©2019

JUAN CARLOS COLLANTES VELA

ALL RIGHTS RESERVED

A Novel CRISPR/RNA-Aptamer-Mediated Base Editing System with Potential Therapeutic Value

By

JUAN CARLOS COLLANTES VELA

A dissertation submitted to the School of Graduate Studies

Rutgers, The State University of New Jersey

In partial fulfillment of the requirements for the degree of Doctor of Philosophy

Graduate Program in Cellular and Molecular Pharmacology

Graduate Programs in Molecular Biosciences

Written under the direction of

Shengkan Jin, PhD

And approved by:

_____

_____

_____

_____

New Brunswick, New Jersey

January 2019

ABSTRACT OF THE DISSERTATION

A Novel CRISPR/RNA-Aptamer-Mediated Base Editing System with Potential Therapeutic Value

by JUAN CARLOS COLLANTES VELA

Dissertation Director: Shengkan Jin

Nuclease-dependent precise genome editing such as correction of point mutations requires introduction of targeted DNA double strand breaks (DSB) and activation of homology dependent repair (HDR), limiting its application to proliferating cells. To expand GE capabilities for therapeutic use in non-dividing somatic cells it is necessary to precisely modify nucleotides avoiding DSBs. Recently, Cas9-cytidine deaminase fusions, also known as based editors (BE), were shown to precisely modify target bases at certain genomic loci. To expand the base editing toolbox, we sought to engineer a novel base editing system based on RNA-aptamer mediated recruitment. To this end, we engineered a nuclease-deficient CRISPR/Cas9 system as a recruitment platform for non-nuclease DNA/RNA editing enzymes that catalyze C·G→T·A conversions by cytidine deamination. Targeted nucleotide modification was achieved with high precision in prokaryotic and eukaryotic cells. In bacteria, we tested our system targeting the rifampicin resistance determining region of the *rpoB* gene. Survival in rifampicin reached over 1000-fold higher than untreated cells. To examine whether the system can correct loss of function mutations in human genome, we treated a stably integrated non-fluorescent *EGFP* gene containing an A·T→G·C mutation on the chromophore sequence.

Fluorescence was efficiently restored in treated cells, detecting around 10% of GFP positive cells after treatment. Next generation sequencing confirmed a G·C→A·T conversion in 60% of reads at the target position, restoring the wild type sequence, with low by-stander effect. Exome-wide sequence analysis revealed no detectable off-target effects. Targeting of endogenous loci also resulted in highly efficient nucleotide conversion at the desired C positions. We also show that our system can destroy the 3' splice acceptor site of intron 50 in human *DMD* gene, potentially inducing exon 51 skipping, providing evidence of a therapeutic application to treat Duchenne muscular dystrophy. Taken together, the data show that our GE system represents a safe and promising technology for editing specific nucleotides, correcting genetic mutations or other clinically relevant applications, independent of DSB and HDR, with potential therapeutic value in non-dividing cells.

ACKNOWLEDGEMENTS

I would like to thank Dr. Shengkan Jin for the wonderful opportunity he has given me to work in this exciting project. With his advice and guidance, this has been the most rewarding professional experience of my career.

I want to also thank the members of my committee, Dr. Debabrata Banerjee, Dr. Joseph Fondell and Dr. Richard Padgett. Their input, advice and positive feedback have been fundamental to guide my work.

This project would not have been possible without the contributions of many friends and colleagues. I want to thank my lab mates that collaborated with me, directly or indirectly, to advance this project. Thanks to Huiting Xu and Victor Tan for their instrumental contribution to lab work. To Dr. Roman Wernyj for his contribution with early experiments that paved the way for the success of this project. To Dr. Hanlin Tao, Dr. Amer Alasadi and Dr. Jingjing Guo for their advice and friendship. To Dr. Katarzyna Tyc and Dr. Jinchuan Xing for their contribution for some of the bioinformatics work. To Husam Al-hraishawi for developing the stable nfEGFP cell line.

I would also wish to extend my appreciation to faculty and staff members of the Pharmacology Department, SGS-Biomedical Sciences and Molecular Biosciences for the continuing the support I have received during my time at Rutgers.

I wish to thank my family, my beautiful wife Meli, my son Gabriel and my daughter Elena (on her way to this world) for their endless love, support and motivation to keep me going through this hard but beautiful journey. I am also grateful to my parents, Carlos

and Ana María, my siblings, Daniela and Sebastian, my cousin Andrés, and my grandma and one of my best friends, Chivita, for always supporting me and encouraging me to pursue this dream.

Finally, this work was supported by Fulbright – SENESCYT grant No. 15130816, US NIH/NCI R21 CA216604, and Rutgers University – School of Graduate Studies "Biomedical Completion and Acceleration Fellowship".

Table of Contents

ABSTRACT OF THE DISSERTATION ii

ACKNOWLEDGEMENTS iv

CHAPTER I 1

Introduction and Rationale 1

CHAPTER II 6

Construction of CRC System and Proof of Concept in Prokaryotic Genome 6

Results 7

CRC System: A Modular Base Editing Platform 7

CRC Proof of Concept in Prokaryotic Cells: Editing an Endogenous Bacterial Gene 10

System Optimization: Engineering Individual Modules Enhances Efficiency 13

Discussion 18

Materials and Methods 20

CHAPTER III 23

Construction of CRC System for Mammalian Expression 23

Results 25

CRC system in mammalian cells 25

CRC Corrects Loss of Function Mutation in Genomic DNA 28

Next Generation Sequencing Analysis Reveals High On-Target Base Editing Efficiency and Undetectable Off-Target Effects 32

Discussion 39

Materials and Methods 42

CHAPTER IV 47

Targeting Endogenous Loci in the Human Genome 47

Results 48

CRC is able to edit target Cs in humane genomic loci 48

APOBEC1 as CRC effector has lower efficiency and purity than AID 53

CRC induces abrogation of 3' splice acceptor site in intron 50 of human DMD 55

Discussion 61

Materials and Methods 65

CHAPTER V 66

Final Discussion and Future Directions 66
Future Directions: CRC Optimization for Improved Expression in Human Cells 71
REFERENCES 75
APPENDIX: SEQUENCES OF PLASMIDS DEVELOPED IN THIS STUDY 81
Bacterial Expression Plasmid Sequences 81
1. **pTet_dCas9-bacteria/p15A** 81
2. **pTet_nCas9$_{H840A}$-bacteria/p15A** 85
3. **pTet_nCas9D10A-bacteria/p15A** 89
4. **pTet_AID_MCP/DF13** 93
5. **sgRNA_MS2-bacteria/pUC19** 97
6. **sgRNA_2xMS2-bacteria/pUC19** 99
Mammalian Expression Plasmid Sequences 102
1. **pCMV_nfEGFP** 102
2. **pUbC_^{A}CRCn** 105
3. **pCMV_^{A}CRCnu** 114
4. **pCMV_^{A}CRCvnu** 120
5. **phU6_gRNA_2xMS2** 126
6. **pCMV_^{A}CRCnu.2** 128
7. **pCMV_^{2NA}CRCnu.2** 134
8. **pCMV_$^{N\Delta A}$CRCnu.2** 140
9. **pCMV_2xA_CRCnu.2** 146
10. **pCMV_^{A1}CRCnu.2** 153
11. **phU6_gRNA_6xMS2** 159
12. **phU6_gRNA_2+4xMS2** 162

CHAPTER I

Introduction and Rationale

Humans have had the ability to manipulate and modify genomic sequences for almost half a century, since the advent of recombinant DNA technology.[1] In the early 1990s, Pavletich and Pabo published the first crystal structure of a zinc finger protein interacting with its target DNA.[2] The elucidation of zinc finger DNA recognition mechanism paved the way to design chimeric DNA binding proteins with novel functions.[2,3] By fusing engineered zinc finger domains to nuclease domains from restriction enzymes such as FokI, also known as zinc finger nucleases (ZFN), researchers were able to introduce programmable and highly specific DNA double strand breaks,[4] thereby launching the gene editing era. In the years that followed, many research applications of these customizable designer nucleases were reported.

Right from the start, the potential use of GE technologies in biomedicine was appreciated.[3] Recently several studies, including a phase 1 clinical trial,[5] have successfully shown that GE technologies may be suitable to treat human diseases, including viral infections as well as inherited disorders.[6] However, this possibility may be fully achieved when existing technologies overcome the challenges and limitations they currently face regarding safety, efficacy and delivery.[7,8]

Currently, there are four GE platforms: ZFN, transcription activator-like effector nucleases (TALENs), meganucleases and clustered regularly interspaced short palindromic repeats (CRISPR)-CRISPR associated protein 9 (CRISPR/Cas9).[6,8,9] A common feature among these methods is the introduction of sequence specific DSBs, which in turn activate DNA repair pathways, mainly HDR and non-homologous end joining (NHEJ).[6,8,10] HDR is an error free pathway, and its activation allows for precise sequence modification when a donor template with the desired modification, flanked by homologous sequences to the target region, is delivered together with the designer nuclease.[10,11] Nonetheless, this method can be applied exclusively to actively dividing cells, specifically throughout S/G2 phases , because HDR is absent in post mitotic cells.[6,12] NHEJ is active in mitotic and post mitotic cells throughout the cell cycle. However, it is an error prone pathway that often introduces insertions and deletions of nucleotides (indels) in the target sequence. If the break occur on coding regions of the target gene, the effective outcome is disruption of the ORF, thus deleting the gene.[6] Therefore, when it comes to treating terminally differentiated cells, current designer nuclease platforms are restricted to gene knockout, which limits their application in the clinical setting. By harnessing the DNA targeting features of GE platforms, independently from their nuclease activity, engineering them as delivery systems of non-nuclease DNA interacting enzymes, we hypothesize that some of such limitations could be potentially overcome.

Monogenic disorders could potentially benefit from clinical applications of GE technologies, since the disease phenotype can be reversed by the correction of a single gene in specific tissues.[13] The World Health Organization (WHO) estimates that over 10,000 human diseases are monogenic (WHO, 2018; retrieved from http://www.who.int/genomics/public/geneticdiseases/en/index2.html), and these disorders affect both dividing and differentiated cells. Some of these disorders arise from changes in single nucleotides, known as single nucleotide polymorphisms (SNPs).[14-16] Developing GE platforms capable of introducing precise modifications in terminally differentiated cells, at the single nucleotide level, will allow researchers treat the disease causing mutation on the DNA sequence, potentially moving forward GE therapeutics.

The recent discovery of the DNA targeting capabilities and ease of use of the bacterial and archaeal adaptive immune system, called CRISPR/Cas9,[17,18] has revolutionized the research field and its potential therapeutic applications are being explored.[6,19,20] In contrast to ZFN, TALENs and meganucleases, CRISPR/Cas9 does not require labor-intensive and costly protein engineering processes to program every new target since it is an RNA-guided engineered endonuclease system (RGEN), greatly facilitating GE experimentation.[17,21] CRISPR/Cas9 can be easily programmed to target specific genomic regions provided that strand complementary to the target sequence (protospacer, normally 20 nucleotides in length) is flanked by a 3' protospacer adjacent motif (PAM), which sequences differ depending on the bacterial species.[22] Most laboratories are

engineering *S. pyogenes*' CRISPR system because of its relatively simple PAM sequence requirement, i.e. NGG.[17,22] However, one major drawback of this and other GE platforms is the introduction of off-target, potentially cytotoxic DSBs.[23] Several strategies to improve the specificity and safety of CRISPR have been reported.[24-33] Despite the robustness and versatility of CRISPR systems, a better safety profile is required in order to improve its specificity and its potential clinical applicability.[10,32] For example, specificity is improved when engineered Cas9 nickases (nCas9) are delivered with paired, offset gRNAs targeting opposite DNA strands.[24,31,34] Off-target effects produced from nickase targeting are likely well tolerated, since single strand breaks (SSB) are resolved efficiently by base excision repair pathway (BER)[35-37] or by an alternative HDR pathway,[38] however the latter might be irrelevant for terminally differentiated cells. Similarly, fusion of catalytically dead Cas9 (dCas9) to the monomeric nuclease domain of FokI targeted with paired gRNAs has been reported as an alternative mechanism to improve CRISPR specificity.[27,28,33] Both of these strategies enhance specificity by relying on CRISPR's targeting capabilities without generating DSBs.

Numerous CRISPR/Cas9 nuclease-independent applications have been described thus far, including regulation of gene expression by transcriptional activation,[39-41] repression[42,43] and epigenome modulation,[44,45] visualization of defined genomic loci,[46,47] immunoprecipitation and isolation of specific genomic sequences,[7] among others. These efforts have opened the path to increase the repertoire of CRISPR technology applications beyond generating DSBs. This prompted us and others[48] to engineer *S.*

pyogenes' CRISPR/Cas9 as a targeted recruitment system for several non-nuclease DNA modifying proteins capable of introducing highly specific sequence modifications at the single nucleotide level, giving rise to a new generation of genome engineering application: base editing. Phage RNA scaffolds that are synthetically fused to the 3' end of the gRNA molecule mediate the recruitment mechanism.[40,41,47] Our platform, called CRISPR-RNA-Aptamer-Mediated Correction/Modification (CRC) system, is able to target and modify specific nucleotides in prokaryotic genome, as well as extrachromosomal DNA and endogenous loci in human cells in a programmable fashion, circumventing the generation of DSB. To test CRC in a clinically relevant target, we also aimed the system to abrogate the 3' splice acceptor site of *DMD*'s intron 50 to induce exon 51 skipping in *DMD* gene. This strategy has been shown to restore dystrophin expression in patients diagnosed with Duchenne muscular dystrophy using a morpholino-based antisense oligonucleotides targeting the mRNA,[49-51] inducing a transient effect. Our results show that it is possible to use CRC to permanently delete exon 51, potentially extending the beneficial effect of therapeutic genome editing. Together, the data suggest that CRC system could be a promising therapeutic tool.

CHAPTER II

Construction of CRC System and Proof of Concept in Prokaryotic Genome

CRISPR/Cas9 are modular DNA targeting systems that have been engineered for targeted gene editing[17], as well as many other nuclease-independent applications such as gene expression regulation[41,52], dynamic genomic labeling[46,53], epigenome editing[54,55], and even as diagnostic tools,[56] etc. In this Chapter, a nuclease-deficient CRISPR/Cas9 system derived from *Streptococcus pyogenes* was used as a recruitment platform for activation-induced cytidine deaminase (AID),[57] a DNA editing enzyme that catalyzes C→U (T) conversions in genomic DNA.[58,59] The recruitment mechanism is mediated by engineered gRNAs synthetically fused to RNA-aptamers[41] which bind to specific RNA-interacting proteins fused to AID. We call our system CRISPR-RNA-Aptamer-Mediated Correction/Modification system (CRC; **Figure 1**). We achieved efficient nucleotide conversion in *E. coli*'s genome using a negative selection system (**Figure 2, 3**). These findings represent a proof of concept for a novel gene editing approach and pave the way for the development of mammalian applications.

During the course of this work, several studies were published by David Liu and colleagues describing a similar gene editing approach, named base-editing.[48,60-62] Their systems, dubbed base editors (BE), use cytidine deaminase APOBEC1 as effector module to catalyze C→T conversions. In base editors, the effector proteins are directly tethered to Cas9. Since BE and CRC have the same goal but key differences in terms of

recruitment mechanism and effector modules, we decided to compare them side-by-side in our mammalian cell experiments (Chapters III and IV).

Results

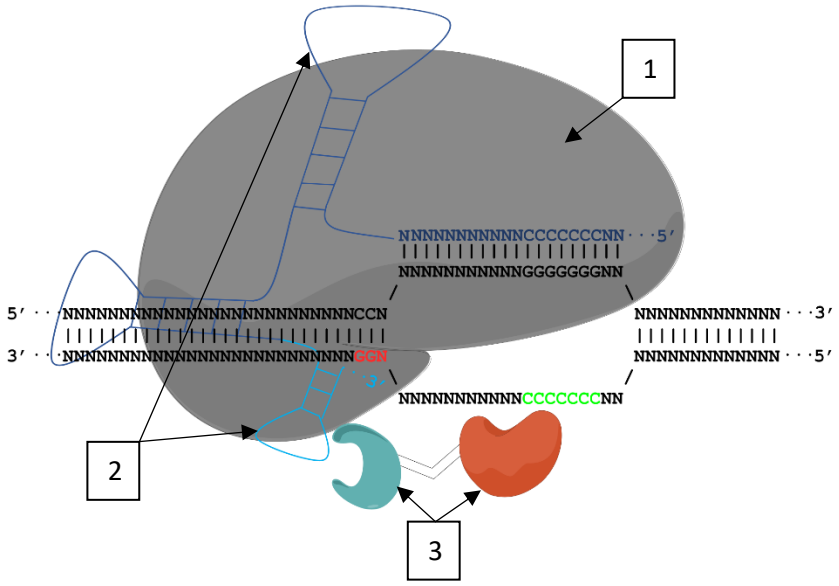
CRC System: A Modular Base Editing Platform

The CRC base editing system contains three functional modules illustrated in **Figure 1A**: (1) a nuclease deficient Cas9 protein; (2) a programmable chimeric RNA scaffold containing gRNA (for sequence recognition and Cas9 binding) and a recruiting RNA motif (aptamer) for effector module recruitment; and (3) the effector module containing a cytidine deaminase (effector) fused to a small RNA binding protein that specifically interacts with the recruiting RNA-aptamer described in (2).

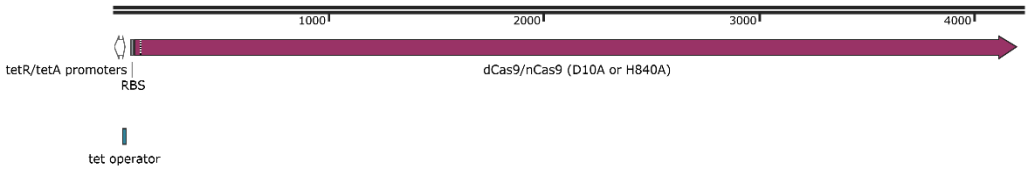
In our prototype system, we used a catalytically dead Cas9 (dCas9) protein derived from *Streptococcus pyogenes* ($nCas9_{D10A}$) as the CRISPR targeting platform. Activation induced cytidine deaminase (AID), an enzyme that catalyzes C→U (T) conversion in DNA context, was used as the effector protein. MS2-MCP (RNA-aptamer – RNA binding protein pair) was used for the RNA mediated effector recruitment.[41] To construct the system, we synthetically fuse MS2 RNA motif to the 3' end of the CRISPR RNA scaffold for recruiting the effector module, AID_MCP fusion protein. As a result, the effector AID

protein could be recruited to any target sequence (specified by the guide RNA) where it could convert a C to U (T). In **Figure 1**, the effector is shown as a monomer; however, in cells, AID forms a dimer at the action site. A potential advantage of the RNA based effector recruitment design compared to a direct fusion of the effector to Cas9[48] is that proteins such as AID and other members of the APOBEC family of cytidine deaminases usually require dimerization or other types of oligomerization for enzymatic activity.[63-65] The AID_MCP fusion monomer protein can form an oligomer relatively efficiently (MCP is a small peptide, 13 kDa) while multimerization for the Cas9-cytidine deaminase fusion is expected to be inefficient (Cas9 is bulky, 160 kDa, and one Cas9-cytidine deaminase monomer is pivoted to DNA).

