cell). In certain aspects, the composition further comprises a pharmaceutically acceptable carrier. In some aspects, the pharmaceutical composition comprises pharmaceutically acceptable carriers, buffer agents, excipients, salts, or stabilizers in the form of lyophilized formulations or aqueous solutions, *See, e.g.*, Remington: The Science and Practice of Pharmacy 20th Ed. (2000) Lippincott Williams and Wilkins, Ed. K. E. Hoover. "Acceptable", as used herein, means that the carrier must be compatible with the active ingredient of the composition and not deleterious to the subject to be treated. In some aspects, the carrier is capable of stabilizing the active ingredient.

[0356] In some aspects, the pharmaceutical composition can be formulated for intratumoral, intrathecal, intramuscular, intravenous, subcutaneous, inhalation, intradermal, intralymphatic, intraocular, intraperitoneal, intrapleural, intraspinal, intravascular, nasal, percutaneous, sublingual, submucosal, transdermal, or transmucosal administration. In some aspects of the disclosure, the pharmaceutical composition can be formulated for intratumoral injection. Intratumoral injection, as used herein, refers to direct injections into the tumor. A high concentration of composition can be achieved in situ, while using small amounts of drugs. Local delivery of immunotherapies allows multiple combination therapies, while preventing significant system exposure and off-target toxicities.

[0357] In some aspects, the pharmaceutical composition can be formulated for intramuscular injection, intravenous injection, or subcutaneous injection.

IX. General Techniques

[0358] The practice of the present disclosure will employ, unless otherwise indicated, conventional techniques of molecular biology (including recombinant techniques), microbiology, cell biology, biochemistry and immunology, which are within the skill of the art. Molecular Cloning: A Laboratory Manual, second edition (Sambrook, et al., 1989) Cold Spring Harbor Press; Oligonucleotide Synthesis (M.J. Gait, ed., 1984); Methods in Molecular Biology, Humana Press; Cell Biology: A Laboratory Notebook (J. E. Cellis, ed., 1998) Academic Press; Animal Cell Culture (R. I. Freshney, ed. 1987); Introduction to Cell and Tissue Culture (J.P. Mather and P.E. Roberts, 1998) Plenum Press; Cell and Tissue Culture: Laboratory Procedures (A. Doyle, J.B. Giffiths, and D.G. Newell, eds., 1993-8) J. Wiley and Sons; Method of Enzymology (Academic Press, Inc.); Handbook of Experimental Immunology (D.M. Weir and C.C. Blackwell, eds.); Gene Transfer Vectors for Mammalian Cells (J.M. Miller and M.P. Calos, eds., 1987); Current Protocols in Molecular Biology (F.M. Ausubel, et al., eds., 1987); PCR: The Polymerase Chain Reaction, (Mullis, et al., eds., 1994); Current Protocols in Immunology (J.E. Coligan et al., eds., 1991); Short Protocols in Molecular Biology (Wiley and Sons, 1999); Immunobiology (C.A. Janeway and P. Travers, 1997); Antibodies (P. Finch, 1997); Antibodies: a practical approach (D. Catty, ed., IRL Press, 1988-1989); Monoclonal antibodies: a practical approach (P. Shepherd and C. Dean, eds., Oxford University Press, 2000); Using antibodies: a laboratory manual (E. Harlow and D. Lane, Cold Spring Harbor Laboratory Press, 1999); The Antibodies (M. Zanette and J.D. Capra, eds., Harwood Academic Publishers, 1995). Without further elaboration, it is believed that one skilled in the art can, based on the above description, utilize the present disclosure to its fullest extent. All publications cited herein (including those listed above and elsewhere in the present disclosure) are incorporated by reference in their entirety.

Table 15: Sequence Table (nnnnnnnnnn indicates an exemplary location for an exonic sequence; mmmmmmmmm indicates an exemplary location for an IRES).

SEQ ID NO	DESCRIPTION	SEQUENCE
42	Anabaena PCC7120	CACTAATACGACTCACTATAGGGAGACCCTCGACCGTCGATTGTCCACTGGTC AACAATAGATGACTTACAATAATCGGAAGGTGCAGAGACTCGACGGGAGCTA CCCTAACGTCAAGACGAGGGTAAAGAGAGAGTCCAATTCTCAAAGCCAATAGG CAGTAGCGAAAGCTGCAAGAGAATGAAAATCCGTTGACCTTAAACGGTCGTGT GGGTTCAGTCCCTCCACCCCCACGCCGGAACGCAATAGCCGAAAAACAAAA AACAAAAAAACAAAAAAACAAAAAAACAAAAACAmmmmmmmmmnnnn nnnnnnAAAAAACAAAAACAAAACGGCTATTATGCGTTACCGGCGAGACGC TACGGACTTAAATAATTGAGCCTTAAAGAAGAAATTCCTTAAGTGGATGCTCT CAAACCTCAGGGAAACCTAAATCTAGTTATAGACAAGGCAATCCTGAGCCAAGC CGAAGTAGTAATTAGTAAGACCAGTGGACAATCGACGGATAACAGCATATCTA GATAAGAAGAGCATCCCATGGTCATAGCTGTTTCCTGGCAGCTCTGGCCCGTG TCTCAAATCTCTGATGTTACATTGCACAAGATAAAATATATCATCATGAAC

		AATAAACTGTCTGCTTACATAAACAGTAATACAAGGGGTGTTATGAGCCATA TTCAACGGGAAACGTCGAGGCCGCGATTAAATTTCCAACATGGATGCTGATTTA TATGGGTATAAATGGGCTCGCGATAATGTCTGGGCAATCAGGTGCGACAATCTA TCGCTTGATATGGGAAGCCCGATGCGCCAGAGTTGTTTCTGAAACATGGCAAAG GTAGCGTTGCCAATGATGTTACAGATGAGATGGTCAGACTAAACTGGCTGACG GAATTTATGCCTCTTCCGACCATCAAGCATTTTATCCGTACTCCTGATGATGC ATGGTTACTCACCCTGCGATCCCCGAAAAACAGCATTTCCAGGTATTAGAAG AATATCCTGATTACAGGTGAAAATATTGTTGATGCGCTGGCAGTGTTTCTGCGC CGGTTGCATTTCGATTCTGTTTGTAAATTGTCTTTTAAACAGCGATCGCGTATT TCGTCTCGCTCAGGCGCAATCACGAATGAATAACGGTTTGGTTGATGCGAGTG ATTTTGTATGACGAGCGTAATGGCTGGCCTGTTGAACAAGTCTGGAAAGAAATG CATAAACTTTTGCCATTCTCACCAGATTCACTCGTCACTCATGGTGATTTCTC ACTTGATAACCTTATTTTGTACGAGGGGAAATTAATAGGTTGTATTGATGTTG GACGAGTCGGAATCGCAGACCGATACCAGGATCTTGCCATCCTATGGAATGCG CTCGGTGAGTTTCTCCTTCATTACAGAAACGGCTTTTTTCAAAAATATGTTAT TGATAATCCTGATATGAATAAATGTCAGTTTCATTTGATGCTCGATGAGTTTT TCTAATCAGAATTGGTTAATTGGTTGTAACACTGGCAGAGCATTACGCTGACT TGACGGGACGGCGCAAGCTCATGACCAAAATCCCTTAACGTGAGTTACGGCTC GTTCCACTGAGCGTCAGACCCCGTAGAAAAGATCAAAGGATCTTCTTGAGATC CTTTTTTTCTGCGCGTAATCTGCTGCTTGCAAACAAAAAACACCGCTACCA GCGGTGGTTTGTGTTGCCGGATCAAGAGCTACCAACTCTTTTTCCGAAGGTAAAC TGGCTTCAGCAGAGCGCAGATACCAAACTGTTCTTCTAGTGTAGCCGTAGT TAGGCCACCACTTCAAGAACTCTGTAGCACCGCTACATACCTCGCTCTGCTA ATCCTGTTACCAAGTGGCTGCTGCCAGTGGCGATAAGTCTGTCTTACCGGGTT GGACTCAAGACGATAGTTACCGGATAAGGCGCAGCGGTGCGGCTGAACGGGGG GTTCTGTGCACACAGCCAGCTTGGAGCGAACGACCTACACCGAACTGAGATAC CTACAGCGTGAGCTATGAGAAAGCGCCACGCTTCCCGAAGGAGAAAGCGGGA CAGGTATCCGGTAAGCGGCAGGGTCGGAACAGGAGAGCGCACGAGGGAGCTTC CAGGGGGAAACGCCTGGTATCTTTATAGTCTGTGCGGTTTTCGCCACCTCTGA CTTGAGCGTCGATTTTGTGATGCTCGTCAGGGGGGCGGAGCCTATGGAAAAA CGCCAGCAACGCGGCTTTTTACGGTTCTTGCCCTTTTGCTGGCCTTTTGCTC ACATGTTCTTTCTGCGTTATCCCTGATTCTGTGGATAACCGTATTACCGCC TTTGAGTGAGCTGATACCGCTCGCCGAGCCGAACGACCGAGCGCAGCGAGTC AGTGAGCGAGGAAGCGGAGGAGCGCCCAATACGCAAACCGCCTCTCCCGCGC GTTGGCCGATTCAATTAATGTCAGTGGCACGACAGGTTTCCCGATGGAAAGCG GGCAGTGAGCGCAACGCAATTAATACGCGTACCGCTAGCCAGGAAGAGTTTGT AGAAACGCAAAAAGGCCATCCGTCAGGATGGCCTTCTGCTTAGTTTGATGCCT GGCAGTTTATGGCGGGCGTCTGCGCGCCACCTCCGGGGCGTTGCTTCACAA CGTTCAAATCCGCTCCCGGCGGATTTGTCTTACTCAGGAGAGCGTTACCCGAC AAACAACAGATAAAACGAAAGGCCAGTCTTCCGACTGAGCCTTTCTGTTTTAT TTGATGCCTGGCAGTTCCCTACTCTCGCGTTAACGCTAGCATGGATGTTTTCC CAGTCACGACGTTGTAAAACGACGGCCAGTCTTAAGCTCGGCTCTTCG
43	Tetrahymena thermophila	CCCGTAGAAAAGATCAAAGGATCTTCTTGAGATCCTTTTTTTCTGCGCGTAAT CTGCTGCTTGCAAACAAAAAACACCGCTACCAGCGGTGGTTTGTGTTGCCGG ATCAAGAGCTACCAACTCTTTTTTCCGAAGGTAACCTGGCTTCAGCAGAGCGCAG ATACCAAATACTGTTCTTCTAGTGTAGCCGTAGTTAGGCCACCACTTCAAGAA CTCTGTAGCACCGCTACATACCTCGCTCTGCTAATCCTGTTACCAAGTGGCTG CTGCCAGTGGCGATAAGTCTGTCTTACCGGGTTGGACTCAAGACGATAGTTA CCGGATAAGGCGCAGCGGTGCGGCTGAACGGGGGGTTCTGTGCACACAGCCAG CTTGAGCGAACGACCTACACCGAAGTACCTACAGCGTGAGCTATGAG AAAGCGCCACGCTTCCCGAAGGGAGAAAGGCGGACAGGTATCCGGTAAGCGGC AGGGTCGGAACAGGAGAGCGCACGAGGGAGCTTCCAGGGGGAAACGCCTGGTA TCTTTATAGTCTGTGCGGTTTTCGCCACCTCTGACTTGAGCGTCGATTTTTGT GATGCTCGTCAGGGGGGCGGAGCCTATGGAACAAACGCGCAGCAACGCGGCTTT TTACGGTTCTTGCCCTTTTGCTGGCCTTTTGCTCACATGTTCTTTCTGCGTT ATCCCTGATTCTGTGGATAACCGTATTACCGCTTTGAGTGAGCTGATACCG CTCGCCGAGCCGAACGACCGAGCGCAGCGAGTCAGTGAGCGAGGAAGCGGAG GAGCGCCCAATACGCAAACCGCTCTTCCCGCGCGTTGGCCGATTCAATTAATG CAGCTGGCACGACAGGTTTCCCGACTGGAAGCGGGGAGGCGCAACGCAAA TTAATACGCGTACCGCTAGCCAGGAAGAGTTGTAGAAACGCAAAAAGGCCAT CCGTCAGGATGGCCTTCTGCTTAGTTTGATGCCTGGCAGTTTATGGCGGGCGT