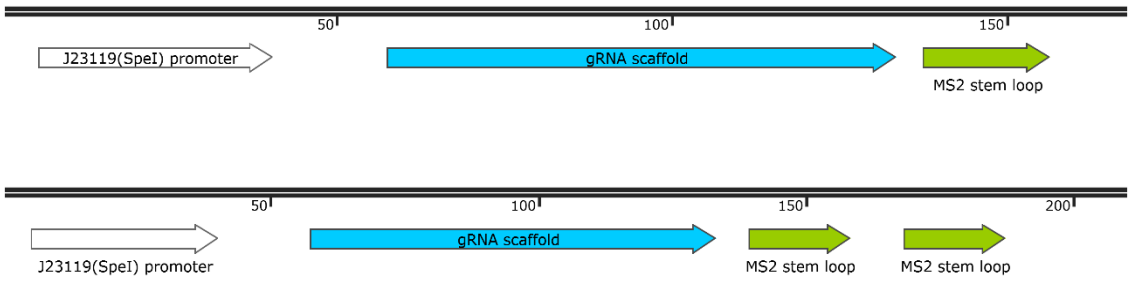
A



B.



C.



D.

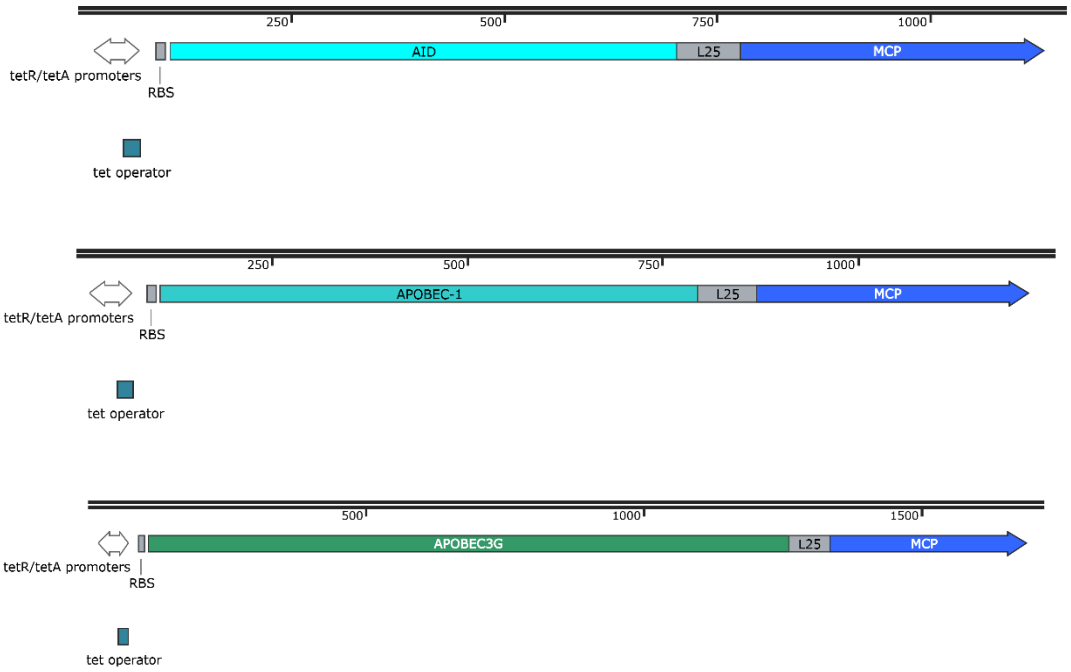


Figure 1. CRC system description

A. Components of the platform: (**1**) DNA targeting module dCas9 or nCas9$_{D10A}$, (**2**) Chimeric RNA scaffold containing a guide RNA motif (for programmable sequence targeting, blue sequence), CRISPR motif (for Cas9 binding, blue scaffold), and recruiting RNA motif (for recruiting effector-RNA binding protein fusion, light blue scaffold), and (**3**) effector-RNA binding fusion protein. In red is shown the PAM sequence. **B – D.** Schematic representation of constructs used in this Chapter, including DNA targeting module encoding for Cas9 variants dCas9, nCas9$_{D10A}$ or nCas9$_{H840A}$ (**B**; component (**1**) in **Figure 1A**); gRNA/recruiting module containing one or two RNA aptamer motifs (**C**, top and bottom, respectively; component (**2**) in **Figure 1A**); and effector module, encoding for fusion proteins AID_MCP, APOBEC1_MCP or APOBEC3G_MCP (**D**, top, middle,

bottom respectively; component (**3**) in **Figure 1A**). For bacterial experiments shown in this chapter, each module is encoded in a single vector.

CRC Proof of Concept in Prokaryotic Cells: Editing an Endogenous Bacterial Gene

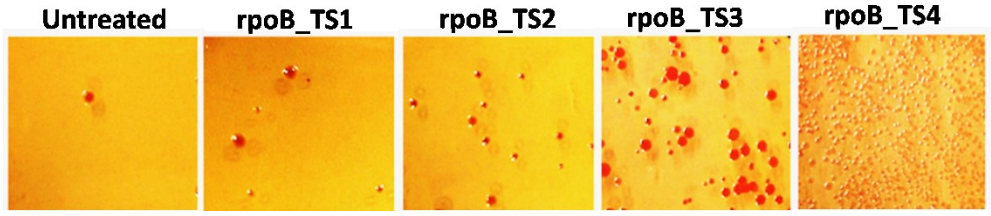
In bacteria, we tested our system using a negative selection approach using the antibiotic rifampicin for selection. This drug binds near the catalytic pocket of the subunit β of bacterial RNA polymerase, encoded by the *rpoB* gene, inhibiting transcription by physically blocking RNA elongation.[66] Well defined mutations along a specific segment of the *rpoB* gene have been associated with rifampicin resistance. This region is known as rifampicin resistance determining region[67] (RRDR; **Figure 2A**). Furthermore, it has been shown that AID is capable of introducing mutations in RRDR, and surviving colonies can be individually assessed for *rpoB* gene mutations,[58] making it an ideal target to test CRC.

Four gRNAs were designed for these experiments targeting the template strand (TS1 – TS4; **Figure 2A, Table 1**), using catalytically dead Cas9 (dCas9) as DNA targeting module and 1xMS2 as recruiting module. The system expressing AID_MCP and dCas9 as effector and targeting modules, respectively, is noted as $^{A}CRC_{d}$. Survival fraction in rifampicin reached 40-fold over the untreated background after CRC treatment with gRNA TS4, while the remaining gRNAs resulted in fewer surviving colonies. The higher efficiency observed in TS4 treated cells might be due to the position of the targeted C within the

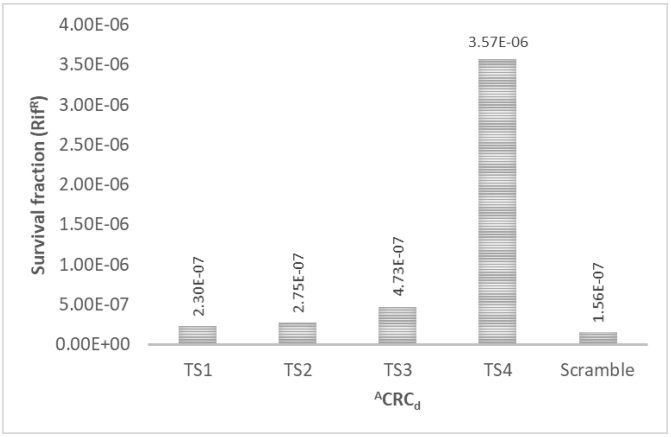
protospacer (the unpaired DNA strand left by CRISPR R-loop), which in this case sits on position 8 from the 5' end. These observations are consistent with base editing activity window, which suggests that for efficient C→T conversion, the relative position of the target C should be between nucleotides 4 and 8 of the unpaired strand.[48] Sequence analysis of isolated colonies treated with $^{A}CRC_{d}$/rpoB_TS4 revealed that the system introduced targeted C→T mutations in codon 531, changing a serine for phenylalanine, a mutation known to induce rifampicin resistance[67,68] (**Figure 2 A – C**). From these results we conclude that CRC's targeting mechanism requires the target C to be on the unpaired DNA strand left by the CRISPR R-loop, closer to the 5' end, as shown in **Figure 1**. Taken together, the data show that targeted nucleotide modification is a feasible approach for gene editing independent of DNA double strand breaks.

A.

B.



C.



D.

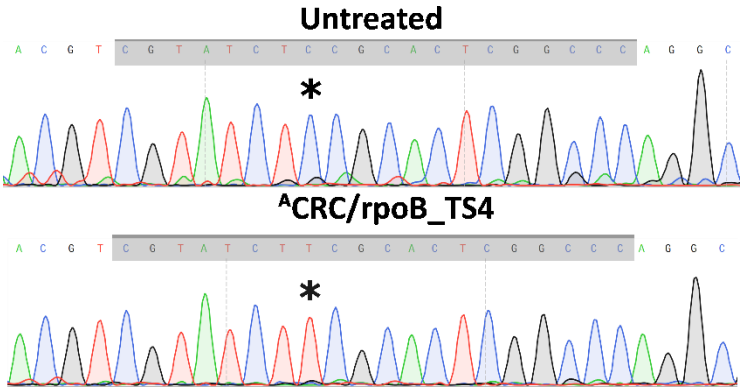


Figure 2. Site-specific nucleotide conversion in bacterial cells. *E. coli* MG1655 cells were treated with four gRNAs and selected in plates containing rifampicin. **A.** Schematic of representation of gRNAs targeting the RRDR Cluster I region of *E. coli*´s *rpoB* gene is shown, where PAM sequences in target DNA are shown in green and mutable positions in red. Black bars represent gRNA targeting sites. **B.** Panels show representative surviving colonies after treatment with CRC targeted with the indicated gRNAs. **C.** Survival fraction was calculated from similar experiments shown in **A. D.** Representative sequencing results from untreated cells (top) and ^{A}CRC treatment with *rpoB*_TS4 gRNA (bottom). Target position in CRC treated and untreated DNA are indicated with black arrows. This C1592>T mutation results in S531F change in protein sequence, a mutation known to induce rifampicin resistance.[68]

Table 1. rpoB targeting gRNA sequences

Name	Target strand	Target gene	Sequence	PAM
TS1	Template	rpoB	GCAGCAGTGAAAGAGTTCTT	CGG
TS2	Template	rpoB	CAGCCAGCTGTCTCAGTTTA	TGG
TS3	Template	rpoB	AAACGTCGTATCTCCGCACT	CGG
TS4	Template	rpoB	CGTATCTCCGCACTCGGCCC	AGG

System Optimization: Engineering Individual Modules Enhances Efficiency

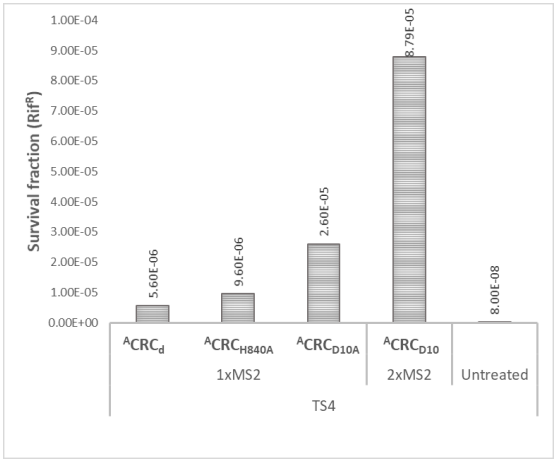
The positive results from our exploratory experiments prompted us to further engineer CRC system to increase its targeting efficiency using gRNA rpoB_TS4. Switching Cas9 module from dCas9 ($^{A}CRC_{d}$) to nickase $Cas9_{D10A}$ ($^{A}CRC_{D10A}$) or $Cas9_{H840A}$ ($^{A}CRC_{H840A}$) dramatically increased the number of surviving colonies (**Figure 3A**). Another strategy to enhance the system's efficiency is to increase the number of recruiting RNA scaffolds. Survival fraction was boosted to 1000-fold higher than untreated cells by doubling the number of MS2 sequences (**Figure 3A**). We decided to continue optimizing the system by testing different spatial configurations of the effector module, in combination with all Cas9 variants developed so far. To this end, various flexible linkers, with different lengths, were used to separate AID from MCP (**Table 2**). The 25 amino acid linker (L25), derived from the hinge region of immunoglobin gamma 3 (IgG3), showed the highest efficiency, although the variations between the different linkers were relatively small, especially for $^{A}CRC_{D10A}$ (**Figure 3B**). These results suggest that the separation between AID and MCP in the effector module does not play a critical role in overall targeting efficiency.

Table 2. Effector module linker sequences

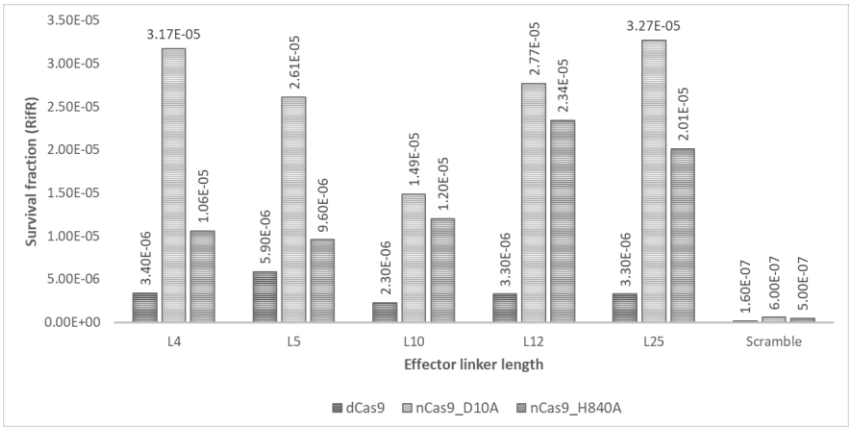
Linker name	Length (aa)	Sequence
L4	4	GSGS
L5	5	GSGRA
L10	10	GSGSGSGSGS
L12	12	GGGGSGGGGSGGGGS
L25	25	ELKTPLGDTTHTSPPCPAPELLGGP

Although $^{A}CRC_{H840A}$ modestly increased the survival fraction compared to $^{A}CRC_{d}$ (**Figure 3A, 3B**), sanger sequencing analysis of individual clones revealed that its targeting accuracy was lower compared to both $^{A}CRC_{d}$ and $^{A}CRC_{D10A}$, which targeted invariably the residue C1592 in codon 531 (**Figure 3C**). On the other hand, $^{A}CRC_{H840A}$ induced mutations not only on the target region, but several nucleotides upstream at high frequency (**Figure 3C**). For this reason, we decided to adopt $nCas_{D10A}$ as targeting module for further research and development in prokaryotic as well as in mammalian cells.

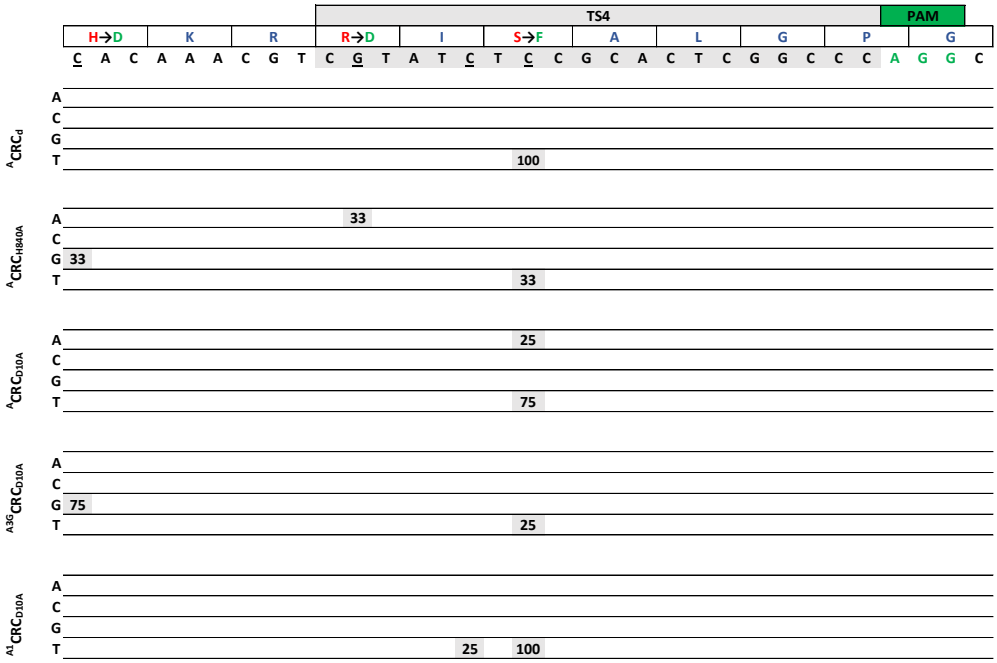
A.



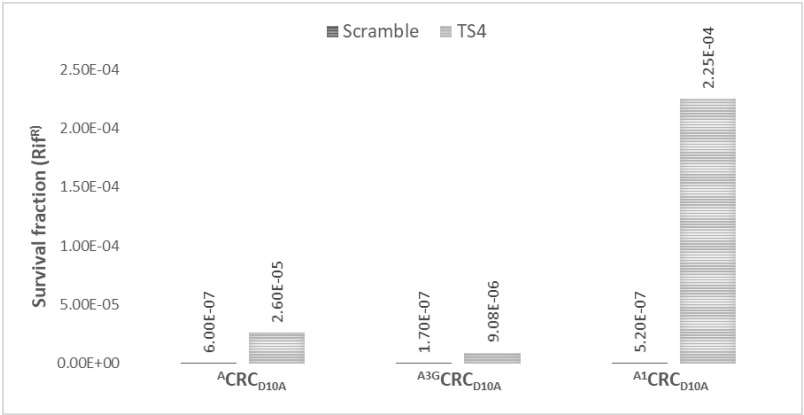
B.


Survival fraction (RifR)
3.50E-05
3.00E-05
2.50E-05
2.00E-05
1.50E-05
1.00E-05
5.00E-06
0.00E+00
3.40E-06
3.17E-05
1.06E-05
5.90E-06
2.61E-05
9.60E-06
2.30E-06
1.49E-05
1.20E-05
3.30E-06
2.77E-05
2.34E-05
3.30E-06
3.27E-05
2.01E-05
1.60E-07
6.00E-07
5.00E-07
L4
L5
L10
L12
L25
Scramble
Effector linker length
dCas9
nCas9_D10A
nCas9_H840A

C.


TS4
PAM
H→D K R R→D I S→F A L G P G
C A C A A A C G T C G T A T C T C C G C A C T C G G C C C A G G C
ACRd
ACRH840A
ACRD10A
A3GCRD10A
A1CRD10A
A C G T
100
33
33
25
75
75
25
25
100

D.


Scramble
TS4
Survival fraction (RifR)
2.50E-04
2.00E-04
1.50E-04
1.00E-04
5.00E-05
0.00E+00
6.00E-07
2.60E-05
1.70E-07
9.08E-06
5.20E-07
2.25E-04
ACRD10A
A3GCRD10A
A1CRD10A

E.

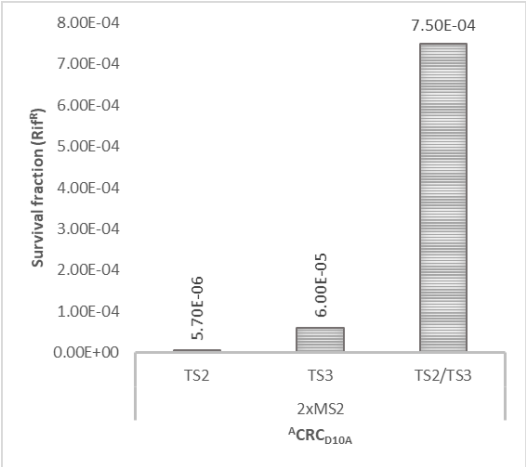


Figure 3. CRC system optimization. A. Efficiency can be increased by using Cas9 nickases ($nCas9_{H840A}$ or $nCas9_{D10A}$) instead of dCas9 as targeting module. Likewise, increasing the number of recruiting motifs from 1xMS2 to 2xMS2 enhances CRC's efficiency. **B.** Effect of linker length variation in effector module. Four linkers of various lengths and flexibilities were testes in these experiments. L4, L5, L10 and L25 are linkers consisting of 4, 5, 10 and 25 amino acids, respectively. **C.** Mutation distribution of clones selected on rifampicin after treatment. Five CRC systems are compared side by side: $^{A}CRC_{d}$, $^{A}CRC_{H840A}$, $^{A}CRC_{D10A}$, $^{A1}CRC_{D10A}$ and $^{A3G}CRC_{D10A}$. rpoB gene from individual clones was PCR amplified and sequenced for genotyping. Numbers represent percentage of clones with a given genotype; n=8 per treatment. **D.** APOEC3G ($^{A3G}CRC_{D10A}$) and APOBEC1 ($^{A1}CRC_{D10A}$) were tested as effectors side by side with the prototype system, $^{A}CRC_{D10A}$. **E.** Survival fraction of cis double nicking treatment with TS2 and TS3 gRNAs compared to single nicking. Targeting efficiency increased with cis double nicking 130- and 12-fold when compared to single targeting with TS2 and TS3 gRNAs.

There are many cytidine deaminases that can be incorporated into CRC system as effectors.[69] In this study, we tested two proteins related to AID from the APOBEC family of cytidine deaminases: APOBEC1 and APOBEC3G ($^{A1}CRC_{D10A}$ and $^{A3G}CRC_{D10A}$, respectively). $^{A1}CRC_{D10A}$ showed greater conversion efficiency, followed by $^{A}CRC_{D10A}$ and finally $^{A3G}CRC_{D10A}$ with the lowest activity (**Figure 3D**). Sequencing analysis revealed that $^{A1}CRC_{D10A}$ induced a high rate of double mutants (25%, **Figure 3C**). The data suggest that CRC system is amenable for further development, and other effectors, such as

adenosine deaminases, DNA glycosylases and other DNA editing proteins, can be easily incorporated into CRC toolbox.

One potential limitation of CRISPR targeting, especially for base editing, is the availability of the protospacer adjacent motif (PAM), a short DNA sequence immediately downstream of the target region that is required for proper Cas9 binding. This limitation is evidenced by the low survival fraction that resulted from treatments using TS2 and TS3 gRNAs, where the target Cs are outside the activity window, specifically on positions 12 and 14, respectively (**Figure 2A – B**). We reasoned that introducing two single strand nicks in cis (i.e. on the same DNA strand), could potentially alleviate this limitation by transiently exposing a longer stretch of single strand DNA, potentially making the target C more accessible to CRC effector module. Double nicking the same strand with two gRNA does not induce double strand DNA breaks or activation of DSB repair pathways.[31] To test this hypothesis, we multiplexed CRC targeting by delivering TS2 and TS3 simultaneously along with $^{A}CRC_{D10A}$ constructs and compare the results to single targeting using either of the gRNAs. The results show that cis double nicking dramatically increased CRC targeting efficiency in a synergistic fashion (**Figure 3E**).

Discussion

We have shown that base editing is feasible in prokaryotic cells using an RNA-aptamer based recruitment mechanism to recruit an effector protein, in this case AID, APOBEC1 and APBEC3G (**Figure 1, 2, 3**). These developments broaden the scope of current base editing systems, by allowing a more flexible configuration compared to direct fusions found in base editors. For example, to form active editosomes, APOBEC1 requires dimerization to become catalytically active.[64] Therefore, CRC's aptamer-based recruitment provides a modular platform where effector multimerization can potentially be achieved more easily, owed to the smaller size of the effector fusion protein.

Several system modifications were tested, highlighting the engineering potential that CRC system can offer. By switching the Cas9 module from dCas9 ($^{A}CRC_{d}$) to $nCas9_{D10A}$ ($^{A}CRC_{D10A}$), targeting efficiency goes from 70- to 325-fold higher than the untreated background (**Figure 3A**). In this configuration we are simultaneously introducing two well tolerated DNA lesions: nicksases introduce single strand breaks (SSB) on the DNA, while AID generates a mismatch. SSBs are among the most common DNA lesions,[35] however this type of DNA damage can become clastogenic if left unrepaired.[36,70] In this way, by introducing these mild insults to target DNA, cells can be biased to repair the most threatening lesion, the DNA nick, thus fixing the single nucleotide editing introduced by the effector module.

Additional effective engineering interventions included increasing the number of recruiting RNA motifs, by which the efficiency went over 1000-fold greater than untreated cells (**Figure 3A**). We also tested variations of the spatial configuration of the effector module. To this end, we engineered the effector module to include linkers of various lengths and sequences shown in **Table 2**. Again, the highest efficiency was observed after $^{A}CRC_{D10A}$ across different treatments (**Figure 3B**). Although the difference was small, the greatest efficiency was achieved with L25 linker, a flexible peptide derived from the hinge region of Immunoglobin gamma 3 protein (**Figure 3B**), and this linker was selected for further development. Finally, we tested different effector proteins in addition to AID ($^{A}CRC_{D10A}$), specifically APOBEC1 ($^{A1}CRC_{D10A}$) and APOBEC3G ($^{A3G}CRC_{D10A}$) (**Figure 3D**). $^{A1}CRC_{D10A}$ showed the highest activity in terms of survival fraction (**Figure 3D**). However, mutational analysis revealed that APOBEC1 is somewhat promiscuous, introducing two simultaneous mutations at high frequency (**Figure 3C**). This could be potentially deleterious, but because bacterial cells have different mechanisms to cope with DNA insults, $^{A1}CRC_{D10A}$ will be revisited in mammalian system.

The nature of RNA-aptamer recruitment also allows for orthogonal engineering. In gene expression regulation[41] and genomic labeling[71] experiments, it is possible to simultaneously recruit two orthogonal RNA binding proteins by tethering heterologous RNA scaffolds to the gRNA. In this way, more complex interventions could be designed

if heteromeric effectors need to be assembled at the target site, or if additional proteins with desirable features (e.g. DNA protection) are incorporated in the system. Taken together, the data show that RNA-aptamer mediated base editing is a feasible approach, and that CRC's modular configuration allows for flexible engineering.

Materials and Methods

Strains: *E. coli* DH5α competent cells were purchased from Thermo Fisher (Cat. No. 18265017) and were used for general cloning purposes. *E. coli* MG1655 strain, used for rpoB gene targeting, was a kind gift from Dr. Stanley Qi (Stanford University). MG1655 cells were made competent using a standard CaCl2 protocol.

Plasmids: pgRNA-bacteria (pUC19, ampicillin resistant; Addgene plasmid # 44251) was engineered to include two offset BbsI restriction sites for guiding sequence cloning, as well as 1 or 2 MS2 stem loop sequences at the 3' end. These modifications were introduced using standard gene synthesis services (Genewiz). The synthesized cassettes were cloned into pUC19 backbone using SpeI and HindIII restriction sites. The effector modules (AID-linker-MCP) were cloned into a pCDFDuet empty vector (DF13, streptomycin resistant; Addgene plasmid # 49796) using BglII and BamHI restriction sites. dCas9-bacteria plasmid (p15A, chloramphenicol resistant; Addgene # 44249), and pwtCas9-bacteria (p15A; Addgene # 44250) were used to generate $nCas9_{D10A}$ and $nCas9_{H840A}$ nickases by swapping portions of the wild type HNH and RuvC active sites,

respectively, from pwtCas9 to dCas9. HNH domain was cloned using Acc65I and BamHI restriction sites. RuvC domain was cloned using XbaI and NheI restriction sites. Cas9 and effector constructs are under the control of a tetracycline inducible promoter. Plasmid sequences are detailed in **Appendix** (Bacterial Expression Plasmid Sequences section).

gRNA design: rpoB targeting gRNAs were designed manually on SnapGene Viewer (GSL Biotech), on or near the rifampicin resistance determining region (RRDR) of *E. coli*'s rpoB gene.[67] gRNA sequences and PAMs are summarized in **Table 1**. Guiding sequences were designed to have 5' overhangs compatible to the overhangs left by BbsI digestion (i.e. Fwd 5'-CTAGN_{20}-3', Rev 5'-AAACN_{20}-3', where N_{20} is the programmable guiding sequence and must be complementary between Fwd and Rev oligos).

Treatments: Chemically competent *E. coli* MG1655 cells were transformed with 9 ng of a 1:1:1 combination of the appropriate plasmids encoding for specific gRNA (ampicillin), AID_MCP (streptomycin) and Cas9 (chloramphenicol) constructs. After transformation, cells were selected overnight in liquid LB media containing working concentrations of ampicillin, streptomycin and chloramphenicol. The day after, cells are diluted in selective media supplemented with 3 µM tetracycline to induce expression of the protein coding modules. After overnight growth, OD is measured, and serial dilutions are performed to plate 10^8-10^3 cells in rifampicin containing LB agar. Plates are incubated at 37° C and monitored for 48h. Surviving fraction is calculated by counting the surviving colonies divided by the number of cells plated.

Mutational analysis: DNA from 8 to 12 colonies from appropriate experiments was extracted. The target region of the rpoB gene (i.e. RRDR region) was PCR amplified using the following primers: Fwd 5'-TTGGCGAAATGGCGGAAAACC-3' and Rev 5'-CACCGACGGATACCACCTGCTG-3'. PCR products were sequenced using Sanger chemistry at Genewiz.

CHAPTER III

Construction of CRC System for Mammalian Expression

With the advent of gene editing systems, for the first time in history humans are able to rationally modify genomes at will. These developments have become instrumental research tools for biotechnology and biological research in general, and have generated hope for the development of targeted therapies for genetic diseases, cancer, HIV, etc. However, gene editing technologies have important limitations that need to be addressed before moving these technologies to the clinic. First, canonical gene editing systems rely on the generation of double strand breaks, which could potentially have deleterious consequences, especially if unintended off-target activity is high.[72,73] Although the development of strategies such as paired nickases[31], catalytically inactive Cas9 fused to dimeric nucleases[27,33] or high-fidelity CRISPR systems[25,26] are thought to mitigate these adverse effects, limitations in on- and off-target mutagenesis detection methods may underestimate the actual effects of gene editing interventions. Importantly, it was recently shown that CRISPR systems induce previously unnoticed deletions and rearrangements that span several kilobases at on-target sites.[74] Likewise, insertional mutagenesis has been observed in experiments using purified Cas9/sgRNA ribonucleoprotein complexes (RNP), a method thought to enhance targeting specificity.[75] Second, in order to introduce precise modifications, it is necessary that the target cells express homology dependent DNA double strand break repair (HDR).[6,10]

Somatic, terminally differentiated cells, however, do not express HDR and instead express the error prone non-homologous end joining (NHEJ) pathway.[76] These findings highlight the limitations of current gene editing systems to develop therapeutics and the need for novel, more accurate ways to modify genomes avoiding potentially pathogenic effects and introducing the right modifications to the right cells.

As mentioned in Chapter II, we and others have developed innovative ways to edit target nucleotides in the genome, bypassing the introduction of DSB by combining the DNA targeting capabilities of CRISPR and the catalytical activity of non-nuclease DNA editing enzymes (e.g. cytidine deaminases). This strategy has been named base editing.[48] One of the most appealing applications for base editing technologies is the development of novel therapeutic tools to tackle the root cause of human genetic diseases. With the information collected from our prokaryotic experiments we sought to construct a mammalian version of our CRC system. In this Chapter, I describe the construction of the first-generation CRC system for mammalian expression, and its application in correcting a loss-of-function mutation in a reporter system. This represents an exciting proof of concept of a therapeutic application, where the function of an inactive gene product is restored with high efficiency by correcting the deficiency-causing mutation, with undetectable off-target effects at exome-wide level.

Results

CRC system in mammalian cells

We recapitulated the $^{A}CRC_{D10A}$ system (for simplicity hereafter called ^{A}CRCn) for mammalian expression by generating a multicistronic vector expressing $nCas9_{D10A}$ and AID_MCP fusion separated by a self-cleavable 2A peptide, under the control of a UbC promoter (**Figure 4A**, top). In order to increase the targeting space, we included a Cas9_VQR variant ($nCas9_{VQR/D10A}$) that recognizes NGA instead of NGG as PAM sequence used by *S. pyogenes* CRISPR/Cas9 system[77] (**Figure 4A**, top). This construct is noted as ^{A}CRCvn. The gRNA vector used in this Chapter have similar architecture than the ones used in the prokaryotic experiments, with 2 MS2 loops at the 3' end of the gRNA scaffold (2xMS2; **Figure 4A**, bottom). To test CRC's base editing activity in mammalian cells, we designed a GFP reporter that harbors an A→G point mutation along the chromophore sequence (**Figure 4B**). This mutation, Y66C, renders the protein non-fluorescent (nfEGFP), thus mimicking a loss of function (LOF) mutation. We designed three gRNAs targeting the non-template strand (NT) around the mutation region (nfEGFP_NT1, NT2 and $NT3^{VQR}$; **Figure 4B**, **Table 3**).

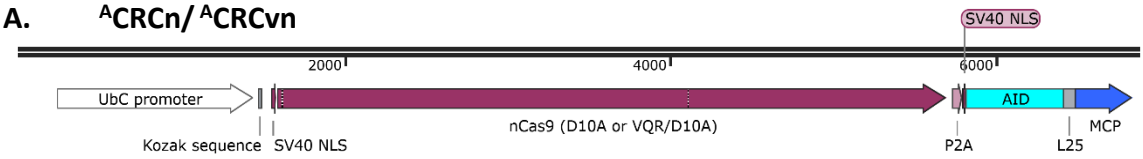
Table 3. nfEGFP targeting gRNA sequences

Name	Target strand	Sequence	PAM
NT1	Non-template	CGCAGGTCAGGGTGGTCACG	AGG
NT2	Non-template	CTGCACGCCGCAGGTCAGGG	TGG
$NT3^{VQR}$	Non-template	GCCGCAGGTCAGGGTGGTCA	CGA

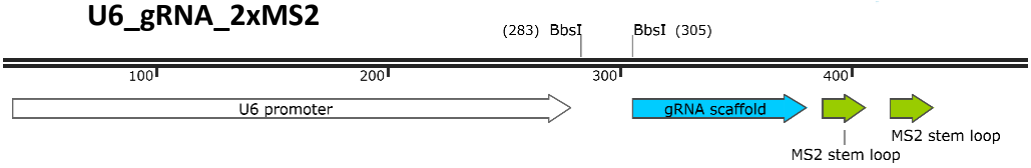
Note: Target C is underlined

First, we sought to correct the LOF mutation in extrachromosomal DNA. To this end, we transiently expressed the target construct in HEK 293 cells together with ^{A}CRCn or ^{A}CRCvn components and the indicated gRNAs (**Figure 4C**). nfEGFP_NT1 treated cells yielded higher GFP conversion than nfEGFP_NT2 treated cells (**Figure 4C**). This result was expected because of the relative position of the target C, which in NT2 lays outside the activity window (position 11) whereas in NT1 the target is more accessible to CRC's effector module (position 3) (**Figure 4B**, **C**). These observations confirmed the notion that the relative position of the target nucleotide within the protospacer sequence is important for efficient conversion, similar to what we observed in our prokaryotic experiments (**Figure 2B**, compare TS3 to TS4). Although ^{A}CRCvn/nfEGFP_NT3^{VQR} treatment showed relatively lower activity than ^{A}CRCn/nfEGFP_NT1 (**Figure 4C**), it demonstrates that CRC base editing activity could be expanded to NGA PAM sequences, thus increasing its targeting space. Finally, for comparison, we tested base editor third generation (BE3)[48] side-by-side with ^{A}CRCn targeted with nfEGFP_NT1. Higher GFP conversion was observed in CRC than in BE3 treated cells. Quantitation by flow cytometry showed 35% GFP positive (GFP+) after ^{A}CRCn/nfEGFP_NT1 treatment, whereas BE3/nfEGFP_NT1 treatment resulted in 19% GFP+ cells (**Figure 4D**). These results confirm that base editing is feasible when treating extrachromosomal DNA in mammalian cells, and that CRC is an efficient alternative to base editors.

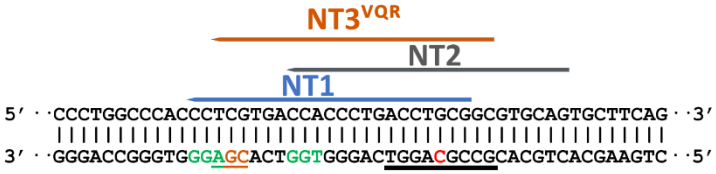
A. ACRCn/ ACRCvn



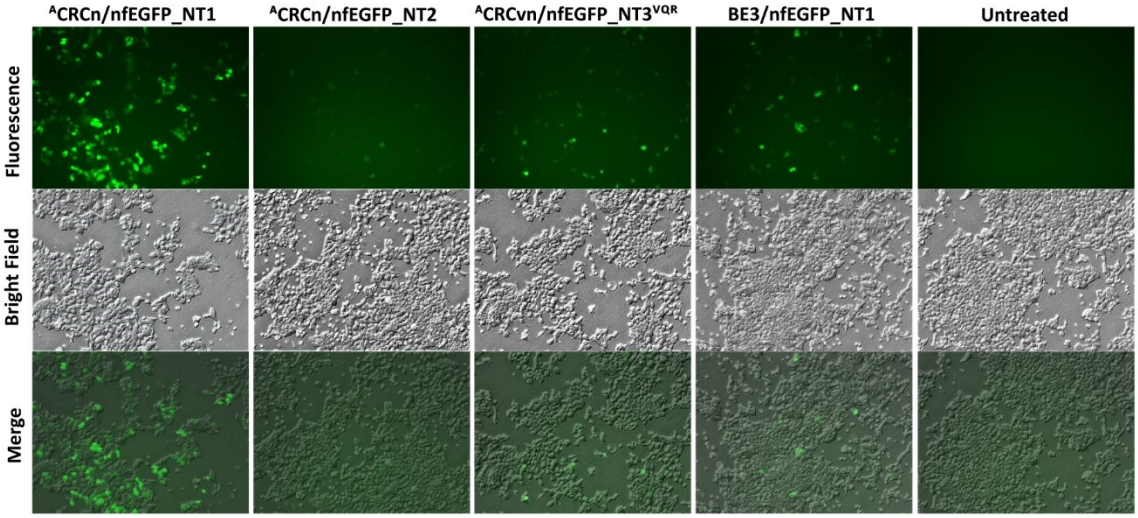
U6_gRNA_2xMS2



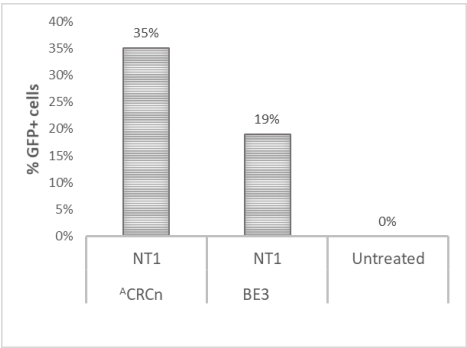
B.