		CCTGCCCGCCACCCTCCGGGCCGTTGCTTCACAACGTTCAAATCCGCTCCCGG CGGATTTGTCTACTCAGGAGAGCGTTACCCGACAAACAACAGATAAAACGAA AGGCCAGTCTTCCGACTGAGCCTTTCGTTTTATTTGATGCCTGGCAGTTCCC TACTCTCGCGTTAACGCTAGCATGGATGTTTTCCAGTCACGACGTTGTA CGACGGCCAGTCTTAAGCTCGGCTCTTCGCACTAATACGACTCACTATAGGGA GtCCCTCGACCGTCGATTGTCCACTGGTCATTTACCTTTGGAGGGAAAAAGTTA TCAGGCATGCACCTGGTAGCTAGTCTTTAAACCAATAGATTGCATCGGTTTAA AAGGCAAGACCGTCAAATTGCGGAAAGGGGTCAACAGCCGTTTCACTACCAAG TCTCAGGGGAAACTTTGAGATGGCCTTGCAAAGGGTATGGTAATAAGCTGACG GACATGGTCCTAACACGCAGCCAAGTCCTAAGTCAACAGATCTTCTGTTGAT ATGGATGCAGTTACAGACTAAATGTGGTTCGGGGAAGATGATTCTTCTCAT AAGATATAGTCGGACCTCTCCTTAATGGGAGCTAGCGGATGAAGTGATGCAAC ACTGGAGCCGCTGGGAACTAATTTGTATGCGAAAGTATATTGATTAGTTTGG AGTACTCGTAAGGTAAAAATCCGTTGACCTTAAACGGTCGTGTGGGTTCAAGT CCCTCCACCCCCACGCCGAAACGCAATAGCCGAAAAACAAAAACAAAAAA ACAAAAAAACCAAAAAACAAAAACACAmmmmmmmmmnnnnnnnnnnAAA AAACAAAAAACAAACGGCTATTATGCGTTACCGCGAGACGCTACGGACCCCT CTAAATAGCAATGACCAGTGGACAATCGACGGATAACAGCATATCTAGATAAG AAGAGCATCCCATGGTCATAGCTGTTTCTGGCAGCTCTGGCCCGTGTCTCAA AATCTCTGATGTTACATTGCACAAGATAAAAAATATATCATCATGAACAATAAA ACTGTCTGCTTACATAAACAGTAATACAAGGGGTGTTATGAGCCATATTCAAC GGGAAACGTCGAGGCCGCGATTAAATTCACATGGATGCTGATTATATGGG TATAAATGGGCTCGCGATAATGTGGGCAATCAGGTGCGACAATCTATCGCTT GTATGGGAAGCCCGATGCGCCAGAGTTGTTTCTGAAACATGGCAAAAGGTAGCG TTGCCAATGATGTTACAGATGAGATGGTCAGACTAACTGGCTGACGGAATTT ATGCCCTTTCGACCATCAAGCATTTTATCCGTA CTCTGATGATGCATGGTT ACTCACCCTGCGATCCCCGGAACAGCATTCAGGTATTAGAAGAATATC CTGATTCAGGTGAAAAATTTGTTGATGCGCTGGCAGTGTTCCTGCGCCGGTTG CATTCGATTCTCTGTTTGTAAATGTCTTTTAAACAGCGATCGCGTATTTCTGCT CGCTCAGGCGCAATACGAATGAATAACGGTTTGGTTGATGCGAGTGAATTTG ATGACGAGCGTAATGGCTGGCCTGTTGAACAAGTCTGGAAGAAATGCATAAA CTTTTGCCATTCTCACCAGATTCAAGTCGTCACTCATGGTGAATTTCTCACTTGA TAACCTTATTTTGTAGCAGGGGAAATTAATAGGTTGTATTGATGTTGGACGAG TCGGAATCGCAGACCGATACAGGATCTTGCCATCCTATGGAATGCCTCGGT GAGTTTCTCTCTTCAATTACAGAAACGGCTTTTCAAAAATATGGTATGATAA TCCTGATATGAATAAATTGCAGTTTCATTTGATGCTCGATGAGTTTTTCTAAT CAGAAATGGTTAATTGGTTGTAACTGGCAGAGCATTACGCTGACTTGACGG GACGGCGCAAGCTCATGACCAAAATCCCTTAACGTGAGTTACGCGTCGTTCCA CTGAGCGTCAGAC
44	Azoarcus	CCCGTAGAAAAGATCAAAGGATCTTCTTGAGATCCTTTTTTTCTGCGCGTAAT CTGCTGCTTGCAAAACAAAAAACCCGCTACCAGCGGTGGTTTGTGTTGCCGG ATCAAGAGCTACCAACTCTTTTTCCGAAGGTAACTGGCTTCAGCAGAGCGCAG ATACCAAATACTGTTCTTCTAGTGTAGCCGTAGTTAGGCCACCACTTCAAGAA CTCTGTAGCACCGCTACATACCTCGCTCTGCTAATCCTGTTACCAGTGGCTG CTGCCAGTGGCGATAAGTCTGTCTTACCGGGTTGGACTCAAGACGATAGTTA CCGGATAAGGCGCAGCGGTTCGGGCTGAACGGGGGGTTCTGTCACACAGCCAG CTTGGAGCGAACGACCTACACCGAAGTACCTACAGCTGAGCTGAGCTATGAG AAAGCGCCACGCTTCCCGAAGGGAGAAAGGCGACAGGTATCCGGTAAGCGGC AGGGTCGGAACAGGAGAGCGCACGAGGGAGCTTCAGGGGGAAACGCCTGGTA TCTTTATAGTCTTGTGCGGTTTCGCCACCTCTGACTTGAGCGTCGATTTTTGT GATGCTCGTCAGGGGGGCGGAGCCTATGGAACAAACGCGAGCAACGCGGCCTTT TTACGGTTCCTGGCCTTTTGTGTCCTTTTGTCTCACATGTTCTTTCTGCGTT ATCCCTGATTCTGTGGATAACCGTATTACCGCCTTTGAGTGAGCTGATACCG CTCGCCGAGCCGAACGACCGAGCGCAGCGAGTCAGTGAGCGAGGAAGCGGAG GAGCGCCCAATACGCAACCGCCTCTCCCGCGCGTTGGCCGATTCAATTAATG CAGCTGGCACGACAGGTTTCCCGACTGGAAGCGGGCAGTGAGCGCAACGCAA TTAATACGCGTACCGCTAGCCAGGAAGAGTTTGTAGAAACGCAAAAGGCCAT CCGTCAGGATGGCCTTCTGCTTAGTTTGTATGCCTGGCAGTTTATGCGGGCGT CCTGCCCGCCACCTTCCGGGCCGTTGCTTCAACAGCTTCAATCCGCTCCCGG CGGATTTGTCTACTCAGGAGAGCGTTACCCGACAAACAACAGATAAAACGAA AGGCCAGTCTTCCGACTGAGCCTTTCGTTTTATTTGATGCCTGGCAGTTCCC

		TACTCTCGCGTTAACGCTAGCATGGATGTTTTCCAGTCACGACGTTGTAAAA CGACGGCCAGTCTTAAGCTCGGCTCTTCGCACTAATACGACTCACTATAGGGA GACCCTCGACCGTCGATTGTCCACTGGTCTttGctCCTTGCGCCGGGAAACCA CGCAAGGGATGGTGTCAAATTCGGCGAAACCTAAGCGCCCCCGGGCGTATG GCAACGCCGAGCCAAGCTTCGGCGCCTGCGCCGATGAAGGTGTAGAGACTAGA CGGCACCCACCTAAGGCAAACGCTATGGTGAAGGCATAGTCCAGGGAGTGGCG AAAGTCACACAAACCGGAAAATCCGTTGACCTTAAACGGTCGTGTGGGTTCAA GTCCCTCCACCCACGCGGAAACGCAATAGCCGAAAAACAAAAACAAAAA AAACAAAAAAACCAAAAAACAAAAACACAmmmmmmmmmnnnnnnnnnnA AAAAACAAAAAACAAACGGCTATTATGCGTTACCGGCGAGACGCTACGGAag TaaaGACCAGTGGACAATCGACGGATAACAGCATATCTAGATAAGAAGAGCAT CCCATGGTCATAGCTGTTTCCCTGGCAGCTCTGGCCCGTGTCTCAAAATCTCTG ATGTTACATTGCACAAGATAAAAAATATATCATCATGAACAATAAACTGTCTG CTTACATAAACAGTAATACAAGGGGTGTTATGAGCCATATTCAACGGGAAACG TCGAGGCCGCGATTAAATTCCAACATGGATGCTGATTTATATGGGTATAAATG GGCTCGCGATAATGTCCGGCAATCAGGTGCGACAATCTATCGCTTGTATGGGA AGCCCGATGCGCCAGAGTTGTTTTCTGAAACATGGCAAAGGTAGCGTTGCCAAT GATGTTACAGATGAGATGGTCAGACTAAACTGGCTGACCGGAATTTATGCCTCT TCCGACCATCAAGCATTTTATCCGTACTCCTGATGATGCATGGTTACTCACCA CTGCGATCCCCGAAAAACAGCATTCAGGTATTAGAAGAAATATCCTGATTCA GGTGAAAATATTGTTGATGCGCTGGCAGTGTTCTCTGCGCCGGTTGCATTGAT TCCTGTTTGTAATTGTCTTTTAAACAGCGATCGCGTATTTCTGCTCGCTCAGG CGCAATCACGAATGAATAACGGTTTGGTTGATGCGAGTGATTTTGATGACGAG CGTAATGGCTGGCCTGTTTGAACAAGTCTGGAAGAAATGCATAAACTTTTGCC ATTCTCACCGGATTTCAGTCGTCACTCATGGTGATTTCTCACTTGATAACCTTA TTTTTGACGAGGGGAAATTAATAGGTTGTATTGATGTTGGACGAGTCGGAATC GCAGACCGATACCAGGATCTTGCCATCCTATGGAACGCTCGGTGAGTTTTC TCCTTCATTACAGAAACGGCTTTTCAAAAATATGGTATTGATAATCCTGATA TGAATAAATTGCAGTTTCATTTGATGCTCGATGAGTTTTTCTAATCAGAATTG GTTAATTGGTTGTAAACACTGGCAGAGCATTACGCTGACTTGACGGGACGGCGC AAGCTCATGACCAAAATCCCTTAAACGTGAGTTACGCGTCGTTCCACTGAGCGT CAGAC
45	Anabaena azollae	CCCGTAGAAAAGATCAAAGGATCTTCTTGAGATCCTTTTTTTCTGCGCGTAAT CTGCTGCTTGCAAACAAAAAACCCCGCTACCAGCGGTGGTTTGTGTGCCGG ATCAAGAGCTACCAACTCTTTTTCCGAAGGTAACTGGCTTCAGCAGAGCGCAG ATACCAAATACTGTCTTCTAGTGTAGCCGTAGTTAGGCCACCACTTCAAGAA CTCTGTAGCACCGCTACATACCTCGCTCTGCTAATCCTGTTACCAGTGGCTG CTGCCAGTGGCGATAAGTCGTGTCTTACCGGGTTGGACTCAAGACGATAGTTA CCGGATAAGGCGCAGCGGTCTGGGCTGAACGGGGGGTTCTGTGCACACAGCCAG CTTGAGCGAACGACCTACACCGAACTGAGATACCTACAGCGTGAGCTATGAG AAAGCGCCACGCTTCCCGAAGGGAGAAAGGCGGACAGGTATCCGGTAAGCGGC AGGGTCGGAACAGGAGAGCGCACAGGGGAGCTTCCAGGGGCAACCGCTTGTA TCTTTATAGTCTGTGCGGTTTCGCCACCTCTGACTTGAGCGTCGATTTTGT GATGCTCGTCAGGGGGGCGGAGCCTATGGAAAACGCCAGCAACGCGGCCTTT TTACGGTTCTTGCCCTTTTGCTGGCCTTTTGCTCACATGTTCTTTCTGCGTT ATCCCCGTGATTCTGTGGATAACCGTATTACCGCCTTTGAGTGAGCTGATACCG CTCGCCGACAGCGAACGACCGAGCGCAGCGAGTCAGTGAGCGAGGAAGCGAG GAGCGCCAATACGCAAAACCGCCTCTCCCCGCGCGTTGGCCGATTCATTAATG CAGCTGGCACGACAGGTTTCCCGACTGGAAAGCGGGCAGTGAGCGCAACGCAA TTAATACGCGTACCGCTAGCCAGGAAGAGTTTGTAGAAACGCAAAAGGCCAT CCGTACAGGATGGCCTTCTGCTTAGTTTGATGCCTGGCAGTTTATGGCGGGCGT CCTGCCCCGCCACCTCCGGGCCGTTGCTTCAACAGTTCAAATCCGCTCCCGG CGGATTGTCTTACTCAGGAGAGCGTTTACCAGCAAAACAAGATAAAAACGAA AGGCCAGTCTTCCGACTGAGCCTTTCGTTTTATTTGATGCCTGGCAGTTCCC TACTCTCGCGTTAACGCTAGCATGGATGTTTTCCAGTCACGACGTTGTAAAA CGACGGCCAGTCTTAAGCTCGGCTCTTCGCACTAATACGACTCACTATAGGGA GACCCTCGACCGTCGATTGTCCACTGGTCAAATCAAATTTAAATCCCAAT TCAAAATTCGGGAAGGTGCAGAGACTCGACGGGAGCTACCCTAACGTAAAGC CGAGGTTAAAGAGAGAGTCCAATTCTCAAAGCCTGAAGTTGCTGAAGCAACAA GGCAGTAGTGAAAGCTGCGAGAGAATGAAAATCCGTTGACCTTAAACGGTCGT GTGGGTTCAAGTCCCTCCACCCACGCGGAAACGCAATAGCCGAAAAACAA