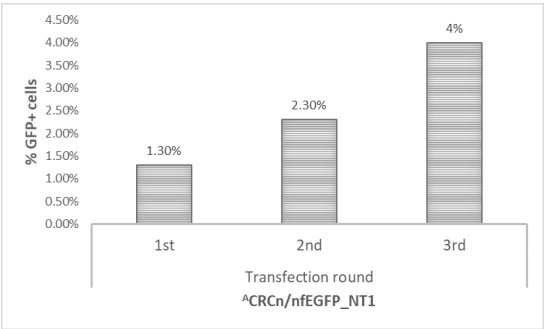
C.



D.



E.



F.

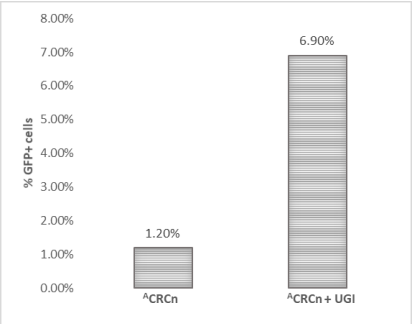


Figure 4. CRC expression and activity in mammalian cells

A. Schematic representation of constructs used in this Chapter. Top: ^{A}CRCn and ^{A}CRCvn are multicistronic constructs expressing AID_L25_MCP fusion protein and nCas9$_{D10A}$ or nCas9$_{VQR/D10A}$, respectively. The two modules are separated by a self-cleavable 2A, and their expression is driven by a UbC promoter. Bottom: gRNA_2xMS2 constructs express from a U6 promoter. Specific guiding sequences are cloned by BbsI digestion **B.** Schematic representation of non-fluorescent EGFP (nfEGFP) target region. A→G LOF mutation occur on the chromophore sequence (underlined in black). Three gRNAs targeting the non-template strand (NT) were designed: NT1, NT2 and NT3^{VQR}. PAM sequences for NT1 and NT2 (NGG) are shown in green, while PAM sequence for NT3^{VQR} (NGA) is orange and green and is underlined with the same colors. **C.** Extrachromosomal treatment. HEK 293 cells were transiently transfected with target DNA containing the nfEGFP mutant together with ^{A}CRCn, ^{A}CRCvu or BE3 components and the indicated gRNAs. **D.** Flow cytometry analysis of cells expressing extrachromosomal nfEGFP gene treated with ^{A}CRCnu and BE3 targeted with nfEGFP_NT1. **E.** nf2.16 cells were subjected to three consecutive treatments with ^{A}CRCn/nfEGFP_NT1, and GFP conversion was measured by flow cytometry. Percentage of GFP positive cells went from 1.3% after a single treatment, to 4% after the third consecutive transfection. **F.** Addition of UGI module to CRC construct enhances GFP conversion efficiency.

CRC Corrects Loss of Function Mutation in Genomic DNA

To examine whether the system can correct loss of function mutations in mammalian genome, we stably integrated the mutant nfEGFP gene into HEK 293 genome (nf2.16 cells). Cells treated with ^{A}CRCn showed 1.3% correction efficiency after a single

transfection (**Figure 4E**). To further enhance base editing efficiency, we hypothesized that chronic treatment would build up the effect. This would mimic chronic treatments in patient's tissues or cells. To this end, nf2.16 cells were subjected to three rounds of treatment with ^{A}CRCn/nfEGFP_NT1. Consecutive treatments resulted in increased GFP conversion, from 1.3% with a single transfection to 4% after three rounds (**Figure 4E**). This approach suggests that if needed, cells can undergo multiple treatment rounds until the desired efficiency is reached.

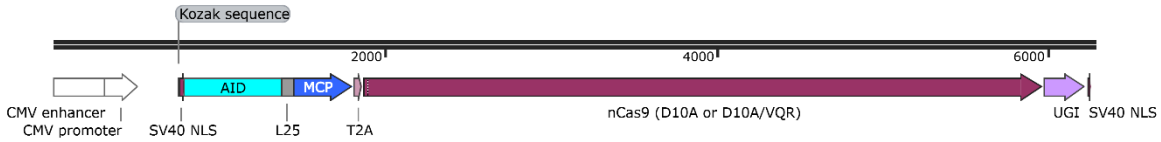
Ideally, however, higher efficiency with a single treatment would be preferable, to avoid accumulation of off-target effects. The observed conversion efficiency could be a balance between target nucleotide deamination and its repair by endogenous DNA repair pathways. In cells, uracil DNA glycosylase (UNG) initiates the repair of U:G mismatches induced by cytidine deamination.[58,78,79] To enhance nucleotide conversion efficiency at the target sites, we fused a bacterial UNG inhibitor peptide (UGI)[80] to Cas9, thus eliciting local UNG inhibition. After a single treatment with this construct guided by nfEGFP_NT1, GFP conversion efficiency reached over 6% (**Figure 4F**). The data show that CRC system is suitable for further engineering and optimization.

We decided to continue the optimization process by restructuring construct architecture and expression. It has been shown that constitutive expression of UbC promoter is inferior to CMV.[81] Therefore, we incorporated the latter into our CRC

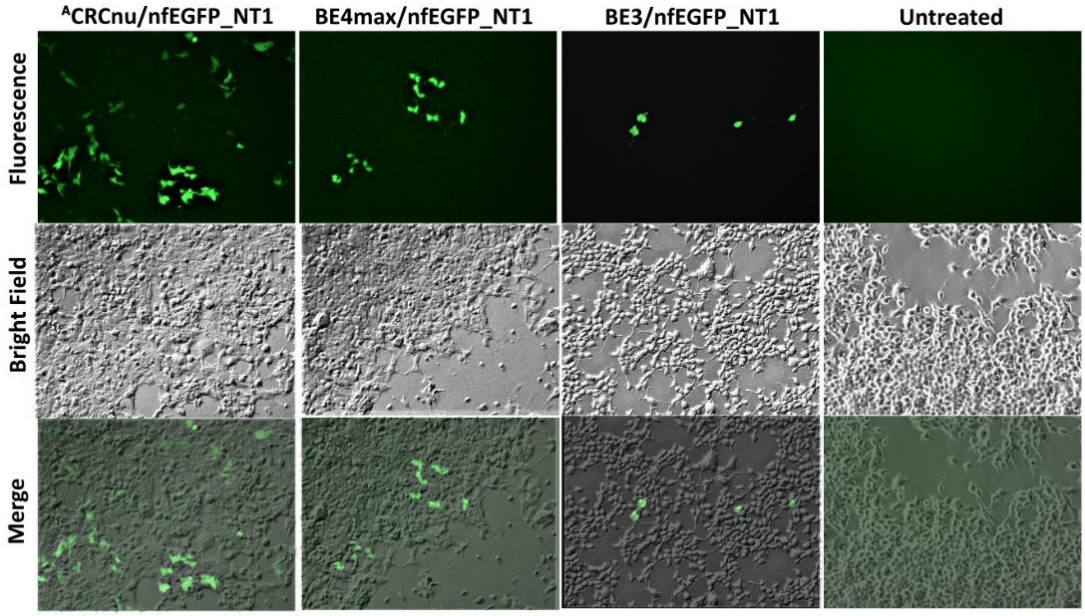
vectors. Next, because Cas9 is a bacterial gene and AID is a human gene, we restructured the multicistronic constructs to enhance expression, cloning AID_MCP fusion upstream of Cas9 (**Figure 5A**), thus potentially improving transcription and translation in HEK 293 cells. The nomenclature for this generation of CRC constructs is: ^{A}CRCnu or ^{A}CRCvnu (A=AID; n=nCas9$_{D10A}$; U=UGI; V=VQR).

To test the latest generation of constructs, cells were treated with ^{A}CRCnu, BE3 and the recently published optimized base editor, BE4max.[62] For this set of experiments, all the base editing systems were targeted with nfEGFP_NT1. Fluorescence was restored at high efficiency after ^{A}CRCnu treatment, with nearly 10% of GFP+ cells after a single transfection (**Figure 5B, C**). As in the extrachromosomal targeting, BE3 treatment resulted in lower GFP conversion, yielding 0.3% of GFP positive cells. BE4max also resulted in modest GFP correction, with 2.7% GFP positive cells (**Figure 5B**, **C**). These results indicate that CRC configuration has greater efficiency than the latest generation of base editors for this target. To visualize the effect of ^{A}CRCnu on genomic nfEGFP gene, we sorted the GFP positive cells by fluorescent activated cell sorting analysis (FACS) followed by sanger sequencing. The results confirmed the G→A conversion at the target position, restoring the wild type sequence (**Figure 5D**). Together, the data demonstrate that CRC mediated base editing is feasible and efficient in mammalian genome, confirming our observations in prokaryotic cells.

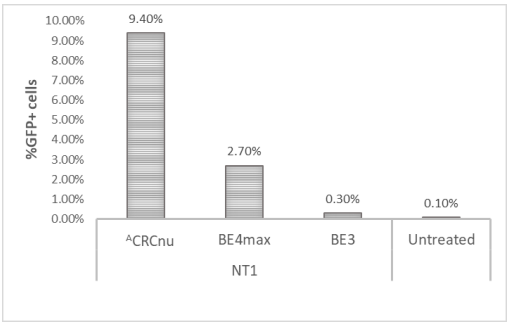
A. **ACRCnu / ACRCvnu**



B.



C.



D.

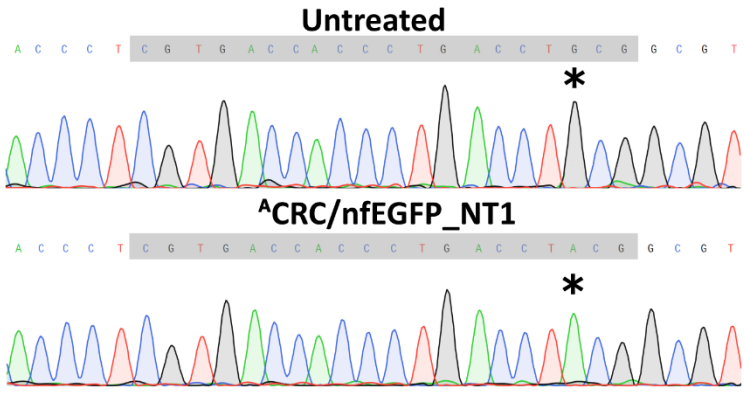


Figure 5. Treating a loss of function mutation in mammalian genome

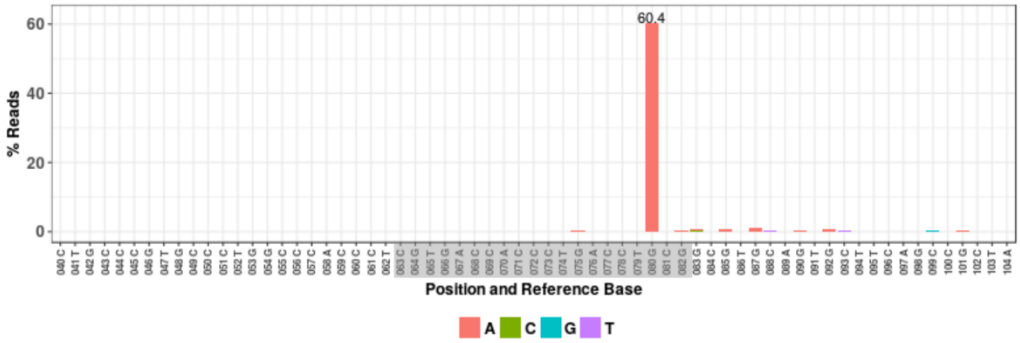
A. Schematic representation of ^{A}CRCnu and ^{A}CRCvnu. **B.** HEK 293 cell stably expressing the non-fluorescent EGFP mutant gene (nf2.16 cells) were treated with ^{A}CRCnu, BE4max and BE3 guided by nfEGFP_NT1 gRNA. **C.** Flow cytometry analysis of cells from **B**. **D.** nf2.16 cell treated with ^{A}CRCnu/nfEGFP_NT1 were subjected to FACS. Sequencing of sorted cells showed specific G→A conversion of the target nucleotide. Highlighted in gray is the targeted region. Black asterisks show the edited position.

Next Generation Sequencing Analysis Reveals High On-Target Base Editing Efficiency and Undetectable Off-Target Effects

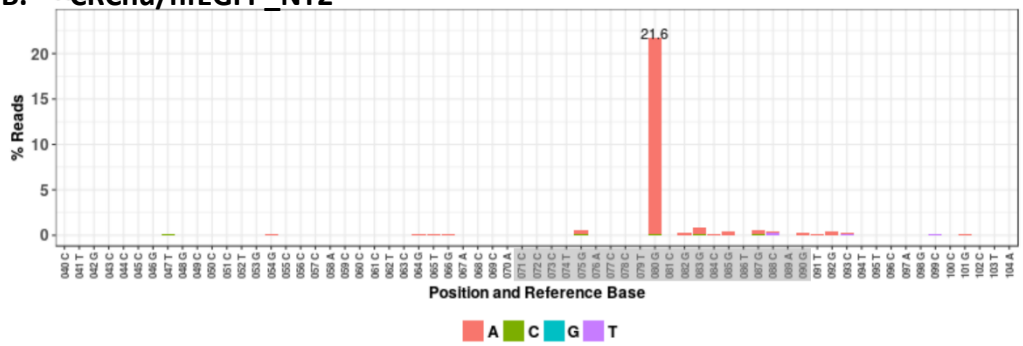
To examine the on–target efficiency of ^{A}CRCnu guided by nfEGFP_NT1, NT2 and NT3^{VQR}, target DNA from treated and untreated cells was amplified by PCR and subjected to high-throughput sequencing analysis. Strikingly, ^{A}CRCnu/nfEGFP_NT1 treatment resulted in over 60% of G→A editing, restoring EGFP wild type sequence, with low bystander effect (**Figure 6A**). Although lower, ^{A}CRCnu/nfEGFP_NT2 and ^{A}CRCvnu/nfEGFP_NT3^{VQR} treatment also resulted in efficient G→A editing, with 21% and 25% conversion efficiency, respectively (**Figure 6B**, **C**). Untreated cells showed background mutations at the target region (**Figure 6G**). In the untreated control,

however, there is a peak at the target nucleotide of approximately 1.4%. This may be due to spontaneous mutation or genetic heterogeneity and could potentially explain the detection of 0.1% of GFP positive cells by flow cytometry (**Figure 5C**, Untreated) that was not observed in the extrachromosomal experiment (**Figure 4D**, Untreated). Contamination was ruled out since no mutation was found in the plasmid encoding nfEGFP, which was amplified and sequenced in parallel with the untreated control (**Figure 6E**). These results confirm that CRC system is efficient, highly specific and suitable for treating loss of function mutations in mammalian genome.

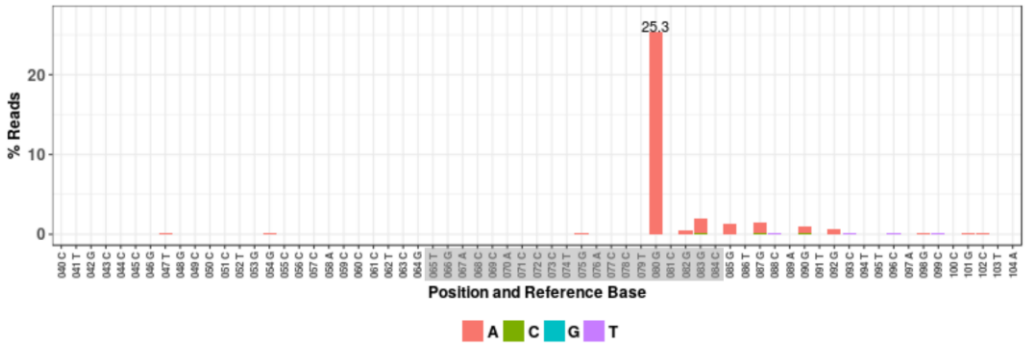
A. ^{A}CRCnu/nfEGFP_NT1



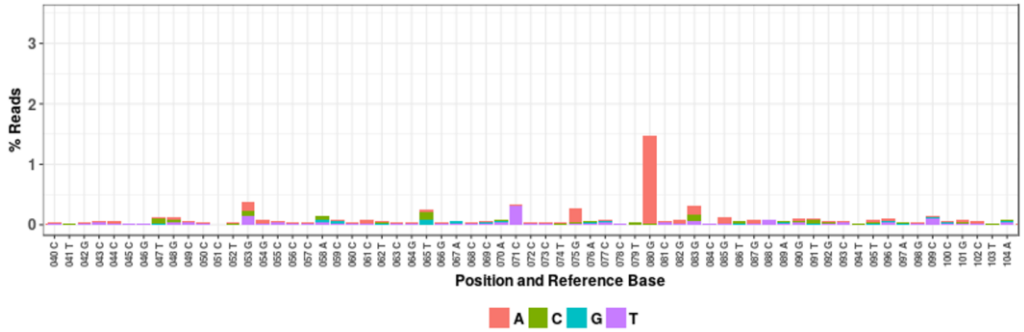
B. ^{A}CRCnu/nfEGFP_NT2



C. ^{A}CRCvnu/nfEGFP_NT3^{VQR}



D. Untreated (genomic)



E. Untreated (plasmid)

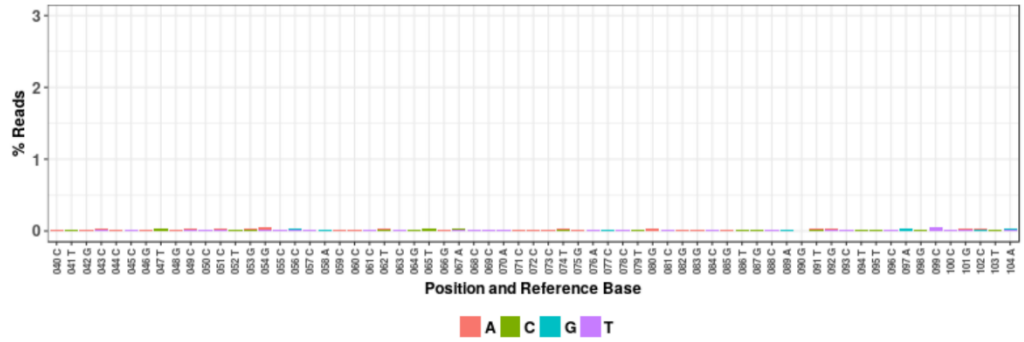
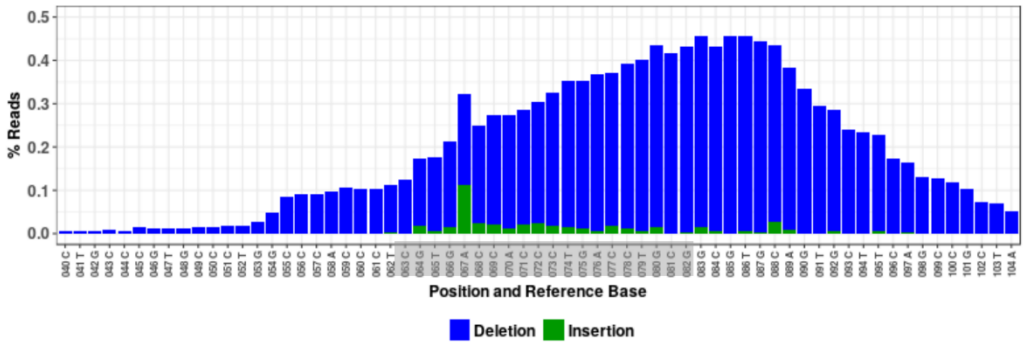


Figure 6. High throughput sequencing analysis of on-target effects. A – D. High-throughput sequencing analysis of target region of stably integrated nfEGFP gene. nf2.16 cells were treated with ^{A}CRCnu/nfEGFP_NT1 (**A**), ^{A}CRCnu/nfEGFP_NT2 (**B**),

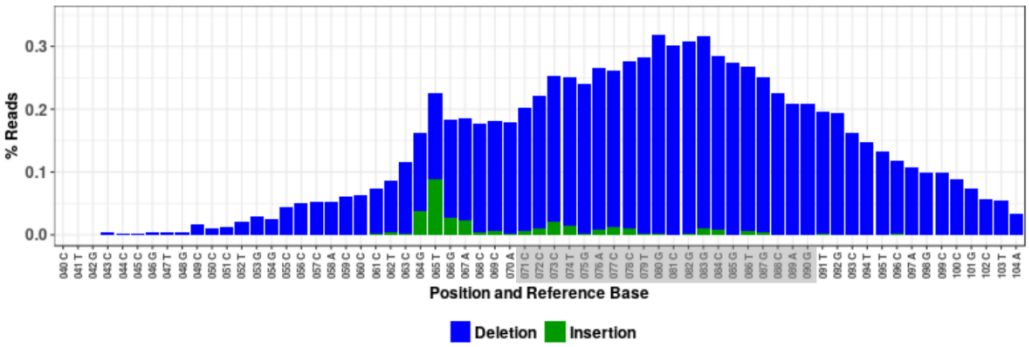
^{A}CRCvnu/nfEGFP_NT3^{VQR} (**C**), or untreated (**D**). In **A – D**, gRNA target sites are highlighted in gray. **E.** Plasmid DNA encoding nfEGFP was amplified and sequenced in parallel to rule out any potential contamination of untreated control (**D**). Numbers below sequences represent the base position within the sequenced amplicon.

Cas9 nickases are largely considered safe because single strand breaks in DNA are well tolerated and efficiently repaired in cells.[35-37] However, it has been shown that nickases can induce mutagenesis[33] and may generate insertion/deletion of bases (indels), though at low rates.[40] Moreover, researchers have found that base editors induce indels at the target site,[48,62] suggesting that these interventions may generate DSBs in a fraction of chromosomes. To determine the extent of indel formation after CRC treatment, we analyzed our data to estimate the frequency of these events in treated and untreated cells. Expectedly, CRC targeting induced indels at low frequencies with all the gRNAs tested, whereas untreated cells showed only background levels of indels (**Figure 7A – D**). Note that the distribution and frequency of indels in treated cells correlates with gRNA target sites. To our knowledge, this is the first report of such effect by Cas9 nickases on target DNA.

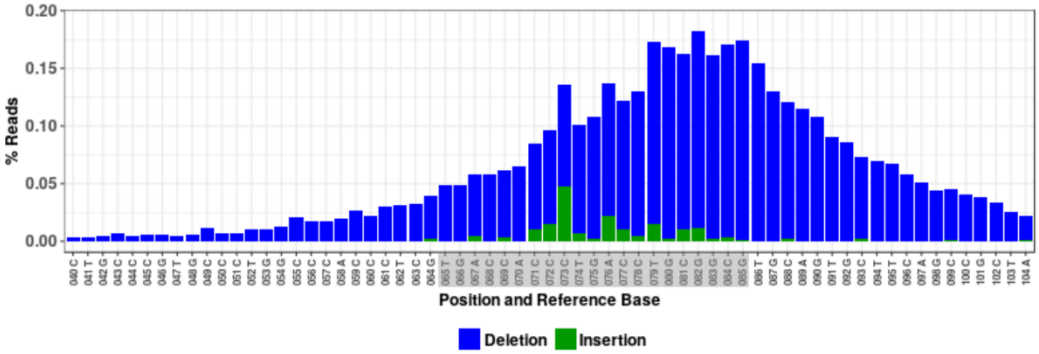
A. ^{A}CRCnu/nfEGFP_NT1



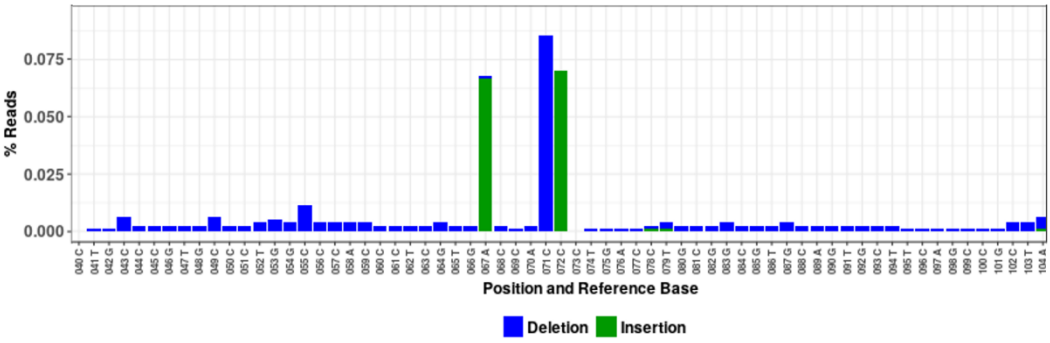
B. ^{A}CRCnu/nfEGFP_NT2



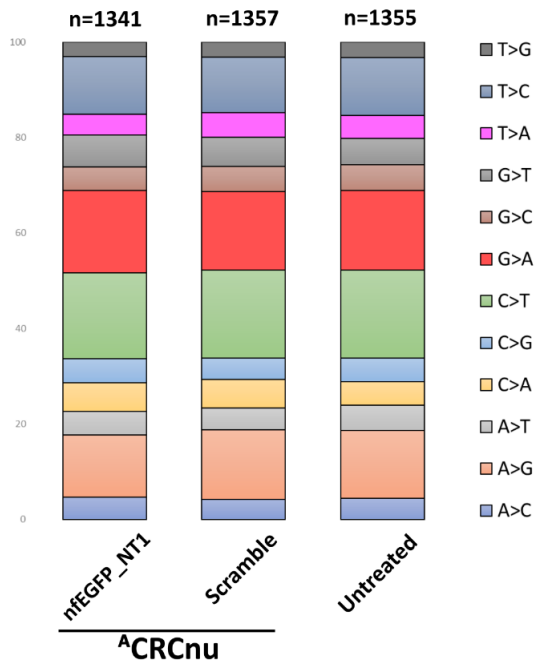
C. ^{A}CRCvnu/nfEGFP_NT3^{VQR}



D. Untreated



E.



F.

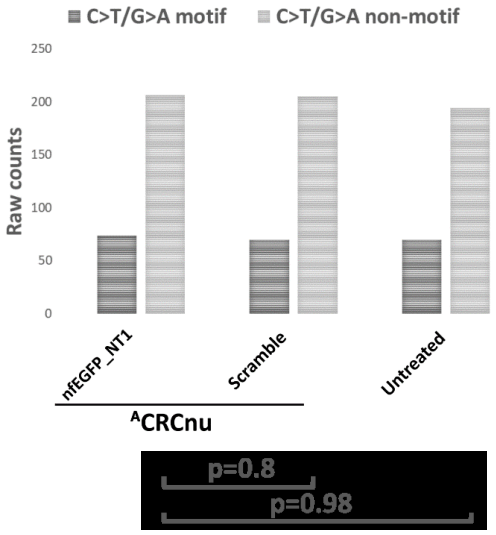


Figure 7. Analysis of on-target mutagenesis and exome-wide off-target effects
A – D. Analysis of indel formation after CRC treatment guided by nfEGFP_NT1 (**A**), nfEGFP_NT2 (**B**), nfEGFP_NT3^{VQR} (**C**) or untreated (**D**). Numbers below sequences represent the base position within the sequenced amplicon. **E.** nf2.16 cells were treated with ^{A}CRCnu/nfEGFP_NT-1, ^{A}CRCnu/Scramble or untreated. Genomic DNA was isolated and subjected to whole exome sequencing. The figure shows the global distribution of single nucleotide polymorphisms of the three samples compared to the human reference genome (hg38), including AID signature mutations C→T/ G→A (green and red,

respectively). The data show no effect of CRC system at exome-wide level. **F.** Using hg38 exome as a reference, occurrences of C>T and G>A events in "AID motif" (WRCH/DGYW, dark gray bars) versus "non-motif" (NNCN/NGNN; light gray bars) were tested. Mutations on CpG sites were not counted to avoid overestimation due to higher mutation rates in these sites. p values were calculated using Chi-square test

To assess potential CRC mediated off-target activity at exome-wide level, nf2.16 cells treated with ^{A}CRCnu/nfEGFP_NT1, ^{A}CRCnu/scramble or untreated cells were subjected to whole exome sequencing. Analysis of point mutations showed no increase in global single nucleotide mutations, including AID signature mutations C→T/ G→A, compared to untreated control, suggesting that AID in the system does not have significant off-target activity (**Figure 7E**). Because AID mutates cytosine residues preferentially within WRCH/DGYW motifs (where the underlined C and G are mutable positions; W: A or T; R: A or G; H: A, C or T; D: A, G or T; Y: C or T),[82] we looked at sites where these residues were mutated with respect to reference human genome, and compared to C or G mutations in non-motifs (NNCN/NGNN) across the different treatments. We found no difference between CRC treated and untreated samples, on both motifs and non-motifs, suggesting that the mutations observed are background mutations and (**Figure 7F).** Taken together, the data show that our GE system represents a promising and safe technology for correcting genetic mutations independent DSB and HDR, with potential therapeutic value in non-dividing somatic cells.

Discussion

Here we have presented a compelling proof of principle showing that CRC is not only active in extrachromosomal and chromosomal DNA in mammalian cells, but more efficient than the third[48] and fourth[62] generation of optimized base editors (**Figure 4 C, D** and **Figure 5 B, C**). The enhanced efficiency of CRC in this specific target compared to BE3 and BE4max could be due to at least three reasons, or a combination thereof. First, CRC uses an RNA-aptamer-based recruitment mechanism, by which the effector fused to a small RNA-aptamer-binding protein is recruited to the target site. Cytidine deaminases are known to form dimers or higher order oligomers to become catalytically active.[64,83-87] The formation of such structures may be hindered when the effector is fused to a bulky protein, such as Cas9 (i.e. in base editors). Therefore, dimer formation could be potentially facilitated in CRC configuration. Second, different cytidine deaminases have certain nucleotide motifs where they deaminate cytosines preferentially. For example, APOBEC1 (effector protein in BE3 and BE4max) deaminates cytosines in a TC context,[48,88] while AID has a more relaxed sequence motif, preferentially deaminating WR<u>C</u>H/D<u>G</u>YW sequences.[82] Third, because of the different effectors and recruitment mechanisms between CRC and BE, the systems may have diverse targeting requirements in terms of the activity window (i.e. the accessible portion of the unpaired strand where effectors can deaminate their target cytosines). Consequently, CRC system increases the base editing tool repertoire, potentially expanding its targets and applications.

In **Chapter II**, we show that CRC's modular design is amenable to optimization by simple changes in vector configuration and treatment modification (**Figure 3A, 3D, 3E**). Here, we decided to take a similar approach to optimize the system for mammalian expression. Functional correction efficiency was dramatically enhanced by using the following strategies: (i) sequential treatment mimicking a multidose therapy (**Figure 4E**); (ii) incorporating a potent UNG inhibitor (UGI) on the Cas9 module (**Figure 4F**); (iii) switching the UbC promoter for CMV and (iv) by restructuring the architecture of the multicistronic construct (**Figure 5B**, **5C**). These simple steps demonstrate the versatility and adaptability of CRC design, and confirm that the system is suitable for further engineering.

We report editing and functional correction of a loss of function mutation in an ectopic EGFP gene, in both extrachromosomal and chromosomal DNA (**Figure 4, 5, 6**). Functional correction of chromosomal GFP reached over 9% efficiency after a single treatment with our improved ^{A}CRCnu guided by nfEGFP_NT1 (**Figure 5B**). Interestingly, we observed that at genomic level, target nucleotide editing was detected in over 60% of reads (**Figure 6A**). These results indicate that CRC is a powerful base editing system with high efficiency in correcting a LOF mutation, highlighting its therapeutic potential.

High base editing efficiency was also observed ^{A}CRCnu/nfEGFP_NT2 (21.6%, **Figure 6B**), although the relative position of the target base is farther from the 5' end compared to

NT1 (**Figure 4B**). We also show that incorporating Cas9 variants with altered PAM recognition requirements[77] (i.e. Cas9_VQR, ^{A}CRCvnu) is feasible and efficient (25%, **Figure 6C**). This is important, because for base editing PAM availability is a major limiting factor since the target base requires to be on the unpaired strand of the CRISPR R-loop, and its position relative to the PAM also determine the editing efficiency. Therefore, adding Cas9 variants that recognize various PAMs could alleviate such limitation, increasing the scope of CRC editing.

Finally, a major concern in gene editing are the unintended effects at on- and off-target sites. It was recently reported that nuclease-dependent CRISPR systems introduce previously unidentified mutagenicity around on-target sites, inducing deletions spanning several kilobases and also triggering complex genomic rearrangements.[74] Moreover, even approaches generally thought to improve editing precision, such as the use of RNPs, have been shown to introduce insertional mutagenesis at on-target sites.[75] Since base editing does not employ nucleases, and therefore it does not directly introduce DSB, this effects should be mitigated. However, indel formation has been observed in base editing experiments,[48] suggesting that these systems introduce unintended DNA breaks. With this in mind, we decided to take a look at indel formation after CRC treatments. As expected, indels were detected in CRC treated cells, although at very low frequencies (**Figure 7A – D**). Additionally, potential off-target effects were analyzed using high-throughput whole exome sequencing. No significant differences were detected in SNPs between treated and untreated cells (**Figure 7E**, **F**), indicating that CRC

has no detectable off-target activity. Taken together, the data show that CRC system can correct a LOF mutation in mammalian genome with high on-target efficiency, and low on- and off-target mutagenic effects, highlighting its therapeutic potential.

Materials and Methods

Plasmids: ^{A}CRCn plasmid was engineered from pLV hUbC-dCas9-T2A-GFP (Addgene plasmid # 53191).[89] P2A-AID_MCP was cloned downstream of Cas9. EGFP was deleted from this construct before performing nfEGFP base editing experiments. nCas9$_{D10A}$ was generated by site directed mutagenesis using 5'-gactacgacgtggatcatatcgtgccccagtc-3' and 5'-gactggggcacgatatgatccacgtcgtagtc-3' as mutagenic primers. To generate ^{A}CRCnu and ^{A}CRCvnu multicistronic constructs, AID_MCP fusion was amplified from ^{A}CRCn and cloned into pCMV_Sports6, followed by nCas9$_{D10A}$ or nCas9$_{VQR/D10A}$ respectively. The two modules are separated by a self-cleavable T2A peptide. To generate gRNA_2xMS2 vector, the gRNA scaffold fused to 2 MS2 loops was synthesized (Genewiz) and cloned into phU6_gRNA (Addgene plasmid # 53188).[89] nfEGFP gene harbors an A→G mutation at nucleotide 200 of the GFP gene. It was synthesized (Genewiz) and cloned into pCMV_Sports6 vector. All plasmid sequences are detailed in **Appendix** (Mammalian Expression Plasmid Sequences section).

gRNA design: nfEGFP targeting gRNAs were designed manually on SnapGene Viewer (GSL Biotech). All gRNAs cover the mutation region (**Figure 4B**) and are listed in **Table 3**.

Cell culture: HEK 293T cells were purchased from ATCC (CRL-3216). Mutant nfEGFP gene was inserted in HEK 293T genome by standard lentiviral transduction and selected by puromycin. Cells expressing mutant nfEGFP were obtained by limiting dilution, generating cell line nf2.16. Cells were grown and maintained at 37° C in Dubelco's Modified Eagle Medium (DMEM, ThermoFisher), supplemented with 10% fetal bovine serum, 1x Glutamine (ThermoFisher) and 1 x Antibiotic-Antimycotic (ThermoFisher).

Transfections: HEK 293T or its derivative nf2.16 cells were plated in 6-well plates the day before experiments. Transfections were performed on cells 75-85% confluent, with a total of 2 µg of combination of DNA from CRC and gRNA constructs in a 3:1 ratio, respectively. Lipofectamine 2000 (ThermoFisher) was used as transfection reagent, following manufacturer's procedure. 72 hours after transfection, fluorescent pictures were taken, GFP signal was quantified by flow cytometry in a Gallios Flow Cytometer instrument (Beckman Coulter) and DNA was purified for downstream analysis using DNeasy Blood and Tissue Kit (Qiagen).

FACS analysis: nf2.16 cells were treated with ^{A}CRCnu/nfEGFP_NT1. 72 hours after transfection, cells were detached from wells and GFP positive cells were sorted on a Beckman Coulter MoFlo XDP Cell Sorter instrument following manufacturer instructions. Sorted cells expressing wild type GFP were cultured, DNA was harvested, and the target region was amplified by PCR followed by Sanger sequencing at Genewiz. Primers used

for PCR were the same as the ones used for high throughput sequencing analysis (see below).

Whole-exome sequencing analysis: WES was carried out by Genewiz (New Jersey, USA). The WES libraries were constructed using Agilent SureSelect Human All Exon (V6 r2) library prep kit and sequenced using Illumina HiSeq with the pair-end 2x150bp format. To estimate potential CRC off-target activity, raw data was analyzed as follows:

- **Variant calling and alternative reference construction**

  WES raw reads were aligned to the human reference genome (hg38) with bwa (version 0.7.15). Variants were identified using Genome Analysis Tool Kit (GATK) version 3.8 roughly following the GATK best practices (REF). Briefly, duplicate reads were first marked with Picard MarkDuplicates. BaseRecalibrator was used to recalibrate base quality, and HaplotypeCaller was then used to call variants on each sample followed by joint genotyping with GenotypeGVCFs. The detected variants in the resulting VCF file were further recalibrated with VariantRecalibrator.

  In the downstream analysis, we only focused on the exonic regions as defined in "SureSelect Human All Exon V6 r2". In the analysis, the overlapping regions were merged using function bedtools merge.

  To construct alternate reference based on the parental cell line T6, we extracted all variants that are genotyped in T6. GATK3.8 FastaAlternateReferenceMaker

was used with default options to construct alternate reference sequence in exonic regions specified in the merged exon-target file.