		AAAACAAAAAAAAACAAAAAAAAACAAAAAAAAACAAACACAmmmmmmmmmnnnnnnnnnnAAAAAACAAAAACAAAAACGGCTATTATGCGTTACCGGCGAGACGCTACGGACTTAGAAAACTGAGCCTTGATAGAGAAATCTTTTAAGTGGAAGCTCTCAAATTCAGGGAAACCTAAATCTGTTTACAGATATGGCAATCCTGAGCCAA GCCAGAAAAATTAGACTTGAGATTGATTTTGGAGTTAGACCAGTGGACAATC GACGGATAACAGCATATCTAGATAAGAAGAGCATCCCATGGTCATAGCTGTTT CCTGGCAGCTCTGGCCCGTGTCTCAAAATCTCTGATGTTACATTGCACAAGAT AAAAATATATCATCATGAACAATAAACTGTCTGCTTACATAAACAGTAATAC AAGGGGTGTTATGAGCCATATTCAACGGGAAACGTCGAGGCCGCGATTAAATT CCAACATGGATGCTGATTTATATGGGTATAAATGGGCTCGCGATAATGTCGGG CAATCAGGTGCGACAATCTATCGCTTGATGGGAAGCCCGATGCGCCAGAGTT GTTTCTGAAACATGGCAAAGGTAGCGTTGCCAATGATGTTACAGATGAGATGG TCAGACTAAACTGGCTGACGGAATTTATGCCTCTTCCGACCATCAAGCATTTT ATCCGTACTCCTGATGATGCATGGTTACTCACCCTGCGATCCCCGAAAAAC AGCATTCAGGTATTAGAAGAATATCCTGATTAGGTGAAAAATATTGTTGATG CGCTGGCAGTGTTCTTGGCCCGGTTGCATTGATTCCTGTTTGTAAATTGTCCT TTTAACAGCGATCGCGTATTTCTGCTCGCTCAGGCGCAATCAGCAATGAATAA CGGTTTGGTTGATGCGAGTGATTTTGATGACGAGCGTAATGGCTGGCCTGTTG AACAACTCTGGAAGAAATGCATAAACTTTTGCCATTCTCACCAGATTAGTC GTCACTCATGGTGATTTCTCACTTGATAACCTTATTTTTGACGAGGGGAAATT AATAGTTGTATTGATGTTGGACGAGTCGGAATCGCAGACCGATACAGGATC TTGCCATCCTATGGAACCTGCTCGGTGAGTTTCTCCTTCAATTACAGAAACGG CTTTTTCAAAAATATGGTATTGATAATCCTGATATGAATAAATTGCAGTTTCA TTTGATGCTCGATGAGTTTTTCTAATCAGAAATTGGTTAATTGGTTGTAAACACT GGCAGAGCATTACGCTGACTTGACGGGACGGCGCAAGCTCATGACCAAAATCC CTTAACGTGAGTTACGCGTCGTTCCACTGAGCGTCAGAC
46	Bacteriophage T4 Td	CCCGTAGAAAAGATCAAAGGATCTTCTTGAGATCCTTTTTTTCTGCGCGTAAT CTGCTGCTTGCAAACAAAAAACCCACCGCTACCAGCGGTGGTTTGTGTTGCCGG ATCAAGAGCTACCAACTCTTTTTCCGAAGGTAACTGGCTTCAGCAGAGCGCAG ATACCAAATACTGTCTTCTAGTGTAGCCGTAGTTAGGCCAACACTTCAAGAA CTCTGTAGCACCGCTACATACCTCGCTCTGCTAATCCTGTTACCAGTGGCTG CTGCCAGTGGCGATAAGTCTGTCTTACCGGGTTGGACTCAAGACGATAGTTA CCGGATAAGGCGCAGCGGTGCGGGCTGAACGGGGGGTTCTGTCACACAGCCAG CTTGGAGCGAACGACCTACACCGAACTGAGATACCTACAGCGTGAGCTATGAG AAAGCGCCACGCTTCCCGAAGGGAGAAAGGCGGACAGGTATCCGTTAAGCGGC AGGGTCGGAACAGGAGAGCGCACGAGGGAGCTTCCAGGGGGAAACGCCTGGTA TCTTTATAGTCCTGTGCGGTTTTCGCCACCTCTGACTTGAGCGTCGATTTTTGT GATGCTCGTCAGGGGGGCGGAGCCTATGGAAAAACGCCAGCAACGCGGCCTTT TTACGGTTCTTGGCCTTTTGTGCTGGCCTTTTGTCTACATGTTCTTCTGCGTT ATCCCCTGATTCTGTGGATAACCGTATTACCGCCTTTGAGTGAGCTGATACCG CTCGCCGACGCGAACGACCGAGCGCAGCGAGTCAGTGAGCGAGGAACGGAG GAGCGCCCAATACGCAAAACCGCCTCTCCCCGCGCGTTGGCCGATTCAATTATG CAGCTGGCACGACAGGTTTCCCGACTGGAAAGCGGGCAGTGAGCGCAACGCAA TTAATACGCGTACCGCTAGCCAGGAAGAGTTTGTAGAAACGCAAAAAGGCCAT CCGTCAGGATGGCCTTCTGCTTAGTTTGATGCCTGGCAGTTTATGGCGGGCGT CCTGCCCCGCCACCTCCGGGCCGTTGCTTCAACAGTTCAAATCCGCTCCCGG CGGATTGTCTTACTCAGGAGAGCGTTTACCAGCAAAACAACAGATAAAACGAA AGGCCAGTCTTCCGACTGAGCCTTTCGTTTTATTGATGCCTGGCAGTTCCC TACTCTCGGTTAACGCTAGCATGGATGTTTTCCAGTCACGACGTTGTAAAA CGACGGCCAGTCTTAAGCTCGGCTCTTCGCACTAATACGACTCACTATAGGGA GACCTCGACCGTCGATTGTCCACTGGTCGGGAGACCTCGAATGGAATTGGT TCTACATAAATGCCTAACGACTATCCCTTTGGGGAGTAGGGTCAAGTGACTCG AAACGATAGACAACTTGCTTTAAACAAGTTGGAGATATAGTCTGCTCTGCATGG TGACATGCAGCTGGATATAATTCGGGGTAAGATTAACGACCTTATCTGAACA TAATGCTAAATCCGTTGACCTTAAACGGTCGTGTGGGTTCAAGTCCCTCCAC CCCCACGCCGAAACGCAATAGCCGAAAAACAAAAACAAAAACAAAAA AAAACCAAAAAACAAAAACAmmmmmmmmmnnnnnnnnnnAAAAACAAAA AACAAAAACGGCTATTATGCGTTACCGGCGAGACGCTACGGAGGTTAATTGAGG CCTGAGTATAAGGTGACTTATACTTGTAATCTATCTAAACGGGGAACCTCTCT AGTAGACAATCCCGTGCTAAATTGTAGGACTAATTCCATTTATCAGATTTCTA GGACCAGTGGACAATCGACGGATAACAGCATATCTAGATAAGAAGAGCATCCC