- **Motif definition and mutation analysis**

AID “WRCH” binding motifs represent a product of ['AT','AG','C','ACT'], and coordinates for any such four consecutive nucleotides were stored. We used python to identify and extract genomic locations of WRCH motifs within reference fasta sequence (either hg38 or alternate reference). Reference fasta sequence was also scanned for sequences complementary to “WRCH”, i.e. “DGYW”, and given by the product of ['AGT','G','CT','AT']. We defined non-WRCH-motif as a four-nucleotide sequence with a C on the third position, which is not WRCH. Similarly, non-DGYW-motif is any four-nucleotide sequence with a G on the second position and not DGYW. In total, there are 12 possible WRCH motifs, 12 DGYW motifs, 52 non-WRCH-motifs, and 52 non-DGYW-motifs. In the mutation analysis we look at WRCH and DGYW categories separately. When looking for potential AID-derived mutated sites a C>T change is categorized as a WRCH motif mutation or a non-WRCH-motif mutation based on its surrounding bases. Similarly, G>A changes are categorized as a DGYW motif mutation or a non-DGYW-motif mutation based on their surrounding bases.

- **Putative CRISPR off-target regions**

Reference genome hg38 was scanned for the putative loci of CRISPR gRNA targeting using CCTop (https://crispr.cos.uni-heidelberg.de/) and CRISPRDesign

(http://crispr.mit.edu/). Together we obtained 54 putative off target regions and variants within these regions were extracted.

High-throughput sequencing analysis: Target region of nfEGFP from treated and untreated samples was amplified using high fidelity Phusion Hot Start Flex DNA polymerase (New England Biolabs) with the following primers: 5'-ggcaagctgaccctgaagttc-3' and 5'-catggcggacttgaagaagtc-3'. PCR products were purified with QIAquick PCR Purification Kit (Qiagen) and submitted to Genewiz for high-throughput sequencing analysis (Aplicon-EZ NGS service). Data analysis, specifically frequency of single nucleotide polymorphisms (SNPs) and insertions/deletions (indels), was performed by Genewiz personnel using a proprietary pipeline.

CHAPTER IV

Targeting Endogenous Loci in the Human Genome

Therapeutic genome editing requires efficient on-target outcomes with reduced off-target effects.[10] Despite the limitations that these technologies currently face for clinical implementation, there are several clinical trials underway, testing safety, efficacy and delivery of genome editing systems in humans.[12,90] Because of its relative ease of use and scalability, CRISPR is poised to dominate the therapeutic landscape, making it the most attractive technology to develop novel applications with therapeutic value. However, most of ongoing clinical trials exploit NHEJ to knock out genes involved in disease by targeted gene deletion.[90] Precise genome modifications, on the other hand, are more difficult to achieve, since NHEJ is the predominant DSB repair pathway, even in dividing cells.[12,91] Moreover, current clinical efforts are aimed to edit cells ex vivo, limiting genome editing interventions to tissues that can be cultured in laboratory settings and amenable of autografts. Therefore, diseases that impact somatic tissues, such as liver, lung, muscle, etc., would not benefit from current genome editing therapeutic applications.

Base editing is a revolutionary genome editing technology that edits single bases without inducing DSB at high rates.[92] These technologies could relieve the limitations that standard CRISPR systems face. . In **Chapters II** and **III**, extensive proof of principle of

the effectiveness and specificity of CRC system was presented. The therapeutic applications of base editing could include correction of disease-causing point mutations, introduction of second-site suppressor mutations, introduction of premature stop codons to knock out genes without breaking the chromosome, deletion of intron-exon junctions for exon skipping approaches. The aim of this Chapter is to explore the capabilities of the CRC system in modifying human genome, including targeting of endogenous genomic sites and an exon skipping proof of principle application, by targeting the human DMD gene, involved in Duchenne muscular dystrophy.

Results

CRC is able to edit target Cs in humane genomic loci

To determine CRC's ability to modify endogenous loci in the human genome, we targeted regions that have been extensively studied for standard nuclease-dependent CRISPR[93,94] as well as base editing applications[48]. These sites and their targeting gRNAs are listed in **Table 4**. As with nfEGFP transgene targeting, CRC treatment resulted in significant C→T conversion, with high purity (i.e. low transversion frequency) and low, but detectable, indel formation (**Figure 8**). These observations further confirm that CRC is a suitable base editing platform, capable of targeting endogenous human genomic sequences.

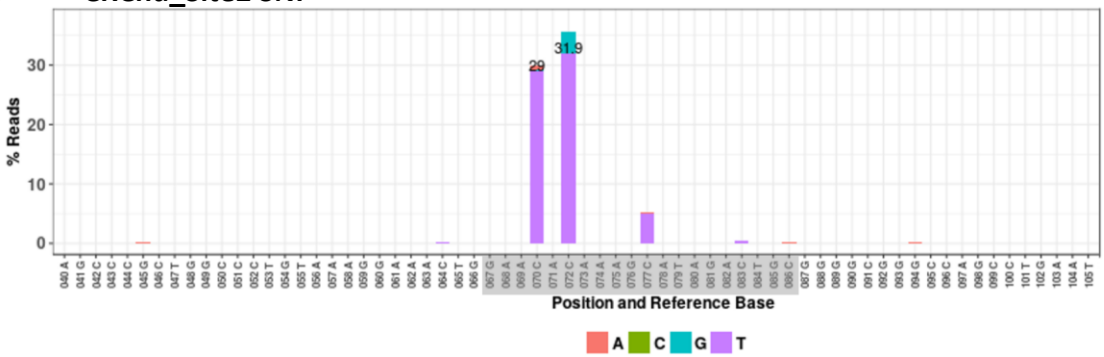
Table 4. Target sequence and genomic locations of endogenous HEK293 loci

Site	Target sequence	PAM	Genomic coordinates (GRCh38/hg38)
2	GAACACAAAGCATAGACTGC	GGG	chr5:87,944,780-87,944,799
3	GGCCCAGACTGAGCACGTGA	TGG	chr9:107,422,339-107,422,358
4	GGCACTGCGGCTGGAGGTGG	GGG	chr20:32,761,950-32,761,969

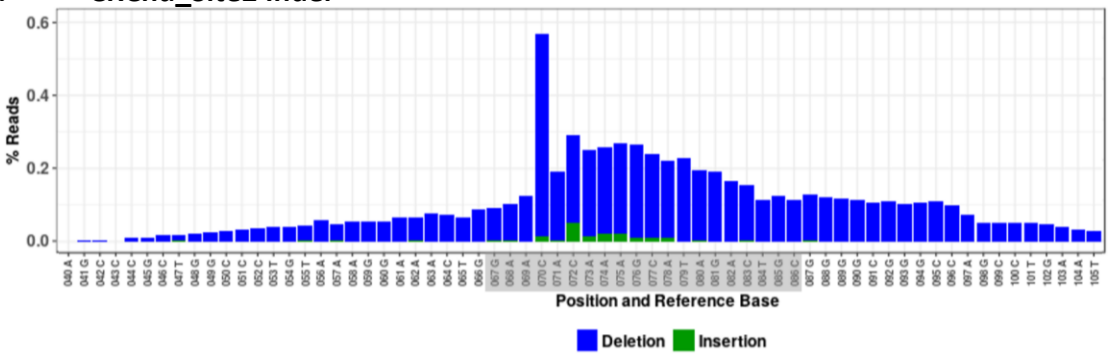
In the case of HEK293 site2 (**Figure 8A**), higher nucleotide conversion was achieved with ^{A}CRCnu treatment compared to the reported BE3 performance (i.e. 29% for C4 and 32% for C6 after CRC treatment, compared to 18% and 15% for the same positions following BE3 treatments).[48] This observation highlights the importance of alternative base editing platforms that may have different targeting performances in specific genomic loci. Similarly, ^{A}CRCnu treatment at site3 and site4 resulted in efficient nucleotide conversion (**Figure 8E, I**). Indel formation followed a similar pattern as observed in nfEGFP targeting (**Figure 7A – E**), which distributions seem to be related to some extent to the gRNA binding sites (**Figure 8B, D, F, H, J**). In Site3 we observed a G→A mutation with high frequency downstream of the target site, however it was also detected in the untreated control (**Figure 8E**, **G**), meaning the this mutation was present in the cell line before the experiments were performed, ruling out any potential unintended bystander effect induced by CRC activity. It is worth noting that, for these targets, CRC seems to have a broader activity window than base editors. In ^{A}CRCnu treatments, inner Cs are being targeted with relatively high efficiency (C11 in Site2, C9 in Site3 and C8 in Site4, **Figure 8A**, **E**, **I**, respectively), whereas BE3 reportedly does not reach these positions.[48] Because base editing is greatly constrained by PAM availability and the relative position of the

target nucleotide within the protospacer, it could be advantageous to have systems with differences in activity window width.

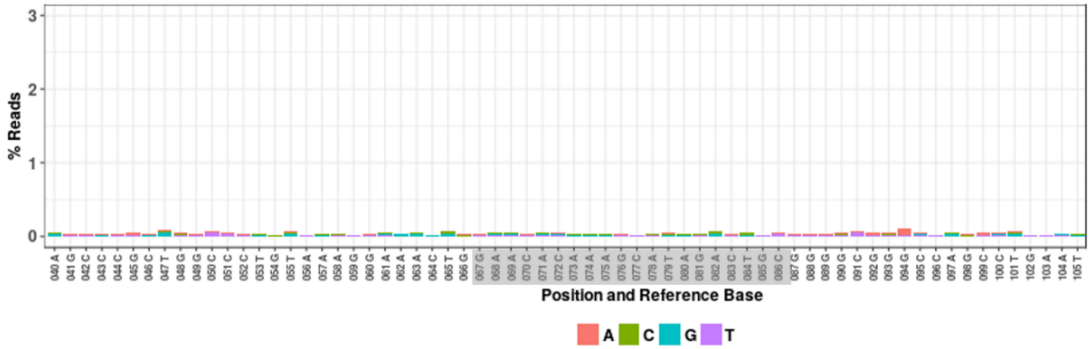
A. **^{A}CRCnu_Site2 SNP**



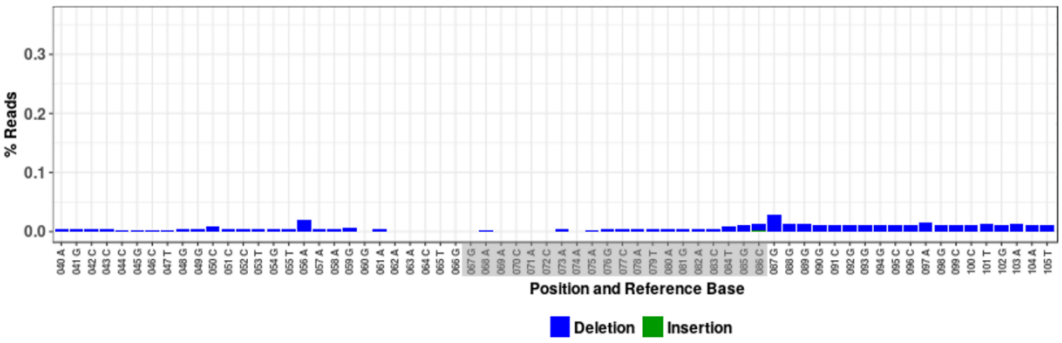
B. **^{A}CRCnu_Site2 Indel**



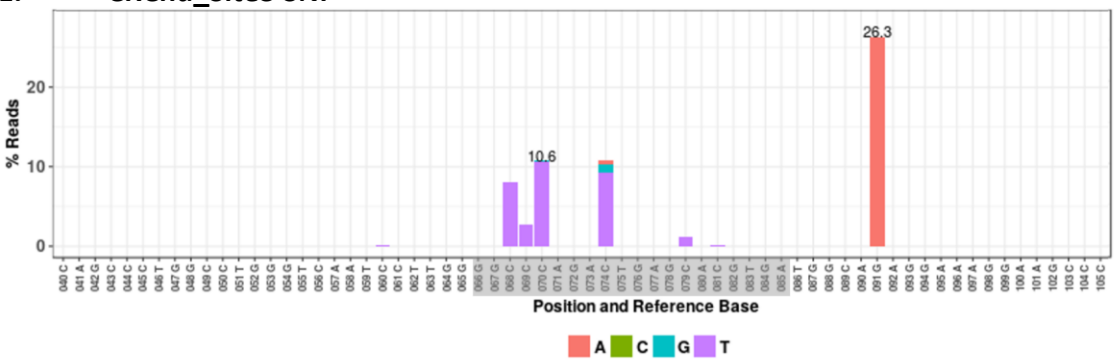
C. **Untreated Site2 SNP**



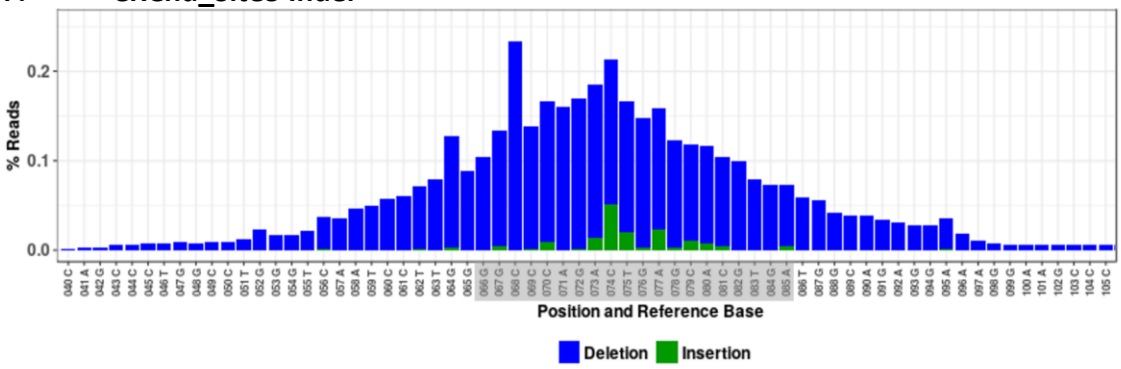
D. Untreated Site2 Indel



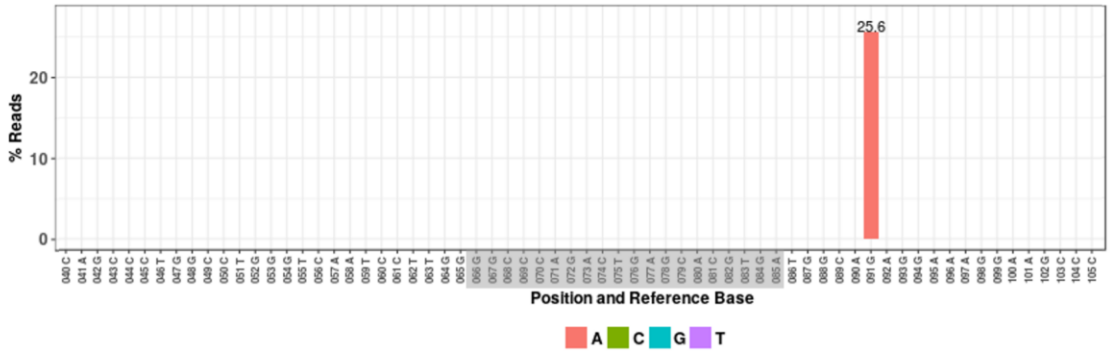
E. ACRCnu_Site3 SNP



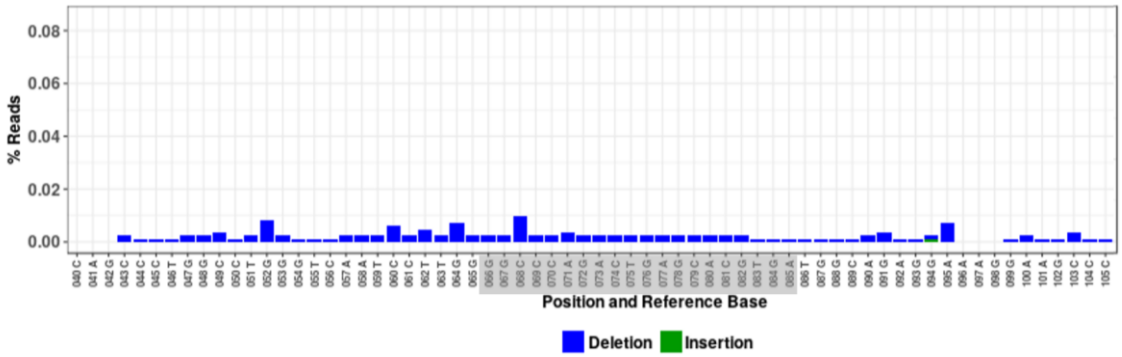
F. ACRCnu_Site3 Indel



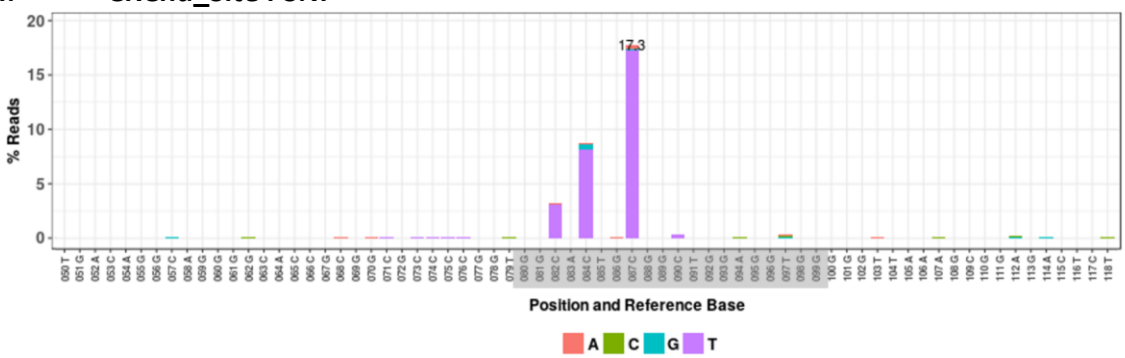
G. Untreated Site3 SNP



H. Untreated Site3 Indel



I. ACRCnu_Site4 SNP



J. ACRCnu_Site4 Indel

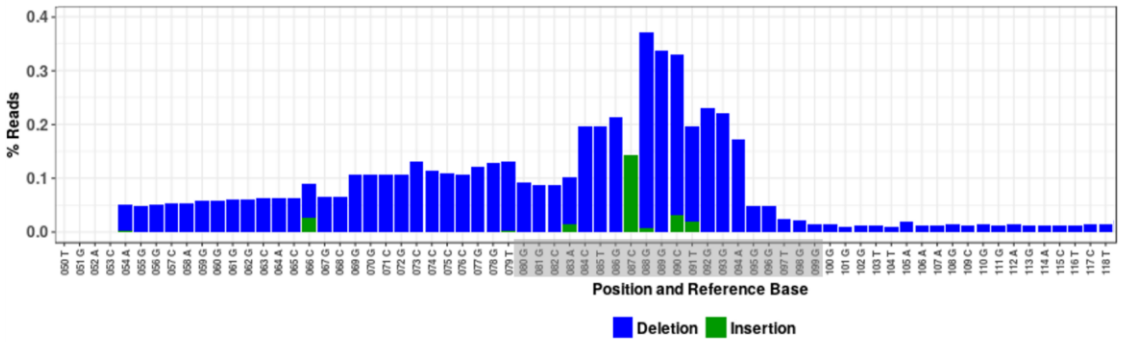


Figure 8. ACRCnu edits endogenous targets in human genome

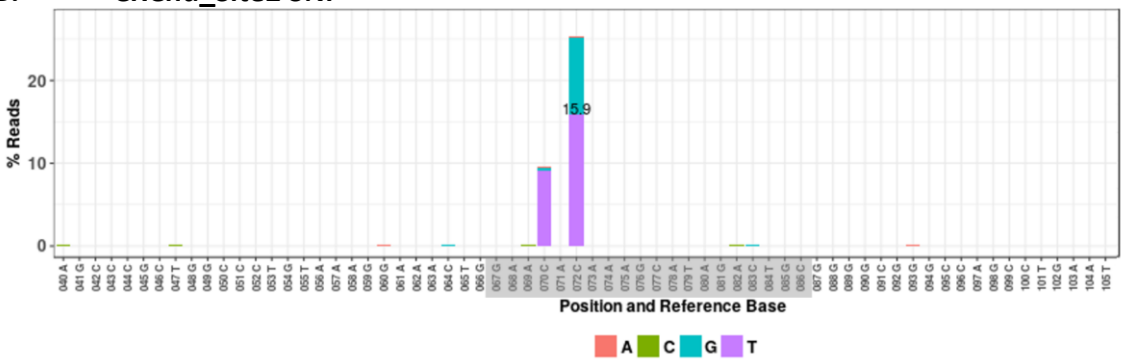
A – J. Base editing outcomes and indel formation after targeting endogenous human loci with ACRCnu and the indicated gRNAs. Numbers below sequences represent the base position within the sequenced amplicon.

APOBEC1 as CRC effector has lower efficiency and purity than AID

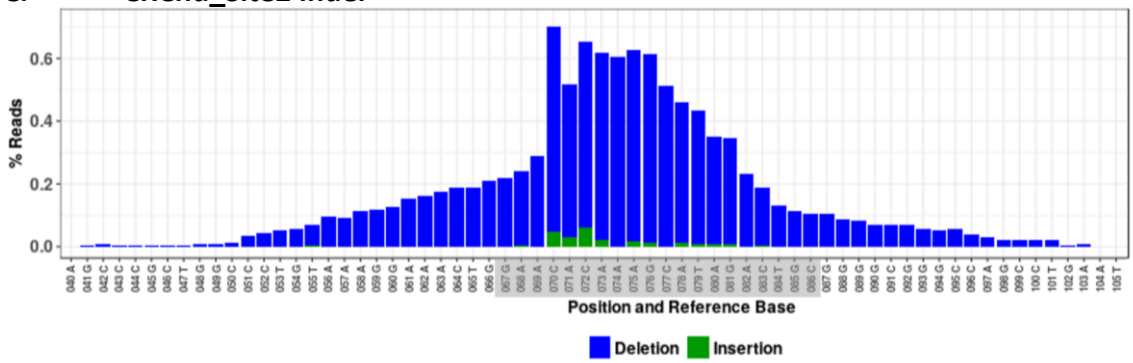
To test if activity window is determined by the effector protein, we decided to construct a CRC system with APOBEC1 as effector (^{A1}CRCnu), following the same general architecture of its AID expressing counterpart (**Figure 9A**). We used the same gRNAs targeting Site2, Site3 and Site4 (**Figure 9B – G**). The results suggest that indeed, ^{A1}CRCnu has a narrower activity widow than ^{A}CRCnu, with internal Cs being edited at much lower frequencies. However, the overall efficiency and base editing purity are lower in ^{A1}CRCnu (**Figure 9B, D, F**). Indel formation was observed in all experiments, suggesting that this phenomenon is not effector specific (**Figure 9C**, **E**, **G**). Because efficiency was lower in APOBEC1 expressing CRC, we decided to continue using ^{A}CRCnu for subsequent experiments.

A.

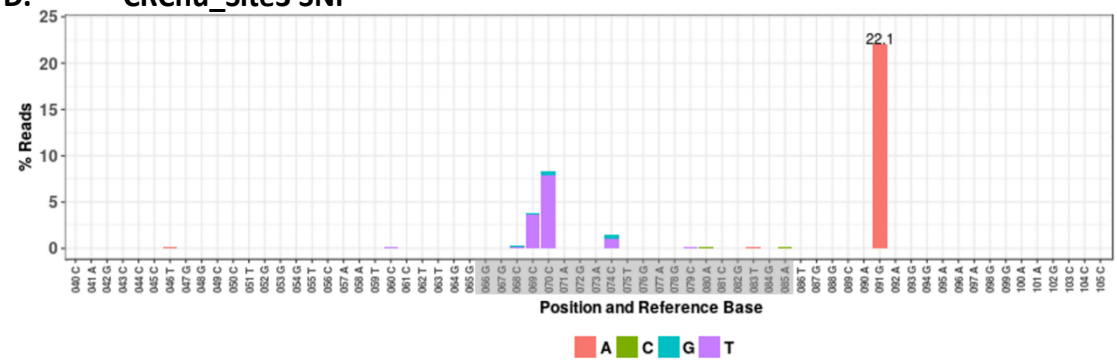
B. **^{A1}CRCnu_Site2 SNP**



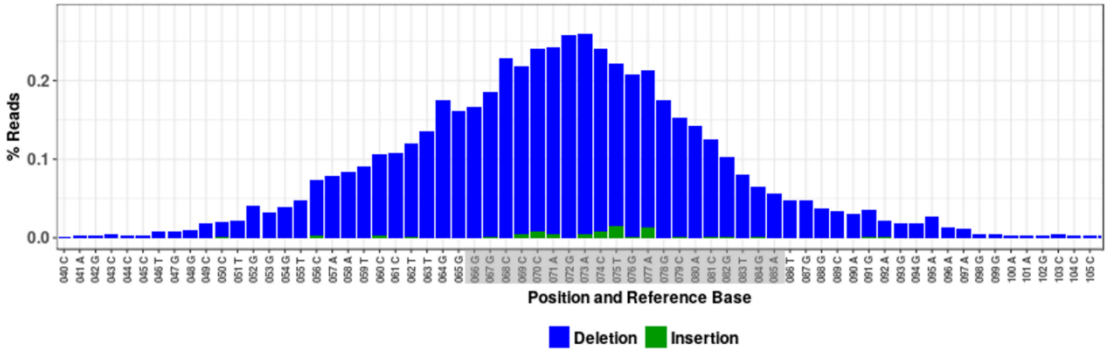
C. **A1CRCnu_Site2 Indel**



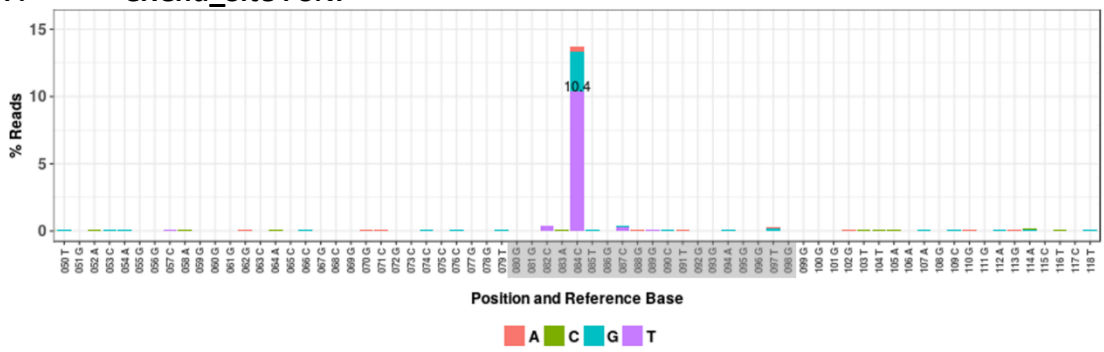
D. **A1CRCnu_Site3 SNP**



E. **A1CRCnu_Site3 Indel**



F. **A1CRCnu_Site4 SNP**



G. ^{A1}CRCnu_Site4 Indel

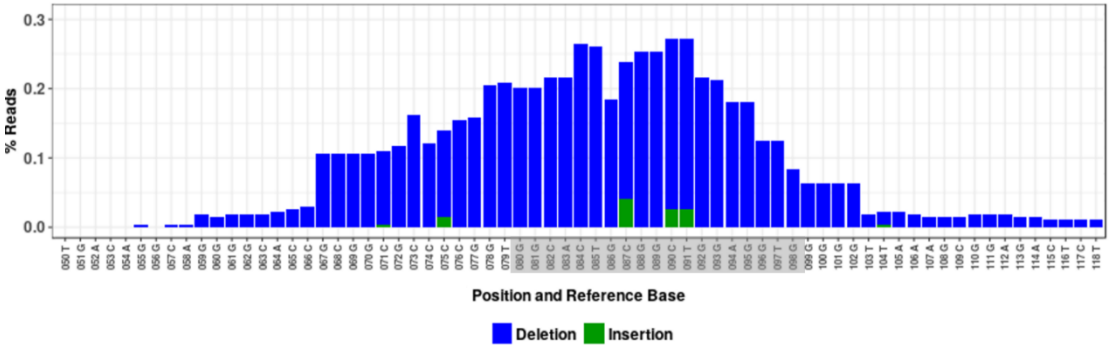


Figure 9. APOBEC1 as CRC effector in the human genome
A – G. Base editing outcomes and indel formation after targeting endogenous human loci with ^{A1}CRCnu and the indicated gRNAs. Numbers below sequences represent the base position within the sequenced amplicon.

CRC induces abrogation of 3' splice acceptor site in intron 50 of human DMD

Duchenne muscular dystrophy (DMD; OMIM 310200) is an X-linked genetic disorder that produces progressive muscular degeneration in affected males.[95] This disease arises from mutations in the *DMD* gene, which encodes for an important muscle cell-specific protein, dystrophin. There is a wide range of mutations involved in DMD, including large deletions of entire exons. For example, 13% of patients present deletions of exon 50, which in turn disrupts the open reading frame (ORF) of the transcript introducing a premature stop codon in exon 51, thereby truncating the protein.[49] Exon 51 skipping restores the ORF and produces a truncated yet functional dystrophin protein, similar to a milder related disease called Becker muscular dystrophy.[50,51] There is an FDA approved antisense oligonucleotide drug based on morpholino chemistry (Eteplirsen, commercially known as Exondys 51), which induces exon 51 skipping, with relatively

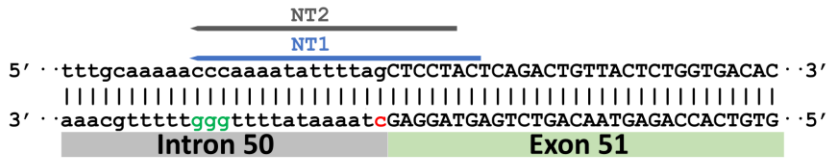
low efficiency (less than 1% improvement in dystrophin expression).[96] Thus, a more efficient and lasting method to induce exon 51 skipping would be beneficial for DMD patients with this particular genetic background. Although nuclease-dependent gene editing has proved effective in achieving this outcome in animal models,[97,98] an approach that avoids introduction of DSB could potentially expedite its clinical application.

As mentioned earlier in this Chapter, a tantalizing application for base editing is to induce therapeutic exon skipping by destroying the 3' splice acceptor site in a target intron – exon junction. The consensus sequence that signal the 3' intron-exon boundary is almost invariably AG in human genes,[99,100] making the G in this motif potentially targetable by CRC, effectively editing the sequence to AA, thereby deleting this regulatory sequence. To test this approach, we designed two gRNAs targeting the intron-exon boundary between intron 50 and exon 51 of the human DMD gene in HEK293 cells (**Figure 10A**). Because there was a single PAM sequence available at the target site, the gRNAs were designed to use the same PAM, varying in length between them (i.e. DMD_NT1 is 19-mer gRNA, DMD_NT2 is 17-mer gRNA; **Figure 10A**). Treatment with both ^{A}CRCnu/DMD_NT1 and NT2 resulted G→A conversion at the target site, with no bystander effect. Although treatment with both gRNAs resulted in similar low efficiencies (around 0.5%; **Figure 10B**, **D**, **F**), this constitutes an exciting proof of concept of the use base editing to treat DMD, considering that the approved antisense oligonucleotide drug has less than 1% of efficiency in restoring dystrophin expression.[96]

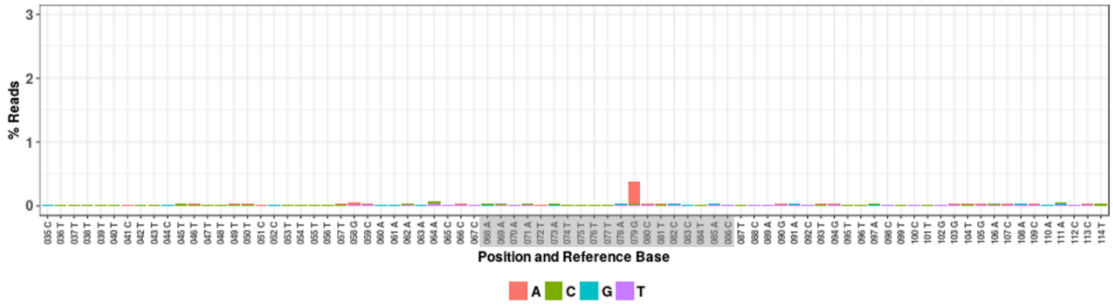
However, higher base editing efficiency is pivotal to develop an effective therapeutic alternative to antisense oligonucleotide-based treatments.

CRC targeting efficiency can be enhanced in multiple ways. For example, chronic treatment could build up the effect, as shown in our GFP reverse mutation assay (**Figure 4E**). Alternatively, by using a novel Cas9 variant with simplified PAM sequence requirement (NG),[101] we could extend the gRNA length to strengthen target binding and to improve the relative position of the target C within the protospacer. Another way to improve targeting efficiency is to optimize the constructs themselves. We have shown that the system is suitable for optimization through vector engineering (**Figure 3, 4F, 5A – B)**. A simple method to enhance construct performance is to optimize the codon usage according to the host cell preferences.[102] Furthermore, codon optimization has been shown to be an effective approach for improving vector expression and editing efficiency in base editing systems.[62,103] This strategy will be discussed in "**Future Directions: CRC Optimization for Improved Expression in Human Cells**" section, in **Chapter V**. Lastly, a possible additional explanation for the observed low conversion efficiency could stem from the cell line used. *DMD* is a muscle-specific gene and these proof of concept experiments were performed in HEK 293 cells, which are kidney-derived, hence these cells do not express *DMD*. Heterochromatin environment present in silenced genes can negatively affect CRISPR targeting,[104-106] therefore a muscle derived cell line with active *DMD* expression, ideally patient-derived, would be a better model for these experiments.

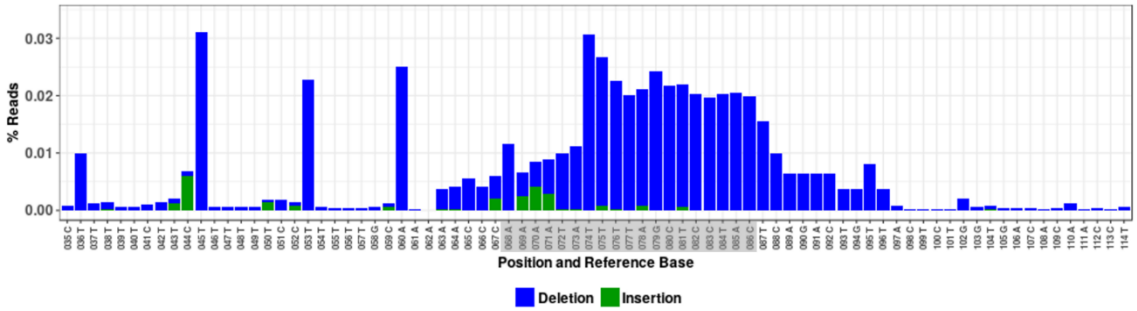
A.



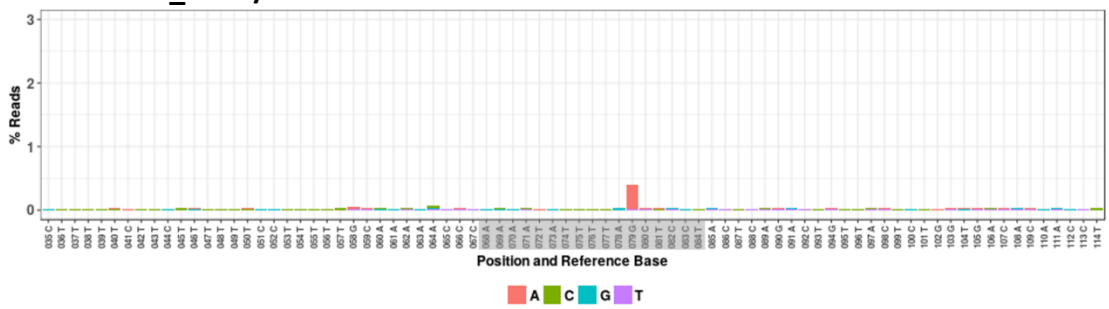
B. **ACRCnu_DMD/NT1 SNP**



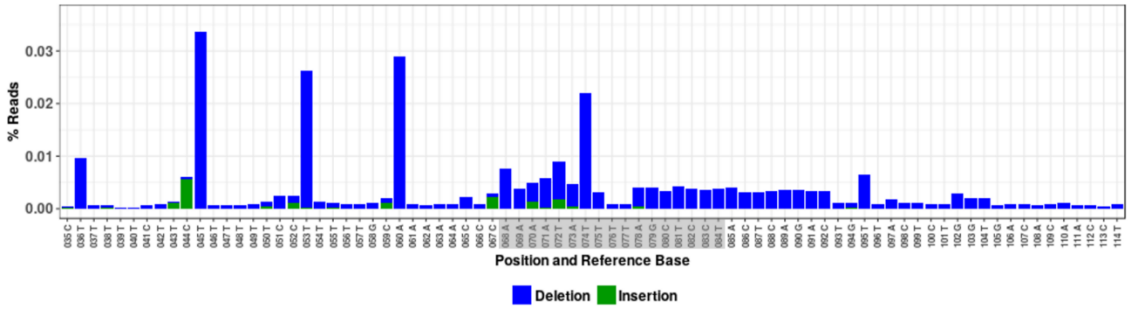
C. **ACRCnu_DMD/NT1 Indel**



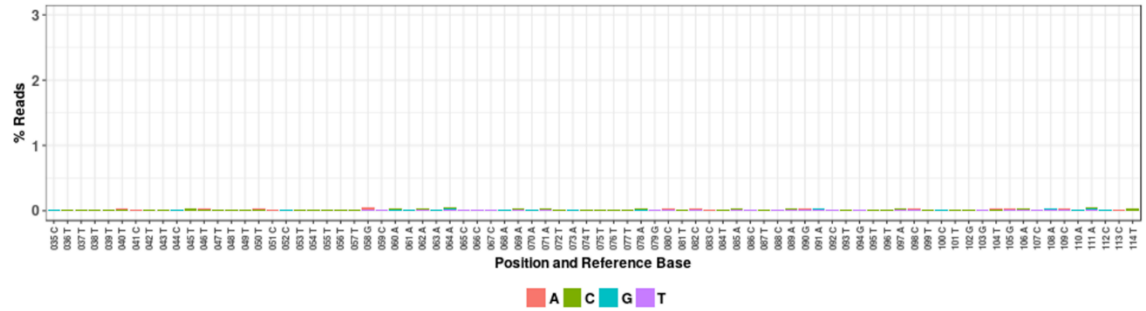
D. **ACRCnu_DMD/NT2 SNP**



E. **^{A}CRCnu_DMD/NT2 Indel**



F. **Untreated DMD In50-Ex51 SNP**



G. **Untreated DMD In50-Ex51 Indel**

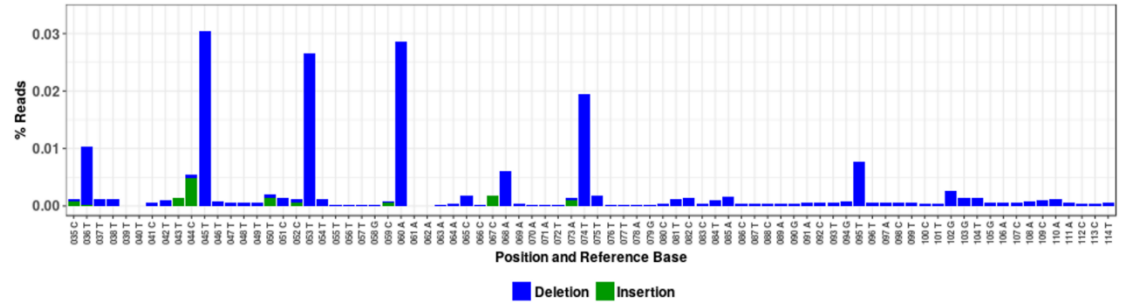


Figure 10. CRC induced abrogation of 3' splice acceptor site in intron 50 of human DMD.