		ATGGTCATAGCTGTTTCTCTGGCAGCTCTGGCCCGTGTCTCAAAATCTCTGATG TTACATTGCACAAGATAAAAAATATATCATCATGAACAATAAACTGTCTGCTT ACATAAACAGTAATAACAAGGGGTGTTATGAGCCATATTCAACGGGAAACGTCG AGGCCCGGATTAATTTCCAACATGGATGCTGATTTATATGGGTATAAATGGGC TCGCGATAATGTCTGGGCAATCAGGTGCGACAATCTATCGCTTGTATGGGAAGC CCGATGCGCCAGAGTTGTTTCTGAAACATGGCAAAGGTAGCGTTGCCAATGAT GTTACAGATGAGATGGTCAGACTAAACTGGCTGACGGAATTTATGCCTCTTCC GACCATCAAGCATTTTATCCGTACTCCTGATGATGCATGGTTACTCACCCTG CGATCCCCGGAAAAACAGCATTCAGGTATTAGAAGAATATCCTGATTACAGGT GAAAAATATTGTTGATGCGCTGGCAGTGTTCCTGCGCCGGTTGCATTTCGATTCC TGTTTGTAAATTGTCCTTTTAAACAGCGATCGCGTATTTTCGTCTCGCTCAGGCGC AATCAGGAATGAATAACGGTTTGGTTGATGCGAGTGATTTTGTATGACGAGCGT AATGGCTGGCCTGTTGAACAAGTCTGGAAAGAAATGCATAAACTTTTGCCATT CTCACCGGATTTCAGTCGTCATCGGTGATTTCTCACTTGATAACCTTATTT TTGACGAGGGGAAATTAATAGGTTGTATTGATGTTGGACGAGTCGGAATCGCA GACCGATAACCAGGATCTTGCCATCCTATGGAAGTGCCTCGGTGAGTTTCTCC TTCATTACAGAAACGGCTTTTTCAAAAATATGGTATTGATAATCCTGATATGA ATAAAATTGCAGTTTTCATTTGATGCTCGATGAGTTTTCCTAATCAGAAATTGGTT AATTGGTTGTAACACTGGCAGAGCATTACGCTGACTTGACGGGACGGCGCAAG CTCATGACCAAAATCCCTTAACGTGAGTTACGCGTCGTTCCACTGAGCGTCAG AC
47	Bacteriophage T4 Td	CCCGTAGAAAAGATCAAAGGATCTTCTTGAGATCCTTTTTTTCTGCGCGTAAT CTGCTGCTTGCAAAACAAAAAACACCGCTACCAGCGGTGGTTTGTGTTGCCGG ATCAAGAGCTACCAACTCTTTTTCCGAAGGTAAGTGGCTTCAGCAGAGCGCAG ATACCAAATACTGTTCTTCTAGTGTAGCCGTAGTTAGGCCACCACTTCAAGAA CTCTGTAGCACCGCCTACATACCTCGCTCTGCTAATCCTGTTACCAGTGGCTG CTGCCAGTGGCGATAAGTCGTGCTTACCAGGTTGGACTCAAGACGATAGTTA CCGGATAAAGGCGCAGCGGTCTGGGCTGAACGGGGGGTTCTGTGCACACAGCCAG CTTGAGCGAAGCAGCTACACCGAAGTACCTACAGCGTGAGCTATGAG AAAGCGCCACGCTTCCCGAAGGGAGAAAGGCGGACAGGTATCCGTAAGCGGC AGGGTCGGAACAGGAGAGCGCACGAGGGAGCTTCCAGGGGGAAACGCCTGGTA TCTTTATAGTCCTGTGCGGTTTTCGCCACCTCTGACTTGAGCGTCGATTTTGT GATGCTCGTCAGGGGGGCGGAGCCTATGGAAAAACGCCAGCAACGCGGCCTTT TTACGGTTCTTGGCCTTTTGCTGGCCTTTTGCTCACATGTTCTTTCTGCGTT ATCCCCTGATTCTGTGGATAACCGTATTACCGCCTTTGAGTGAGCTGATACCG CTCGCCGACGCGAACCAGACCGAGCGCAGCGAGTCAGTGAGCGAGGAAGCGGAG GAGCGCCCAATACGCAAACCGCCTCTCCCCGCGCGTTGGCCGATTATTAATG CAGCTGGCACGACAGGTTTCCCGACTGGAAAGCGGGCAGTGAGCGCAACGCAA TTAATACGCGTACCCTAGCCAGGAAGAGTTTGTAGAAACGCAAAAGGCCAT CCGTACAGGATGGCCTTCTGCTTAGTTTGTATGCTTGGCAGTTTATGGCGGGCGT CCTGCCCGCCACCCTCCGGGCGGTTGCTTCACAACGTTCAAACTCCGCTCCCGG CGGATTTGTCTTACTCAGGAGAGCGTTTACCAGACAAACACAGATAAAACGAA AGGCCAGTCTTCCGACTGAGCCTTTGTTTTATTTGATGCCTGGCAGTTCCC TACTCTCGCGTTAACGCTAGCATGGATGTTTTCCAGTCACGACGTTGTAAAA CGACGGCCAGTCTTAAGCTCGGCTCTTCGCACTAATACGACTCACTATAGGGA GACCTCGACCGTCGATTGTCCACTGGTCGGGAGACCCTCGAATGGAATTGGT TCTACATAAAATGCCTAACGACTATCCCTTTGGGGAGTAGGGTCAAGTGACTCG AAACGATAGACAACCTGCTTTAAACAAGTTGGAGATATAGTCTGCTCTGCATGG TGACATGCAGCTGGATATAATTCGGGGTAAGATTAACGACCTTATCTGAACA TAATGCTACCGTTTAATATTGCGTCAGGTAGTAACTACTAATAACCTGC TGAAGCAmmmmmmmmnnnnnnnnnnCAGAGATGTTTTCTTGGGTAAATTGA GGCCTGAGTATAAGGTGACTTATACTTGTAACTATCTAAACGGGGAACCTCT CTAGTAGACAATCCCGTGCTAAATTTGTAGGACTAATTCATTTATCAGATTTCT TAGGACCAGTGGACAATCGACGGATAACAGCATATCTAGATAAGAAGAGCATC CCATGGTCATAGCTGTTTCTTGGCAGCTCTGGCCCGTGTCTCAAAATCTCTGA TGTTACATTGCACAAGATAAAAAATATATCATCATGAACAATAAACTGTCTGC TTACATAAACAGTAATAACAAGGGGTGTTATGAGCCATATTCAACGGGAAACGT CGAGGCCGCGATTAAATTTCCAACATGGATGCTGATTTATATGGGTATAAATGG GCTCGGATAAATGTCTGGGCAATCAGGTGCGACAATCTATCGCTTGTATGGGAA GCCCCGATGCGCCAGAGTTGTTTCTGAAACATGGCAAAGGTAGCGTTGCCAATG ATGTTACAGATGAGATGGTCAGACTAAACTGGCTGACGGAATTTATGCCTCTT

		CCGACCATCAAGCATTTTATCCGTACTCCTGATGATGCATGGTTACTCACCAC TGCGATCCCCGAAAAACAGCATTCAGGTATTAGAAGAATATCCTGATTTCAG GTGAAAAATATTGTTGATGCGCTGGCAGTGTTCTTGC GCCGGTTGCATTTCGATT CCTGTTTGTAATTGTCTTTTAAACAGCGATCGCGTATTTCTGCTCAGGC GCAATCACGAATGAATAACGGTTTGGTTGATGCGAGTGATTTTGTATGACGAGC GTAATGGCTGGCCTGTTGAACAAGTCTGGAAAGAAATGCATAAACTTTTGCCA TTCTCACCGGATTTCAGTCGTCACTCATGGTGATTTCTCACTTGATAACCTTAT TTTTGACGAGGGGAAATTAATAGGTTGTATTGATGTTGGACGAGTCGGAATCG CAGACCGATACCAGGATCTTGCCATCCTATGGAAGTGCCTCGGTGAGTTTCT CCTTCATTACAGAAACGGCTTTTCAAAAAATATGGTATTGATAATCCTGATAT GAATAAATTGCAGTTTCATTTGATGCTCGATGAGTTTCTTAATCAGAAATTGG TTAATTGGTTGTAAACACTGGCAGAGCATTACGCTGACTTGACGGGACGGCGCA AGCTCATGACCAAAATCCCTTAACGTGAGTTACGCGTCGTTCCTGAGCGTCA AGAC
48	Anabaena PCC7120	TAATACGACTCACTATAGGGAGACCCTCGACCGTCGATTGTCCACTGGTCAAC AATAGATGACTTACAACTAATCGGAAGGTGCAGAGACTCGACGGGAGCTACCC TAACGTCAAGACGAGGGTAAAGAGAGAGTCCAATTCTCAAAGCCAATAGGCAG TAGCGAAAGCTGCAAGAGAATGAAAAATCCGTTGACCTTAAACGGTCGTGTGGG TTCAAGTCCCTCCACCCACGCGGAAACGCAATAGCCGAAAAACAAAAAAC AAAAAAACAAAAAAACAAAAAAACAAAAACACAmmmmmmmmmmmmmmmmmmm nnnnAAAAAAACAAAAACAAAAACGGCTATTATGCGTTACCGGCGAGACGCTAC GGACTT
49	Azoarcus	TAATACGACTCACTATAGGGAGACCCTCGACCGTCGATTGTCCACTGGTCTtt GctCCTTGCGCCGGGAAACACGCAAGGGATGGTGTCAAATTCGGCGAAACCT AAGCGCCCGCCGGGCGTATGGCAACGCGGAGCCAAGCTTCGGCGCCTGCGCC GATGAAGGTGTAGAGACTAGACGGCACCCACCTAAGGCAACGCTATGGTGAA GGCATAGTCCAGGGAGTGGCGAAAGTCACACAAACCGGAAAAATCCGTTGACCT TAAACGGTCGTGTGGGTTCAAGTCCCTCCACCCACGCGGAAACGCAATAG CCGAAAAACAAAAACAAAAAAACAAAAAAACAAAAAAACAAAAACACA mm nnnnAAAAAAACAAAAACAAAAACGGCTATTATGCGTTACCGGCGAGACGCTAC ACCGGCGAGACGCTACGGAAGT
50	Anabaena azollae	TAATACGACTCACTATAGGGAGACCCTCGACCGTCGATTGTCCACTGGTCAAC TCAAAATTTAAATCCCAAATTCAAAATTCGGGAAGGTGCAGAGACTCGACG GGAGCTACCCCTAACGTAAAGCCGAGGGTAAAGAGAGAGTCCAATTCTCAAAGC CTGAAGTTGCTGAAGCAACAAGGCAGTAGTGAAAGCTGCGAGAGAATGAAAAAT CCGTTGACCTTAAACGGTCGTGTGGGTTCAAGTCCCTCCACCCACGCGGGA AACGCAATAGCCGAAAAACAAAAACAAAAAAACAAAAAAACAAAAAAACAAAAAA ACAAAAACACAmmm nnnnAAAAAAACAAAAACAAAAACGGCTATTATGCGTTACCGGCGAGACGCTAC CGGACTT
51	Bacteriophage T4 Td	TAATACGACTCACTATAGGGAGACCCTCGACCGTCGATTGTCCACTGGTCGGG AGACCCTCGAATGGAATTGGTTCTACATAAATGCCTAACGACTATCCCTTTGG GGAGTAGGGTCAAGTGACTCGAAACGATAGACAACCTTGCTTTAAACAAGTTGGA GATATAGTCTGCTCTGCATGGTGACATGCAGCTGGATATAAATCCGGGGTAAAG ATTAACGACCTTATCTGAACATAATGCTAAAATCCGTTGACCTTAAACGGTCG TGTGGGTTCAAGTCCCTCCACCCACGCGGAAACGCAATAGCCGAAAAACA AAAAACAAAAAAACAAAAAAACAAAAAAACAAAAAAACAAAAACACAmmmmmmmmmmm nnnnnnnnnnAAAAAAACAAAAACAAAAACGGCTATTATGCGTTACCGGCGAGA CGCTACGGAGGT
52	Anabaena PCC7120	GGAGACCCTCGACCGTCGATTGTCCACTGGTCAACAATAGATGACTTACAAC AATCGGAAGGTGCAGAGACTCGACGGGAGCTACCCCTAACGTCAAGACGAGGGT AAAGAGAGAGTCCAATTCTCAAAGCCAATAGGCAGTAGCGAAAGCTGCAAGAG AATGAAAATCCGTTGACCTTAAACGGTCGTGTGGGTTCAAGTCCCTCCACCCC CACGCCGAAACGCAATAGCCGAAAAACAAAAACAAAAAAACAAAAAAACAAAAAA ACAAAAAAACAAAAACACAmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmm AAAAACGGCTATTATGCGTTACCGGCGAGACGCTACGGAATTAATAATTGAGC CTTAAAGAAGAAATTTCTTAAAGTGGATGCTCTCAAACCTCAGGGAAACCTAAAT CTAGTTATAGACAAGGCAATCCTGAGCCAAGCCGAAGTAGTAATTAGTAAGAC CAGTGACAATCGACGGATAACAGCATATCTAGATA
53	Tetrahymena thermophila	GGAGTCCCTCGACCGTCGATTGTCCACTGGTCATTTACCTTTGGAGGGAAAAAG TTATCAGGCATGCACCTGGTAGCTAGTCTTTAAACCAATAGATTGCATCGGTT TAAAAGGCAAGACCGTCAAATTCGGGGAAGGGGTCAACAGCCGTTCACTAGTACC

[illegible]

		AGACGGCACCCACCTAAGGCAAACGCTATGGTGAAGGCATAGTCCAGGGAGTG GCGAAAGTCACACAAACCGGAAAATCCGTTGACCTTAAACGGTCGTGTGGGTT CAAGTCCCTCCACCCCGACGCCGAAACGCAATAGCCGAAAAACAAAAACAA AAAAACAAAAAAAACCAAAAAACAAACACAmmmmmmmmmnnnnnnnn nnAAAAACAAAAACAAACGGCTATTATGCGTTACCGGCGAGACGCTACGG AagT
60	Anabaena azollae	GGAGACCCTCGACCCTCGATTGTCCACTGGTCAACTCAAAATTTAAAATCCCA AATTCAAAATTCGGGAAGGTGCAGAGACTCGACGGGAGCTACCTAACGTAA AGCCGAGGGTAAAGAGAGAGTCCAATTCTCAAAGCCTGAAGTTGCTGAAGCAA CAAGGCAGTAGTGAAAGCTGCGAGAGAATGAAAATCCGTTGACCTTAAACGGT CGTGTGGGTTCAAGTCCCTCCACCCCGACGCCGAAACGCAATAGCCGAAAA CAAAAAACAAAAAAAACAAAAAAAACCAAAAAACAAACACAmmmmmmmmm mmnnnnnnnnnnAAAAACAAAAACAAACGGCTATTATGCGTTACCGGCGA GACGCTACGGACTT
61	Bacteriophage T4 Td	GGAGACCCTCGACCCTCGATTGTCCACTGGTCGGGAGACCTCGAATGGAATT GGTTCTACATAAAATGCCTAACGACTATCCCTTTGGGGAGTAGGGTCAAGTGAC TCGAAACGATAGACAACTTGCTTTAAACAAGTTGGAGATATAGTCTGCTCTGCA TGGTGACATGCAGCTGGATATAATTCCGGGGTAAGATTAAACGACCTTATCTGA ACATAATGCTAAAAATCCGTTGACCTTAAACGGTCGTGTGGGTTCAAGTCCCTC CACCCCGACGCCGAAACGCAATAGCCGAAAAACAAAAACAAAAAACAAAA AAAAAACCAAAAAACAAACACAmmmmmmmmmnnnnnnnnnnAAAAACA AAAAACAAACGGCTATTATGCGTTACCGGCGAGACGCTACGGAGGT
62	Anabaena PCC7120	GGAGACCUCGACCGUCGAUUGUCCACUGGUCACAAUAGAUGACUUACAACU AAUCGGAAGGUGCAGAGACUCGACGGGAGCUACCCUAACGUCAGACGAGGU AAAGAGAGAGUCCAAUUCUCAAAGCCAAUAGGCAGUAGCGAAAGCUGCAAGAG AAUGAAAAUCCGUUGACCUUAAACGGUCGUGUGGUUUAAGUCCCUCCACCCC CACGCCGAAACGCAUAGCCGAAAAACAAAAACAAAAAACAAAAAAA ACCAAAAAACAAACACAmmmmmmmmmnnnnnnnnnnAAAAACAAAAAAC AAAACGGCUAUUAUGCGUUAACGGCGAGACGCUACGGACUUAAUAAUUGAGC CUUAAAGAAGAAAUUCUUUAAAGUGGAUGCUCUCAAACUCAGGGAAACCUAAU CUAGUUAUAGACAAGGCAAUCCUGAGCCAAGCCGAAGUAGUAAUUAUAGAAC CAGUGGACAAUCGACGGAUAACAGCAUAUCUAGAU
63	Tetrahymena thermophila	GGAGuCCUCGACCGUCGAUUGUCCACUGGUCAUUUACCUUUGGAGGGAAAAAG UUAUCAGGCAUGCACCUGGUAGCUAGUCUUUAAACCAUAGAUGCAUCCGUU UAAAAGGCAAGACCGUCAAAUUGCGGGAAGGGGUCAACAGCCGUUCAGUACC AAGUCUCAGGGGAAACUUUGAGAUUGGCCUUGCAAGGGUUAUGGUAAUAGCUG ACGGACAUGGUCCUAAACACGCAGCCAAGUCCUAAAGUCAACAGAUUCUUGUU GAUAUGGAUGCAGUUCACAGACUAAAUGUCGGUCGGGGAAGAUUAUUCUUCU CAUAAAGAUUAGUCGACCUUCUUAUUGGAGCUAGCGGAUGAAGUGAUGC AACACUGGAGCCGUCGGGAACUAAUUGUAUGCGAAAGUAUAUAGAUUAGUUU UGGAGUACUCGUAAGGUAAAAUCCGUUGACCUUAAACGGUCGUGUGGGUUA AGUCCCUCCACCCCGACGCCGAAACGCAAUAGCCGAAAAACAAAAACAAAA AAAAACAAAAAAAACCAAAAAACAAACACAmmmmmmmmmnnnnnnnnnn AAAAACAAAAAACAAACGGCUAUUAUGCGUUAACGGCGAGACGCUACGGAC CCUCUAAUAGCAAUGACCAGUGGACAAUCGACGGAUAACAGCAUAUCUAGAU A
64	Azoarcus	GGAGACCUCGACCGUCGAUUGUCCACUGGUCuuuGcuCCUUGCGCCGGGAAA CCACGCAAGGGAUGGUGUCAAAUUCGGCGAAACCUAAGCGCCCGCCGGGCGU AUGGCAACGCCGAGCCAAGCUUCGGCGCCUGCGCGAUGAAGGUGUAGAGACU AGACGGCACCCACCUAAGGCAACGCUAUGGUGAAGGCAUAGUCCAGGGAGUG GCGAAAGUCACACAAACCGGAAAUCCGUUGACCUUAAACGGUCGUGUGGGUU CAAGUCCCUCCACCCCGACGCCGGAACGCAUAGCCGAAAAACAAAAACAA AAAAACAAAAAAAACCAAAAAACAAACACAmmmmmmmmmnnnnnnnnnn nnAAAAACAAAAACAAACGGCUAUUAUGCGUUAACGGCGAGACGCUACGG AagUaaaGACCAGUGGACAAUCGACGGAUAACAGCAUAUCUAGAU
65	Anabaena azollae	GGAGACCUCGACCGUCGAUUGUCCACUGGUCACUCAAAAUUAAAAUCCCA AAUUCAAAAUUCGGGAAGGUGCAGAGACUCGACGGGAGCUACCCUAACGUAA AGCCGAGGGUAAAGAGAGAGUCCAAUUCUCAAGCCUGAAGUUGCUGAAGCAA CAAGGCAGUAGUGAAAGCUGCGAGAGAAUGAAAAUCCGUUGACCUUAAACGGU CGUGUGGGUUAAGUCCUCCACCCCGACGCCGGAACGCAAUAGCCGAAAA CAAAAAACAAAAAAAACAAAAAAAACCAAAAAACAAACACAmmmmmmmmm

		mmnnnnnnnnnnAAAAAACAAAAACAAAACGGCUAUUAUGCGUUAACGGCGA GACGCUACGGACUUAAGAAAACUGAGCCUUGAUAGAGAAUUCUUUAAGUGGAA GCUCUCAAAUUCAGGGAAACCUAAAUCUGUUUACAGAUUAGGCAAUCCUGAGC CAAGCCCAGAAAAUAGACUUGAGAUUUGAUUUUGGAGUUAGACCAGUGGACA AUCGACGGAUAACAGCAUAUCUAGAUA
66	Bacteriophage T4 Td	GGAGACCCUCGACCGUCGAUUGUCCACUGGUCGGGAGACCCUCGAAUGGAAU GGUUCUACAUAUAAUGCCUAACGACUAUCCCUUUGGGGAGUAGGGUCAAGUGAC UCGAAACGAUAGACAACUUGCUUUAACAAGUUGGAGAUUAGUCUGCUCUGCA UGGUGACAUGCAGCUGGAUAUAAUCCGGGGUAGAUAUACGACCUUAUCUGA ACAUAAUGCUAAAAUCCGUUGACCUUAAACGGUUCGUGUGGGUUAAGUCCUC CACCCCCACGCCGGAACGCAAUAGCCGAAAAACAAAAACAAAAAACAAA AAAAAACCAAAAAACAAAACACAmmmmmmmmmnnnnnnnnnnAAAAACA AAAAACAAAACGGCUAUUAUGCGUUAACGGCGAGACGCUACGGAGGUUAUUG AGGCCUGAGUAUAGGUGACUUAUACUUGUAAUCUAUCUAAACGGGGAACCUC UCUAGUAGACAAUCCGUGCUAAAUUGUAGGACUAAUCCAUUUUACAGAUUU CUAGGACCAGUGGACAAUCGACGGAUAACAGCAUAUCUAGAUA
67	Bacteriophage T4 Td	GGAGACCCUCGACCGUCGAUUGUCCACUGGUCGGGAGACCCUCGAAUGGAAU GGUUCUACAUAUAAUGCCUAACGACUAUCCCUUUGGGGAGUAGGGUCAAGUGAC UCGAAACGAUAGACAACUUGCUUUAACAAGUUGGAGAUUAGUCUGCUCUGCA UGGUGACAUGCAGCUGGAUAUAAUCCGGGGUAGAUAUACGACCUUAUCUGA ACAUAAUGCUACCGUUAUUAUUGCGUCAGGUAGUAAACUACUACUACAACC UGCUGAAGCAmmmmmmmmmnnnnnnnnnnCAGAGAUGUUUUCUUGGGUUAU UGAGGCCUGAGUAUAGGUGACUUAUACUUGUAAUCUAUCUAAACGGGGAACC UCUCUAGUAGACAAUCCGUGCUAAAUUGUAGGACUAAUCCAUUUUACAGAU UUCUAGGACCAGUGGACAAUCGACGGAUAACAGCAUAUCUAG
68	Anabaena PCC7120	GGAGACCCUCGACCGUCGAUUGUCCACUGGUCACAAUAGAUGACUUAACAACU AAUCGGAAGGUGCAGAGACUCGACGGGAGCUACCCUAACGUAAGACGAGGGU AAAGAGAGAGUCCAAUUCUCAAGCCAAUAGGCAGUAGCGAAAGCUGCAAGAG AAUGAAAAUCCGUUGACCUUAAACGGUCGUGUGGGUUAAGUCCUCCACCCC CACGCCGGAACGCAAUAGCCGAAAAACAAAAACAAAAAAACAAAAAA ACCAAAAAACAAAAACAmmmmmmmmmnnnnnnnnnnAAAAACAAAAAAC AAAACGGCUAUUAUGCGUUAACGGCGAGACGCUACGGACUU
69	Azoarcus	GGAGACCCUCGACCGUCGAUUGUCCACUGGUCuuuGcuCCUUGCGCCGGGAAA CCACGCAAGGGAUGGUGUCAAAUUCGGCGAAACCUAAGCGCCCGCCGGGCGU AUGGCAACGCCGAGCCAAGCUUCGGCGCCUGCGCGGAUGAAGGUGUAGAGACU AGACGGCACCCACCUAAGGCAAACGCUAUGGUGAAGGCAUAGUCCAGGGAGUG GCGAAAGUCACACAAACCGGAAAAUCCGUUGACCUUAAACGGUCGUGUGGGUU CAAGUCCUCCACCCCCACGCCGGAACGCAAUAGCCGAAAAACAAAAACAA AAAAACAAAAAAACCAAAAAACAAAAACACAmmmmmmmmmnnnnnnnnnn nnAAAAACAAAAACAAAAACGGCUAUUAUGCGUUAACGGCGAGACGCUACGG AagU
70	Anabaena azollae	GGAGACCCUCGACCGUCGAUUGUCCACUGGUCACCUAAAAUUAUAAUCCCA AAUUCAAAAUUCGGGAAGGUGCAGAGACUCGACGGGAGCUACCCUAACGUAA AGCCGAGGGUAAAGAGAGAGUCCAAUUCUCAAGCCUGAAGUUGCUGAAGCAA CAAGGCAGUAGUGAAAGCUGCGAGAGAAUGAAAAUCCGUUGACCUUAAACGGU CGUGUGGGUUAAGUCCUCCACCCCCACGCCGGAACGCAAUAGCCGAAAA CAAAAAACAAAAAAACAAAAAAACCAAAAAACAAAAACAmmmmmmmmm mmnnnnnnnnnnAAAAACAAAAACAAAAACGGCUAUUAUGCGUUAACGGCGA GACGCUACGGACUU
71	Bacteriophage T4 Td	GGAGACCCUCGACCGUCGAUUGUCCACUGGUCGGGAGACCCUCGAAUGGAAU GGUUCUACAUAUAAUGCCUAACGACUAUCCCUUUGGGGAGUAGGGUCAAGUGAC UCGAAACGAUAGACAACUUGCUUUAACAAGUUGGAGAUUAGUCUGCUCUGCA UGGUGACAUGCAGCUGGAUAUAAUCCGGGGUAGAUAUACGACCUUAUCUGA ACAUAAUGCUAAAAUCCGUUGACCUUAAACGGUCGUGUGGGUUAAGUCCUC CACCCCCACGCCGGAACGCAAUAGCCGAAAAACAAAAACAAAAAAACAAA AAAAAACCAAAAAACAAAACACAmmmmmmmmmnnnnnnnnnnAAAAACA AAAAACAAAACGGCUAUUAUGCGUUAACGGCGAGACGCUACGGAGGU
84	GFP	atggtgagcaagggcgaggagctgttcacggggtggtgcccatactggtcga gctggacggcgacgttaacggccacaagttcagcgtgtccggcgagggcgagg gcgatgccacctacggcaagctgacctgaagttcatctgcaccaccggcaag ctgcccgtgcccggcccacctcgtgaccacctgacctacggcggtgcagtg

		cttcagccgctaccccgaccacatgaagcagcagcacttcttcaagtccgcca tgcccgaaggctacgtccaggagcgcaccatcttcttcaaggacgacggcaac tacaagaccgcgcgcgaggtgaagttcgagggcgacaccctggtgaaccgcat cgagctgaagggcatcgacttcaaggaggacggcaacatcctggggcacaagc tggagtacaactacaacagccacaacgtctatatcatggccgacaagcagaag aacggcatcaaggtgaacttcaagatccgccacaacatcgaggacggcagcgt gcagctcgccgaccactaccagcagaacacccccatcgggcagcggccccgtgc tgctgcccgacaaccactacctgagcaccagtcggccctgagcaaagacccc aacgagaagcgcgatcacatggtcctgctggagtctcgtagccgcgcgggat cactctcgccatggacgagctgtacaagtaa
--	--	---

Examples

[0359] The following experiments (Examples 1-4) describe pre-circRNA constructs and methods of making circRNAs using pre-circRNA constructs. The non-limiting examples of pre-circRNA constructs include, but are not limited to: the constructs as exemplified in **Fig. 1A – 1E**.

Example 1: generation of the DNA genetic constructs

[0360] DNA genetic constructs comprising a 3' intronic sequence and a 5' intronic sequence of SEQ ID NO 1-5 were chemically synthesized in the form of double-stranded DNA.

[0361] The genetic constructs were cloned in a DNA vector (pUC plasmid vector). The genetic constructs comprise SEQ ID NOs: 42-51. The intron performs a catalytic splicing reaction that removes the 3' intronic sequence and the 5' intronic sequence from the transcribed RNA sequence and circularizes the exons sequence that is bookended by the introns.

[0362] The mechanism the circularization reaction is as follows:

- 1) a free exogenous guanosine (GTP) binds to the catalytic core in the permuted 3' intron,
- 2) the guanosine performs a nucleophilic attack on the 5' splice site of the permuted 5' intron, cleaving the intron from the exonic sequence, and
- 3) the 3'-OH of the uracil located at the 5' splice site then performs a nucleophilic attack of the phosphate between the intron and the omega guanosine – the terminal guanosine in the exonic sequence – at the 3' splice site, cleaving the 3' intron from the sequence and ligating the exon ends together into a circular RNA (**Fig. 2**).

[0363] The intron comprised in the generated genetic constructs are a group I introns. The 5' splice site, the 3' splice site, and the catalytic core are all involved in coordinating the sequential cleavage-ligation reactions of group I introns. The exogenous G first binds the catalytic core, also known as the G-site, in a base triple with the intron. Magnesium ions help coordinate the nucleophilic attack on the phosphate backbone in the 5' splice site and later with the 3' splice site. The 5' splice site is located in a region known as the P1 helix, or P1 domain. Intron and exonic sequences base pair with each other in this region, which creates the specificity needed for proper

cleavage at the 5' splice site. The intron component is known as the internal guide sequence (IGS) – 5 bp of complementarity to the exonic sequence and a G across from the uracil at the 5' splice site. The IGS sequence is the exonic sequence complementary to the IGS.

[0364] The group I introns comprised in the genetic constructs of the disclosure derive from tRNA^{LEU} of *Anabaena azollae* (SEQ ID NO: 4), the tRNA^{ILE} of *Azoarcus* (SEQ ID NO: 3), or modified Phage T4 td (SEQ ID NO: 5). *Anabaena PCC7120* (SEQ ID NO: 1) were used as controls.

[0365] The 3' and 5' intronic sequences of the group I introns were cloned in a plasmid vector (**FIGs. 3A-D**), flanking an exonic sequence that contains a 5' spacer sequence, a 5' polyA sequence, an IRES, a GFPencoding nucleotide sequence, a 3' polyA sequence, and a 3' spacer sequence.

Example 2: Efficiency of Circularization

[0366] The efficiency of circularization was assayed by gel electrophoresis. Given their different spatial conformation and different molecular weight, linear and circularized RNA can be identified as distinct bands on the electrophoresis gel. The transcribed RNAs were assayed pre- and post- circularization in the presence of GTP. Genetic constructs comprising intronic sequences derived from *Anabaena azollae*, *Azoarcus*, or modified Phage T4 td introns showed a higher circularization efficiency compared to genetic constructs comprising intronic sequences derived from *Anabaena PCC7120*. Additionally, Genetic constructs comprising intronic sequences derived from *Anabaena azollae*, *Azoarcus*, or modified Phage T4 td introns showed a reduced formation of concatemers compared to constructs comprising intronic sequences derived from *Anabaena PCC7120* permuted introns (**Table 16**, and **FIGs. 4, 5A-D**).

Table 16. Efficiency of Circularization

Intron	Azo*	Ana Az*	Td*	Ana
mRNA name	m2008	m2076	m2077	m852
Unspliced size (bp)	1911	1995	2011	1972
Spliced size (bp)	1637	1637	1639	1637
Intron size (bp)	232	189	220	163
	42	172	155	154
Post GTP circ				
% spliced	84.53	88.78	85.28	67.38
% unspliced	13.34	9.52	0.81	20.09
% concatemers	2.13	1.70	9.85	12.53
% nicked	-	-	4.06	-

Example 3: Expression Efficiency of the Circularized RNAs

[0367] To assess the expression efficiency of the generated circularized RNAs, HEK293T cells were transfected with circularized RNAs synthesized using genetic constructs comprising intronic sequences derived from *Azoarcus*, *Anabaena azollae*, modified T4 phage Td, and *Anabaena PCC7120*. GFP expression was evaluated by flow cytometry. Circularized RNAs derived from genetic constructs comprising intronic sequences derived from *Azoarcus*, *Anabaena azollae*, modified T4 phage Td, and *Anabaena PCC7120* all showed comparable GFP expression levels (**FIG. 6**).

[0368] These data demonstrated the feasibility of synthesizing circularized RNAs with intronic sequences derived from *Azoarcus*, *Anabaena azollae*, T4 phage Td, previously not used for this scope. Additionally, the intronic sequences derived from *Azoarcus*, *Anabaena azollae*, T4 phage Td showed advantages compared to intronic sequences derived from the known *Anabaena PCC7120*. Intronic sequences derived from *Azoarcus*, *Anabaena azollae*, T4 phage Td permuted introns showed increased transcription and circularization efficiency, and a reduced concatemers formation.

[0369] It is to be appreciated that the Detailed Description section, and not the Summary and Abstract sections, is intended to be used to interpret the claims. The Summary and Abstract sections can set forth one or more but not all exemplary aspects of the present disclosure as contemplated by the inventor(s), and thus, are not intended to limit the present disclosure and the appended claims in any way.

[0370] The present disclosure has been described above with the aid of functional building blocks illustrating the implementation of specified functions and relationships thereof. The

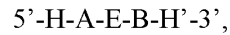
boundaries of these functional building blocks have been arbitrarily defined herein for the convenience of the description. Alternate boundaries can be defined so long as the specified functions and relationships thereof are appropriately performed.

[0371] The foregoing description of the specific aspects will so fully reveal the general nature of the disclosure that others can, by applying knowledge within the skill of the art, readily modify and/or adapt for various applications such specific aspects, without undue experimentation, without departing from the general concept of the present disclosure. Therefore, such adaptations and modifications are intended to be within the meaning and range of equivalents of the disclosed aspects, based on the teaching and guidance presented herein. It is to be understood that the phraseology or terminology herein is for the purpose of description and not of limitation, such that the terminology or phraseology of the present specification is to be interpreted by the skilled artisan in light of the teachings and guidance

[0372] The breadth and scope of the present disclosure should not be limited by any of the above-described exemplary aspects, but should be defined only in accordance with the following claims and their equivalents.

What is claimed is:

1. A genetic construct comprising a formula of:



wherein H comprise a 5' homology arm, A comprises a 3' intronic sequence, E comprises an exonic sequence, B comprises a 5' intronic sequence comprising an internal guide sequence (IGS), and H' comprises a 3' homology arm, wherein the 3' intronic sequence comprises a catalytic core, wherein the 3' terminus of A and the 5' terminus of E form a 3' splice site and the 3' terminus of E and 5' terminus of B form a 5' splice site, and wherein the genetic construct is capable of forming a circular RNA that does not contain any exonic sequence from the genome from which the 5' intronic sequence and the 3' intronic sequence are derived ("excision scar").

2. The genetic construct of claim 1, wherein the 5' intronic sequence and/or the 3' intronic sequence are derived from a group I intron.
3. The genetic construct of claim 1 or 2, wherein the 3' intronic sequence and the 5' intronic sequence together are less than the full intron.
4. The genetic construct of any one of claims 1 to 3, wherein the 5' intronic sequence and/or the 3' intronic sequence are derived from an intron of *thymidylate synthase* (td) from T4 bacteriophage; of pre-tRNA^{LEU} of the cyanobacteria *Anabaena azollae*, *Anabaena* PCC7120, *Aphanizomenon flos-aquae* strain NIVA-CYA 142, or of *Synechococcus*; of pre-tRNA^{ILE} of the nitrogen-fixing bacteria *Azoarcus*; of 26S ribosomal RNA from *Tetrahymena thermophila*; of *COX2* gene from *Trimorphomyces papilionaceus*; or any combination thereof.
5. The genetic construct of any one of claims 1 to 4, wherein the 5' intronic sequence and/or the 3' intronic sequence are derived from an intron of *thymidylate synthase* (td) from T4 bacteriophage.
6. The genetic construct of any one of claims 1 to 5, wherein the 5' splice site comprises a site between two nucleotides comprising U and N, where N is any nucleotide.
7. The genetic construct of claim 6, wherein the N is A, T, G, or C.
8. The genetic construct of any one of claims 1 to 7, wherein the 3' splice site comprises a site between two nucleotides comprising G and N, wherein N is any nucleotide.

9. The genetic construct of claim 8, wherein the N is C.
10. The genetic construct of any one of claims 5 to 9, wherein the catalytic core comprises ACG, CU, or both.
11. The genetic construct of any one of claims 1 to 10, wherein the excision scar comprises SEQ ID NO: 80 or SEQ ID NO: 82.
12. The genetic construct of any one of claims 1 to 10, wherein the excision scar comprises SEQ ID NO: 81 or SEQ ID NO: 83.
13. The genetic construct of any one of claims 1 to 11, wherein the excision scar comprises SEQ ID NO: 80 and SEQ ID NO: 82.
14. The genetic construct of any one of claims 1 to 10, and 12, wherein the excision scar comprises SEQ ID NO: 81 and SEQ ID NO: 83.
15. The genetic construct of any one of claims 1 to 14, wherein the 3' intronic sequence comprises a nucleotide sequence having at least about 70%, at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, at least about 99%, or about 100% sequence identity with a 3' intronic sequence selected from Table 4 (SEQ ID NOs: 11 – 22).
16. The genetic construct of any one of claims 1 to 4 and 15, wherein the 5' intronic sequence comprises a nucleotide sequence having at least about 70%, at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 95%, at least about 95%, at least about 97%, at least about 98%, at least about 99%, or about 100% sequence identity with a 5' intronic sequence selected from Table 4 (SEQ ID NOs: 23-30).
17. The genetic construct of any one of claims 1 to 4, wherein the 5' intronic sequence and/or the 3' intronic sequence are derived from an intron of *tRNA^{LEU}* of the cyanobacteria *Anabaena azollae*.
18. The genetic construct of claim 17, wherein the 5' splice site comprises a site between two nucleotides comprising U and N, wherein N is any nucleotide.
19. The genetic construct of claim 18, wherein the N is A.

20. The genetic construct of any one of claims 17 to 19, wherein the 3' splice site comprises a site between two nucleotides comprising G and N, wherein N is any nucleotide.
21. The genetic construct of claim 20, wherein the N is C.
22. The genetic construct of any one of claims 17 to 21, wherein the catalytic core comprises GAG, CC, or both.
23. The genetic construct of any one of claims 17 to 22, wherein the 3' intronic sequence comprises a nucleotide sequence having at least about 70%, at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 95%, at least about 95%, at least about 97%, at least about 98%, at least about 99%, or about 100% sequence identity with a 3' intronic sequence in Table 4, for example, SEQ ID NO: 14 or 20.
24. The genetic construct of any one of claims 17 to 23, wherein the 5' intronic sequence comprises a nucleotide sequence having at least about 70%, at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 95%, at least about 95%, at least about 97%, at least about 98%, at least about 99%, or about 100% sequence identity with a 5' intronic sequence in Table 4, for example, SEQ ID NO: 25 or 29.
25. The genetic construct of any one of claims 1 to 4, wherein the 5' intronic sequence and the 3' intronic sequence are derived from an intron of *tRNA^{LEU}* of *Azoarcus*.
26. The genetic construct of claim 25, wherein the 5' splice site comprises a site between two nucleotides comprising U and N, wherein the N is any nucleotide.
27. The genetic construct of claim 26, wherein the N is A.
28. The genetic construct of any one of claims 25 to 27, wherein the 3' splice site comprises a site between two nucleotides comprising G and N, wherein N is any nucleotide.
29. The genetic construct of claim 28, wherein the N is G.
30. The genetic construct of any one of claims 25 to 29, wherein the catalytic core comprises GAG, CC, or both.
31. The genetic construct of any one of claims 25 to 30, wherein the 3' intronic sequence comprises a nucleotide sequence having at least about 70%, at least about 75%, at least about 80%,

at least about 85%, at least about 90%, at least about 95%, at least about 95%, at least about 97%, at least about 98%, at least about 99%, or about 100% sequence identity with a 3' intronic sequence in Table 4, for example, SEQ ID NO: 13 or 19.

32. The genetic construct of any one of claims 25 to 31, wherein the 5' intronic sequence comprises a nucleotide sequence having at least about 70%, at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 95%, at least about 95%, at least about 97%, at least about 98%, at least about 99%, or about 100% sequence identity with a 5' intronic sequence in Table 4, for example, nucleotide sequence AAA.

33. The genetic construct of any one of claims 1 to 4, wherein the 5' intronic sequence and the 3' intronic sequence are derived from an intron of 26S ribosomal RNA from *Tetrahymena thermophila*.

34. The genetic construct of claim 33, wherein the 5' splice site comprises a site between two nucleotides comprising U and N, wherein N is any nucleotide.

35. The genetic construct of claim 34, wherein the N is A.

36. The genetic construct of any one of claims 33 to 35, wherein the 3' splice site comprises a site between two nucleotides comprising G and N, wherein the N is any nucleotide.

37. The genetic construct of claim 36, wherein the N is U.

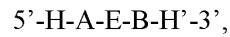
38. The genetic construct of any one of claims 33 to 37, wherein the catalytic core comprises a catalytic core selected from Table 5.

39. The genetic construct of any one of claims 33 to 38, wherein the 3' intronic sequence comprises a nucleotide sequence having at least about 70%, at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 95%, at least about 95%, at least about 97%, at least about 98%, at least about 99%, or about 100% sequence identity with a 3' intronic sequence in Table 4, for example, SEQ ID NO: 12 or 18.

40. The genetic construct of any one of claims 33 to 39, wherein the 5' intronic sequence comprises a nucleotide sequence having at least about 70%, at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 95%, at least about 95%, at least about 97%,

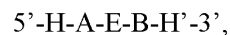
at least about 98%, at least about 99%, or about 100% sequence identity with a 5' intronic sequence in Table 4, for example, SEQ ID NO: 24 or 28.

41. A genetic construct comprising a formula of:



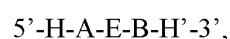
wherein H comprise a 5' homology arm, A comprises a 3' intronic sequence, E comprises an exonic sequence, B comprises a 5' intronic sequence comprising an internal guide sequence (IGS), and H' comprises a 3' homology arm, wherein the 3' intronic sequence comprises a catalytic core, wherein the 3' terminus of A and the 5' terminus of E form a 3' splice site and the 3' terminus of E and 5' terminus of B form a 5' splice site, wherein the genetic construct is capable of forming a circular RNA that does not contain any exonic sequence from the genome from which the 5' intronic sequence and the 3' intronic sequence are derived ("excision scar"), and wherein the genetic construct comprises a nucleotide sequence having at least about 70%, at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 95%, at least about 95%, at least about 97%, at least about 98%, at least about 99%, or about 100% sequence identity with SEQ ID NO: 113 or 123.

42. A genetic construct comprising a formula of:



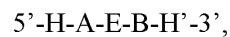
wherein H comprise a 5' homology arm, A comprises a 3' intronic sequence, E comprises an exonic sequence, B comprises a 5' intronic sequence comprising an internal guide sequence (IGS), and H' comprises a 3' homology arm, wherein the 3' intronic sequence comprises a catalytic core, wherein the 3' terminus of A and the 5' terminus of E form a 3' splice site and the 3' terminus of E and 5' terminus of B form a 5' splice site, wherein the genetic construct is capable of forming a circular RNA that does not contain any exonic sequence from the genome from which the 5' intronic sequence and the 3' intronic sequence are derived ("excision scar"), and wherein the genetic construct comprises a nucleotide sequence having at least about 70%, at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 95%, at least about 95%, at least about 97%, at least about 98%, at least about 99%, or about 100% sequence identity with SEQ ID NO: 114 or 124.

43. A genetic construct comprising a formula of:



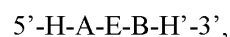
wherein H comprise a 5' homology arm, A comprises a 3' intronic sequence, E comprises an exonic sequence, B comprises a 5' intronic sequence comprising an internal guide sequence (IGS), and H' comprises a 3' homology arm, wherein the 3' intronic sequence comprises a catalytic core, wherein the 3' terminus of A and the 5' terminus of E form a 3' splice site and the 3' terminus of E and 5' terminus of B form a 5' splice site, wherein the genetic construct is capable of forming a circular RNA that does not contain any exonic sequence from the genome from which the 5' intronic sequence and the 3' intronic sequence are derived ("excision scar"), and wherein the genetic construct comprises a nucleotide sequence having at least about 70%, at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 95%, at least about 95%, at least about 97%, at least about 98%, at least about 99%, or about 100% sequence identity with SEQ ID NO: 115 or 125.

44. A genetic construct comprising a formula of:



wherein H comprise a 5' homology arm, A comprises a 3' intronic sequence, E comprises an exonic sequence, B comprises a 5' intronic sequence comprising an internal guide sequence (IGS), and H' comprises a 3' homology arm, wherein the 3' intronic sequence comprises a catalytic core, wherein the 3' terminus of A and the 5' terminus of E form a 3' splice site and the 3' terminus of E and 5' terminus of B form a 5' splice site, wherein the RNA construct is capable of forming a circular RNA that does not contain any exonic sequence from the genome from which the 5' intronic sequence and the 3' intronic sequence are derived ("excision scar"), and wherein the genetic construct comprises a nucleotide sequence having at least about 70%, at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 95%, at least about 95%, at least about 97%, at least about 98%, at least about 99%, or about 100% sequence identity with SEQ ID NO: 116 or 126.

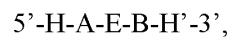
45. A genetic construct comprising a formula of:



wherein H comprise a 5' homology arm, A comprises a 3' intronic sequence, E comprises an exonic sequence, B comprises a 5' intronic sequence comprising an internal guide sequence (IGS), and H' comprises a 3' homology arm, wherein the 3' intronic sequence comprises a catalytic core, wherein the 3' terminus of A and the 5' terminus of E form a 3' splice site and the 3' terminus of E and 5' terminus of B form a 5' splice site, wherein the genetic construct is capable of forming a

circular RNA that does not contain any exonic sequence from the genome from which the 5' intronic sequence and the 3' intronic sequence are derived ("excision scar"), and wherein the genetic construct comprises a nucleotide sequence having at least about 70%, at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 95%, at least about 95%, at least about 97%, at least about 98%, at least about 99%, or about 100% sequence identity with SEQ ID NO: 117 or 127.

46. A genetic construct comprising a formula of



wherein H comprises a 3' homology arm, A comprises a 3' homology arm, a 3' intronic sequence, E comprise an exonic sequence, B comprises a 5' intronic sequence comprising an internal guide sequence, and H' comprising a 5' homology arm, wherein the 3' intronic sequence comprises a catalytic core, wherein the 3' terminus of A and the 5' terminus of E form a 3' splice site and the 3' terminus of E and 5' terminus of B form a 5' splice site, and wherein the 5' intronic sequence and/or the 3' intronic sequence are derived from pre-tRNA^{LEU} of the cyanobacteria *Anabaena azollae*, cyanobacteria *Azoarcus*, or *Synechococcus* or from 26S ribosomal RNA of *Tetrahymena thermophila*, or and are capable of forming a circular RNA.

47. The genetic construct of any one of claims 1 and 41 to 46, wherein the 5' intronic sequence and/or the 3' intronic sequences are derived from tRNA^{LEU} of the cyanobacteria *Anabaena azollae*.

48. The genetic construct of claim 47, wherein the 5' splice site comprises a site between two nucleotides comprising U and any nucleotide (N), optionally A.

49. The genetic construct of claim 47 or 48, wherein the 3' splice site comprises a site between two nucleotides comprising G and any nucleotide (N), optionally C.

50. The genetic construct of any one of claim 47 to 49, wherein the catalytic core comprises GAG, CC, or both.

51. The genetic construct of any one of claims 47 to 50, wherein the 5' intronic sequence comprises a nucleotide sequence having at least about 70%, at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 95%, at least about 95%, at least about 97%, at least about 98%, at least about 99%, or about 100% sequence identity with SEQ ID NO: 25 or 29.

52. The genetic construct of any one of claims 47 to 51, wherein the 3' intronic sequence comprises a nucleotide sequence having at least about 70%, at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 95%, at least about 95%, at least about 97%, at least about 98%, at least about 99%, or about 100% sequence identity with SEQ ID NO: 14 or 20.
53. The genetic construct of any one of claims 1 and 41 to 46, wherein the 5' intronic sequence and/or the 3' intronic sequences are derived from tRNA^{LEU} of the cyanobacteria *Azoarcus*.
54. The genetic construct of claim 53, wherein the 5' splice site comprises a site between two nucleotides comprising U and any nucleotide, optionally A.
55. The genetic construct of claim 53 or 54, wherein the 3' splice site comprises a site between two nucleotides comprising G and any nucleotide, optionally C.
56. The genetic construct of any one of claim 53 to 55, wherein the catalytic core comprises GAG, CC, or both.
57. The genetic construct of any one of claims 53 to 56, wherein the 5' intronic sequence comprises a nucleotide sequence of AAA.
58. The genetic construct of any one of claims 53 to 57, wherein the 3' intronic sequence comprises a nucleotide sequence having at least about 70%, at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 95%, at least about 95%, at least about 97%, at least about 98%, at least about 99%, or about 100% sequence identity with SEQ ID NO: 13 or 19.
59. The genetic construct of any one of claims 1 and 41 to 46, wherein the 5' intronic sequence and the 3' intronic sequences are derived from tRNA^{LEU} of *Synechococcus*.
60. The genetic construct of claim 59, wherein the 5' splice site comprises a site between two nucleotides comprising U and any nucleotide (N), optionally A.
61. The genetic construct of claim 59 or 60, wherein the 3' splice site comprises a site between two nucleotides comprising G and any nucleotide (N), optionally C.
62. The genetic construct of any one of claim 59 to 61, wherein the catalytic core comprises GAG, CC, or both.

63. The genetic construct of any one of claims 59 to 62, wherein the 5' intronic sequence comprises a nucleotide sequence having at least about 70%, at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 95%, at least about 95%, at least about 97%, at least about 98%, at least about 99%, or about 100% sequence identity with a 5' intronic sequence derived from *Synechococcus* intron as set forth in SEQ ID NO: 140 or 143.

64. The genetic construct of any one of claims 59 to 63, wherein the 3' intronic sequence comprises a nucleotide sequence having at least about 70%, at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 95%, at least about 95%, at least about 97%, at least about 98%, at least about 99%, or about 100% sequence identity with a 5' intronic sequence derived from *Synechococcus* intron as set forth in SEQ ID NO: 140 or 143.

65. The genetic construct of any one of claims 1 and 41 to 46, wherein the 5' intronic sequence and the 3' intronic sequences are derived from 26S ribosomal RNA of *Tetrahymena thermophila*.

66. The genetic construct of claim 65, wherein the 5' splice site comprises a site between two nucleotides comprising U and any nucleotide (N), optionally A.

67. The genetic construct of claim 65 or 66, wherein the 3' splice site comprises a site between two nucleotides comprising G and any nucleotide (N), optionally C.

68. The genetic construct of any one of claim 65 to 67, wherein the catalytic core comprises GAG, CC, or both.

69. The genetic construct of any one of claims 65 to 68, wherein the 5' intronic sequence comprises a nucleotide sequence having at least about 70%, at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 95%, at least about 95%, at least about 97%, at least about 98%, at least about 99%, or about 100% sequence identity with SEQ ID NO: 24 or 28.

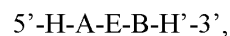
70. The genetic construct of any one of claims 65 to 69, wherein the 3' intronic sequence comprises a nucleotide sequence having at least about 70%, at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 95%, at least about 95%, at least about 97%, at least about 98%, at least about 99%, or about 100% sequence identity with SEQ ID NO: 12 or 18.

71. The genetic construct of any one of claims 46 to 70, wherein the circular RNA does not comprise an excision scar.

72. The genetic construct of any one of claims 1 to 71, wherein the circular RNA has a reduced immunogenicity.

73. The genetic construct of claim 72 wherein the circular RNA has a reduced immunogenicity compared to a circular RNA that contains the excision scar.

74. A genetic construct comprising a formula of



wherein H comprises a 3' homology arm, A comprises a 3' homology arm, a 3' intronic sequence, E comprise an exonic sequence, B comprises a 5' intronic sequence comprising an internal guide sequence, and H' comprising a 5' homology arm, wherein the 3' intronic sequence comprises a catalytic core, wherein the 3' terminus of A and the 5' terminus of E form a 3' splice site and the 3' terminus of E and 5' terminus of B form a 5' splice site, and wherein the 5' intronic sequence and/or the 3' intronic sequence are derived from pre-tRNA^{LEU} of the cyanobacteria *Anabaena azollae*, cyanobacteria *Azoarcus*, or *Synechococcus* or from 26S ribosomal RNA of *Tetrahymena thermophila* and are capable of forming a circular RNA and wherein the genetic construct comprises a nucleotide sequence having at least about 70%, at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 95%, at least about 95%, at least about 97%, at least about 98%, at least about 99%, or about 100% sequence identity with SEQ ID NO: 118 or 128.

75. A genetic construct comprising a formula of:



wherein A comprises a 3' intronic sequence, E comprises an exonic sequence, wherein the 3' intronic sequence comprises a catalytic core, wherein the 3' terminus of A and the 5' terminus of E form a 3' splice site, and wherein the genetic construct is capable of forming a circular RNA.

76. A genetic construct comprising a formula of:



wherein A comprises a 3' intronic sequence, E comprises an exonic sequence, wherein the 3' intronic sequence comprises a catalytic core, wherein the 3' terminus of A and the 5' terminus of

E form a 3' splice site, wherein the 3' terminus of A and the 5' terminus of E form a 3' splice site, wherein the genetic construct comprises a nucleotide sequence having at least about 70%, at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 95%, at least about 95%, at least about 97%, at least about 98%, at least about 99%, or about 100% sequence identity with SEQ ID NO: 119 or 129, and wherein the genetic construct is capable of forming a circular RNA.

77. A genetic construct comprising a formula of



wherein A comprises a 3' intronic sequence comprising a catalytic core (CC), and E comprises an exonic sequence comprising a first portion of the exonic sequence (E'); a first homology arm (HI'); a second portion of the exonic sequence (E''); a second homology arm (HI''); and a third exonic sequence (E''') comprising a catalytic nucleotide (CN); wherein the 3' terminus of A and the 5' terminus of E form a 3' splice site, wherein the genetic construct comprises a nucleotide sequence having at least about 70%, at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 95%, at least about 95%, at least about 97%, at least about 98%, at least about 99%, or about 100% sequence identity with SEQ ID NO: 120 or 130, and wherein the genetic construct is capable of forming a circular RNA.

78. A genetic construct comprising a formula of



wherein A comprises a 3' intronic sequence comprising a catalytic core (CC), and E comprises an exonic sequence comprising a first portion of the exonic sequence (E'); a first homology arm (HI'); a second portion of the exonic sequence (E''); a second homology arm (HI''); and a third exonic sequence (E''') comprising a catalytic nucleotide (CN); wherein the 3' terminus of A and the 5' terminus of E form a 3' splice site, wherein the genetic construct comprises a nucleotide sequence having at least about 70%, at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 95%, at least about 95%, at least about 97%, at least about 98%, at least about 99%, or about 100% sequence identity with SEQ ID NO: 121 or 131, and wherein the genetic construct is capable of forming a circular RNA.

79. A genetic construct comprising a formula of



wherein H comprises a 3' homology arm, A comprises a 3' intronic sequence comprising a catalytic core (CC), and E comprises an exonic sequence comprising a first portion of the exonic sequence (E'); a first homology arm (HI'); a second portion of the exonic sequence (E''); a second homology arm (HI''); and a third exonic sequence (E''') comprising a catalytic nucleotide (CN); wherein the 3' terminus of A and the 5' terminus of E form a 3' splice site, wherein the genetic construct comprises a nucleotide sequence having at least about 70%, at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 95%, at least about 95%, at least about 97%, at least about 98%, at least about 99%, or about 100% sequence identity with SEQ ID NO: 122 or 132, and wherein the genetic construct is capable of forming a circular RNA.

80. A genetic construct comprising a formula of



wherein A comprises a 3' intronic sequence comprising a catalytic core (CC), and E comprises an exonic sequence comprising a first portion of the exonic sequence (E'); a first homology arm (HI'); a second portion of the exonic sequence (E''); a second homology arm (HI''); and a third exonic sequence (E''') comprising a catalytic nucleotide (CN); wherein the genetic construct comprises a nucleotide sequence having at least about 70%, at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 95%, at least about 95%, at least about 97%, at least about 98%, at least about 99%, or about 100% sequence identity with SEQ ID NO: 181 and wherein the genetic construct is capable of forming a circular RNA.

81. A genetic construct comprising a formula of



wherein A comprises a 3' intronic sequence, 3S comprises a 3' splice site, E comprises an exonic sequence, B comprises a 5' intronic sequence, and 5S comprises a 5' splice site, wherein the 3' intronic sequence comprises a catalytic core, wherein B-A comprise the full intron, wherein the genetic construct comprises a nucleotide sequence having at least about 70%, at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 95%, at least about 95%, at least about 97%, at least about 98%, at least about 99%, or about 100% sequence identity with SEQ ID NO: 182 or 183, and wherein the genetic construct is capable of forming a circular RNA.

82. The genetic construct of any one of claims 1 to 81, which is RNA.

83. The genetic construct of any one of claims 1 to 82, which is DNA.

84. The genetic construct of claim 83, which further comprises a promoter.
85. The genetic construct of claim 84, wherein the promoter comprises a T7 promoter, a T3 promoter, or a Sp6 promoter.
86. The genetic construct of claim any one of claims 1-85, which further comprises an IRES.
87. The genetic construct of any one of claims 1 to 86, wherein the exonic sequence further comprises a binding site for a polyA binding protein at the 5' terminus.
88. The genetic construct of any one of claims 1 to 87, wherein the 5' homology arm comprises a nucleotide sequence at least about 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% identical to SEQ ID NO: 72 (CCGTCGATTGTCCACTGGT).
89. The genetic construct of any one of claims 1 to 88, wherein the 3' homology arm comprises a nucleotide sequence at least about 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% identical to SEQ ID NO: 73 (ACCAGTGGACAATCGACGG).
90. The genetic construct of any one of claims 1 to 89, wherein the exonic sequence is linked to one or more untranslated regions at the 3' and/or 5' of the exonic sequence.
91. The genetic construct of claim 90, wherein the one or more untranslated regions comprise IRES.
92. The genetic construct of claim 91, wherein the IRES comprises a sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity with any one of SEQ ID NO: 87-110 and 145-168.
93. The genetic construct of claim 92 wherein the IRES is linked to a polyA.
94. The genetic construct of any one of claims 90-93, wherein the one or more untranslated regions comprise a polyA.
95. The genetic construct of claim 94, wherein the exonic sequence is linked to a polyA.
96. The genetic construct of any one of claims 93-95, wherein the polyA comprises a sequence having at least about 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% identity with SEQ ID NO: 85 or 86.

97. The genetic construct of any one of claims 90-96, wherein the one or more untranslated regions further comprise a spacer sequence.

98. The genetic construct of claim 97, wherein the spacer comprises a sequence having at least about 70%, at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 95%, at least about 95%, at least about 97%, at least about 98%, at least about 99%, or about 100% sequence identity with any one of SEQ ID NO: 74-75 and 78-79.

99. The genetic construct of any one of claims 1 to 98, wherein the exonic sequence encodes a protein.

100. The genetic construct of claim 99, wherein the circular RNA exhibits an increased expression of the protein, a prolonged expression of the protein in vitro and/or in vivo, a reduced activation of an immune signaling pathway, an increased immune activation, or any combination thereof.

101. The genetic construct of claim 99 or 100, wherein the protein comprises a therapeutic protein.

102. The genetic construct of any one of claims 99 to 101, wherein the protein comprises a cytokine, a receptor, a ligand, an immunomodulatory, a growth factor, an antigen, an RNA binding domain, or any combination thereof.

103. The genetic construct of any one of claims 99 to 102, wherein the protein comprises an antibody, a fusion protein, or any combination thereof.

104. The genetic construct of any one of claims 99 to 103, wherein the protein comprises CD3, CD4, CDS, CD19, CD20, CD22, CD30, CD34, HER2, HER3, HER4, LFA-I, MoI, pl50, 95, VLA-4, ICAM-I, VCAM, alpha v/beta 3 integrin, human macrophage inflammatory protein (MIP-1 - alpha), erythropoietin (EPO), NGF-beta, platelet-derived growth factor (PDGF), fibroblast growth factors, epidermal growth factor (EGF), transforming growth factors (TGF), insulin-like growth factors-I and -II (IGF-I and IGF-II), des(1-3)-IGF-I (brain IGF-1), factor VIII, tissue factor, von Willebrands factor, protein C, alpha-1-antitrypsin, plasminogen activators, such as urokinase and tissue plasminogen activator ("t-PA"), bombazine, thrombin, thrombopoietin, M-CSF, GM-CSF, G-CSF, flk2/flt3 receptor, obesity (OB) receptor, LDL receptor, growth hormone receptors, thrombopoietin receptors ("TPO-R," "c-mpl"), glucagon receptors, interleukin receptors, interferon

receptors, T-cell receptors, stem cell factor receptors, OX40L, bone-derived neurotrophic factor (BDNF), neurotrophin-3, -4, -5, or -6 (NT-3, NT-4, NT-5, or NT-6), relaxin A-chain, relaxin B-chain, prorelaxin, interleukins and interleukin receptors, including IL-1 to IL-33 and IL-1 to IL-33 receptors, such as the IL-8 receptor, viral antigens, lipoproteins, calcitonin, glucagon, atrial natriuretic factor, lung surfactant, tumor necrosis factor-alpha and -beta, enkephalinase, RANTES (regulated on activation normally T-cell expressed and secreted), mouse gonadotropin-associated peptide, DNase, inhibin, activin, PUF domain, Cas protein, or any combination thereof.

105. A composition comprising the genetic construct of any one of claims 1 to 104 and a GTP.

106. The composition of claim 105, wherein the GTP is in a sufficient amount to cleave the 5' intron and the 3' intronic sequences from the genetic construct.

107. A method of making a circular RNA, comprising contacting a GTP and the genetic construct of any one of claims 1 to 104.

108. A circular RNA prepared by the method of claim 107.

109. A method of treating a disease or condition in a subject in need thereof, comprising administering the circular RNA of claim 108 to the subject.