A. Schematic representation of the target site at the intron-exon boundary of intron 50 and exon 51 of the human *DMD* gene. Two gRNAs were designed to target the single PAM motif available at the target site (shown in green). DMD_NT1 and NT2 are 19-mer and 17-mer gRNAs, respectively. In red is shown the target C on the complementary stand of the 3' splice acceptor site (AG). **B – G.** Base editing outcomes and indel formation after targeting endogenous human loci with ^{A}CRCnu and the indicated gRNAs. Note that NT2 treatment (17-mer gRNA) induces lower indel frequency than NT1 (19-mer gRNA), almost at background level (**C**, **E**, **G**), while maintaining the same on-target

base editing efficiency (**B**, **D**, **F**). Numbers below sequences represent the base position within the sequenced amplicon.

Interestingly, the shorter gRNA (DMD_NT2) induced less indels than DMD_NT1 (**Figure 10C**, **E**, **G**). Truncated gRNAs (tru-gRNAs) of 17 nucleotides in length are known to increase specificity by weakening off-target binding while maintaining on-target performance.[30] This could explain the similar on-target efficiency observed with both gRNAs (**Figure 10B**, **D**), while NT2 (a tru-gRNA) reduced indel frequency almost to background levels (**Figure 10C**, **E**, **G**). This suggests that tru-gRNAs could be used to guide base editing platforms in order to reduce unintended on-target mutagenesis by mitigating indel formation, in addition to reducing off-target activity. Note that although indel distribution follows a similar pattern in DMD_NT1 treatment as previously observed, the overall indel frequency, is 10-20-fold lower than those observed in our earlier experiments (compare **Figure 10C**, **E** to **Figure 7A – C** and **Figure 8B**, **F**, **J**). This lower indel formation could be correlated with the possible lack of access to this genomic locus in HEK 293 cells, since these cells do not express *DMD*. Taken together, the data show that therapeutic exon skipping in DMD is a feasible application for CRC-mediated base editing. Nonetheless, the system requires optimization to increase efficiency. Some strategies to improve CRC system will be discussed in **Chapter V**.

Discussion

To test CRC's ability to modify human genomic sites, we targeted three well characterized endogenous loci[93] (**Table 4**). ^{A}CRCnu-mediated base editing was observed in all the targeted sequences tested (**Figure 8**), showing that our base editing system can efficiently modify the human genome. Indel formation was detected around the target sites (**Figure 8B**, **F**, **J**), with a distribution apparently related to gRNA binding, confirming our prior observations in our GFP reporter experiments (**Figure 7 A – D**). Although indel frequencies are relatively low after CRC treatment, and have been observed in other base editing systems,[48] it is necessary to address this phenomenon and find ways to mitigate it. Strategies such as titrating the delivered constructs to find optimal dose, or using truncated gRNAs (**Figure 10E**, see further discussion below) could reduce indels. Additionally, adding DNA binding peptides such as Gam protein has been shown to also reduce base editing-related indels.[60] It would be simple to incorporate such protein into CRC configuration by using heterologous recruiting aptamers (e.g. PP7 or Com stem loops) in the gRNA module and fusing Gam to their respective RNA-binding proteins (e.g. PCP and Com peptides, respectively). In this way, gRNAs fused to heterologous aptamers could simultaneously recruit the effector module as well as Gam fusion protein.

We previously showed that $^{A1}CRC_{D10A}$, where APOBEC1 is the effector protein, has high activity in prokaryotic cells, even inducing double mutations as observed in our

rifampicin negative selection assay (**Figure 3C, D**). Thus, in addition to AID as effector protein in our human experiments, we decided to test the related protein APOBEC1 (^{A1}CRCnu) targeted to the same genomic sites. The result show that ^{A1}CRCnu was able to edit target bases, with a narrower activity widow than ^{A}CRCnu, though the overall base editing efficiency and product purity was lower in ^{A1}CRCnu. Additionally, ^{A1}CRCnu induced similar frequency of indel formation than ^{A}CRCnu, suggesting that this effect is not effector specific, but rather might be the result of the cellular response to the DNA lesions that CRC introduce. ^{A1}CRCnu will be revisited for the second-generation CRC constructs, described in "**Future Directions: CRC Optimization for Improved Expression in Human Cells**" section in **Chapter V**.

To continue testing CRC's editing capabilities in a clinically relevant target, we aimed the system to the human *DMD* gene. Mutations in this gene are involved in the development of the most common deadly inherited disorder, Duchenne muscular dystrophy (DMD).[51] Around 13% of patients suffering from this disease have large deletions in the *DMD* gene that disrupt the ORF and truncate the protein by introducing a premature stop codon in exon 51.[49] The only approved drug designed to restore the reading frame in this genetic background, the antisense oligonucleotide Eteplirsen, induces skipping of exon 51 generating a truncated though functional dystrophin protein,[50,51] with less than 1% of efficiency.[96] One of the downsides of this strategy is that it targets the *DMD* transcript and not the gene, meaning that the effect is transient and patients need to take this drug for life. Thus, a treatment with a more lasting

therapeutic effect could dramatically improve the lives of these patients. Motivated by this, we targeted ^{A}CRCnu to the intron 50 – exon 51 boundary using two gRNAs varying in length (i.e. 19-mer and 17-mer guides) to eliminate the AG consensus motif at the 3' splice acceptor site (**Figure 10A**). Both gRNAs used in this experiment resulted in low yet detectable target editing, with approximately 0.5% of the reads showing AG→AA editing outcome. It is important to note that these experiments were performed in HEK 293 cells, which do not express *DMD*, therefore the chromatin environment around the target site may interfere with targeting efficiency. Efforts to increase the efficiency in this target are already underway, and include construct expression optimization, use of patient-derived muscle cells and non-integrating viral delivery (see **Future Directions: CRC Optimization for Improved Expression in Human Cells** in the next Chapter).

An interesting finding in our *DMD* targeting experiments is the lower indel formation observed with the 17-mer gRNA (DMD_NT2) compared to the 19-mer (DMD_NT1) gRNA (**Figure 10C**, **E**). Truncated gRNAs, also known as Tru-gRNAs, which contain targeting sequences with shortened 5' ends (as short as 17 nucleotides in length), have been shown to improve CRISPR specificity. The shorter binding interface is thought to reduce the energy of gRNA-DNA base pairing, making the gRNA/Cas9 complex more sensitive to mismatches, thus reducing off-target activity. This reduced binding energy at the target site could potentially be behind the lower indel frequency observed in ^{A}CRCnu/DMD_NT2 treatment, however more research in needed to confirm and fully characterize these observations.

One question that remains to be addressed is the extent of CRC-induced off-target effects. In our nfEGFP reverse mutation experiments in **Chapter III**, we showed that CRC treatment does not generate detectable off-target effects at exome-wide level (**Figure 7E**, **F**). However, a more thorough analysis should be performed. To this end, known off-target sites of HEK293 sites 2, 3 and 4 will be analyzed by deep sequencing in CRC treated and untreated samples. Likewise, potential off-target sites for the *DMD* targeting gRNAs have been identified using the CRISPR target prediction tool CCTop,[107] and these sites will also be subjected to high throughput sequencing analysis to assess off-target activity.

Together, the data presented in this Chapter clearly show that CRC is a versatile base editing system capable of modifying endogenous human loci. However, higher efficiency and reduced on-target mutagenicity need to be addressed. This is particularly important to develop therapeutic applications. To boost efficiency, several efforts are being currently taken, such as improving vector expression by codon optimization. More research is needed to better understand the mechanisms involved in editing, the cellular response to CRC treatment, the influence of chromatin environment in base editing outcomes, etc.

Materials and Methods

All methods used in this Chapter are described in the **Material and Methods** section of **Chapter III**.

High throughput sequencing: High throughput sequencing analysis was performed as described in **Chapter III**. To generate the amplicons, the primers described below were used. HEK 293 site2 was amplified with 5'- CCTGGCTGAGCTAACTGTGACAG-3' and 5'-GTCAAACTGTGCGTATGACATCATCAG-3'. HEK 293 site3 was amplified with 5'-GCATGCATTTGTAGGCTTGATGC-3' and 5'-GCCCCTGTCTAGGAAAAGCTGTC-3'. HEK 293 site4 was amplified with 5'-CTGGGTGGAAGGAAGGGAGGAAG-3' and 5'-TCAACCCGAACGGAGACACACAC-3'. Human *DMD* intron 50 – exon 51 boundary was amplified with 5'-GAAATTGGCTCTTTAGCTTGTGTTTC and 5'-GCATTTCTAGTTTGGAGATGGCAG-3'. All PCR amplification were performed with high fidelity Phusion Hot Start DNA Polymerase (New England Biolabs), as per manufacturer's instructions.

CHAPTER V

Final Discussion and Future Directions

Here we present a modular base editing system built by a combination of an engineered *S. pyogenes* CRISPR system and RNA-aptamer based recruitment of non-nuclease DNA editing enzymes, such as cytidine deaminases, expanding the scope of base editing systems. Our platform is called CRC (CRISPR/RNA-Aptamer-Mediated Correction and Modification System; **Figure 1A**). We systematically tested CRC's editing capabilities, starting with a bacterial model. Here we aimed the system to modify the bacterial *rpoB* gene, which encodes the β subunit of RNA polymerase. Mutants of this gene, particularly in a region called RRDR (rifampicin resistance determining region), can survive in the presence of rifampicin, a potent antibiotic that binds and inhibits the bacterial RNA polymerase protein, shutting down global transcription.[67] Thus, it is possible to estimate base editing efficiency by calculating the survival fraction after treatment.

By the time this project was conceived, nothing was known about base editing or its underlying mechanisms, consequently we designed our gRNAs manually, using the available PAM sequences in RRDR near critical residues involved in rifampicin resistance that could be edited by cytidine deamination (**Figure 2A**). Efficient base editing was achieved when the gRNA was targeted to the complementary strand of the target

cytosine, with higher efficiency if the latter is positioned closer to the 5' end of the protospacer (i.e. in the PAM distal region of the unpaired strand left by the CRISPR R-loop; compare rpoB_TS3 and TS4 in **Figure 2B**, **C**). Importantly, efficiency can be enhanced by simple engineering processes, such as using Cas9 nickase instead of catalytically dead Cas9, increasing the number of recruiting stem loops, using alternative effector protein or introducing double nicks in cis for difficult targets (**Figure 3**). The data show not only that CRC-mediated base editing is feasible in the prokaryotic genome, but importantly that the system is suitable for enhancement to improve base editing outcomes.

With the knowledge gathered in our bacterial experiments, we reconstructed the system for eukaryotic expression. First, as a proof of principle, we designed a GFP-based reverse mutation assay. GFP fluorescence was abolished by introducing a point A→G point mutation destroying the chromophore sequence (generating a non-fluorescent GFP, nfEGFP), mimicking a loss-of-function mutation. Various gRNAs covering the target region were designed to restore the wild-type sequence, thus functionally recovering GFP fluorescence (**Figure 4A**). Extrachromosomal and chromosomal targeting resulted significant GFP signal detection, with a functional correction of 35% and 9.4%, respectively (**Figure 4C**, **D** and **Figure 5B**, **C**). Interestingly, CRC treatment resulted in higher efficiency than third-[48] and optimized fourth-generation[62] base editors (0.3% and 2.7% GFP positive cells for BE3 and BE4max, respectively; **Figure 4C**, **D** and **Figure 5B**, **C**). Besides regular *S. pyogenes* Cas9, we also tested a Cas9^{VQR} variant in CRC configuration

(^{A}CRCvnu), which recognizes NGA instead of the standard NGG PAM sequence.[77] Although functional correction of GFP after ^{A}CRCvnu was relatively lower than ^{A}CRCnu (**Figure 4B**), these results show that the targeting space can be expanded by incorporating novel Cas9 variants into the CRC toolbox. On-target effect analysis by high-throughput sequencing revealed highly efficient editing of the target nucleotide, up to 60% with ^{A}CRCnu/nfEGFP_NT1 and 25% with ^{A}CRCvnu/nfEGFP_NT3^{VQR} treatments (**Figure 6**), while no off-target effects were detected at exome-wide range (**Figure 7E**, **F**), suggesting that CRC is a safe and efficient base editing system. Together, these results show strong evidence that CRC can perform base editing tasks efficiently in human cells, expanding the available gene editing tools.

Encouraged by the results in our GFP reporter experiments, we probed endogenous sites in the human genome. We show that both ^{A}CRCnu and ^{A1}CRCnu edit these sites, though they show differences in efficiency, activity window width and product purity. ^{A}CRCnu had higher base editing efficiency and better editing purity (i.e. most sequencing reads showed C→T conversion, while a fraction of them were transversions) than ^{A1}CRCnu (**Figure 8** and **Figure 9**, respectively). However, the base editing window appears to be wider in ^{A}CRCnu treated cells than in ^{A1}CRCnu. This observation suggests that the activity window might be determined by the specific effector protein, rather than system architecture, since the activity window of ^{A1}CRCnu is similar to reports in BE3 and BE4max (where effector protein is also APOBEC1).[48,62] Exploring more effectors

and incorporating them in CRC could further improve and expand the base editing scope.

Recently, reports showed that Cas9-mediated DSB induced previously unnoticed on-target mutagenicity that could be potentially deleterious.[74,75] Base editing is designed to avoid DNA breaks altogether. However, we and others[48] have found that base editing interventions still generate indels at low frequencies in human DNA, suggesting that DNA breaks still occur, probably as a result of cellular response to the DNA damage induced by these systems rather than by their intrinsic gene editing mechanism. Interestingly, the distribution and frequency of indels in CRC treated DNA seem to relate to gRNA binding site, with the highest frequency toward the center in most cases (**Figure 7A – D**; **Figure 8B**, **F**, **J**; **Figure 9C**, **E**, **G**). As we note in our *DMD* targeting experiments, using shorter gRNAs could be an appealing strategy to reduce indels at on-target sites (**Figure 10**). Additionally, incorporating the bacteriophage double strand DNA binding protein Gam to base editing systems mitigates indel formation.[60] Therefore, it would be worth exploring other proteins and tactics that could potentially protect the DNA from breaking after base editing treatments to improve the safety profile of these systems.

To continue testing CRC base editing performance, we decided to aim the system to a clinically relevant target: *DMD*, a gene involved in Duchenne muscular dystrophy (DMD), the most common fatal genetic disease.[51] We show that CRC is able to delete the 3'

splice acceptor site of intron 50, which would induce exon 51 skipping. This approach is relevant for 13% of patients who harbor deletions that introduce a premature stop codon in exon 51. Antisense oligonucleotide therapy has been approved to treat DMD patients by inducing exon skipping, however the effect is modest, with less than 1% of dystrophin recovery.[96] Although the base editing outcome after CRC treatment in HEK 293 cells resulted in merely 0.5% G→A conversion at the 3' splice AG motif, this is exciting evidence showing that base editing could be used to induce therapeutic exon skipping in DMD patients, with longer lasting effects than antisense oligonucleotide therapeutics. The low efficiency observed in these is most likely due to target accessibility in HEK 293 cells, since chromatin environment in silenced genes can prevent CRISPR targeting, probably by physically impeding access to target DNA.[104-106] To confirm these results in a more relevant model, future experiments will be performed in patient-derived immortalized myofibroblast cells. In addition, to determine the universality of CRC-based exon skipping treatment, other in-frame exons (i.e. exons multiples of 3 to avoid nonsense mediated mRNA decay) in highly expressed genes should be targeted.

A key feature of CRC is its modularity. Both in bacteria and human experiments we have shown that the system can be revamped by engineering individual modules. Strategies like increasing the number of recruiting RNA-aptamers; incorporating different Cas9 variants for enhanced efficiency, improved specificity or different PAM recognition; testing several effector proteins; vector expression optimization; construct delivery;

etc., can improve CRC's efficiency and increase its targeting scope. This highlights the need for continuing the engineering efforts to tailor the system to specific targets, a desirable attribute for genome editing systems aimed toward personalized medicine applications. Some of these strategies are currently in progress and are described in the following section.

Future Directions: CRC Optimization for Improved Expression in Human Cells

Expression optimization through optimal codon usage has been shown to improve construct expression and base editing outcomes.[62,103] To generate the second-generation CRC system (CRCnu.2), we designed several new constructs, all of them codon optimized for human expression, using GeneScript and Genewiz online codon optimization tools for the effector module, while Cas9_UGI module was derived from the already optimized BE4max construct.[62] The main difference between the novel Cas9_UGI module and the one present in CRCnu constructs, besides codon optimization, is that the former contains two UGI genes and two nuclear localization sequences (NLS), whereas the latter has only one of each. CRCnu.2 constructs are summarized in **Figure 11**.

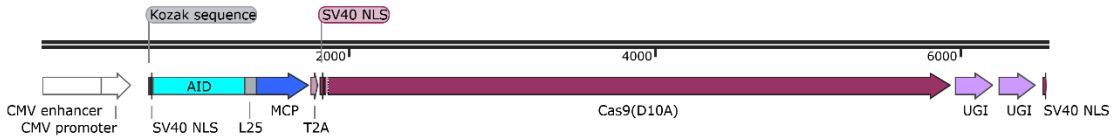
We first recapitulated a basic ^{A}CRCnu construct, generating ^{A}CRCnu.2 (**Figure 11A**). Next, to enhance nuclear presence of the AID_MCP effector module, we incorporated an

additional NLS downstream of MCP (^{2NA}CRCnu.2, **Figure 11B**). AID has an intrinsic nuclear export signal (NES) that could limit its dwelling time in the nucleus[108,109], thus an alternative strategy to accumulate AID_MCP in the nucleus is to delete the C-terminal NES from AID ($^{N\Delta A}$CRCnu.2, **Figure 11C**). Because AID and other deaminases from the APOBEC family work as multimers,[64,83-87] we designed an effector module expressing two non-homologous AID sequences fused through a flexible 32 amino acid linker (L32) to facilitate oligomerization at the site of action (^{2xA}CRCnu.2, **Figure 11D**). Finally, to further expand the CRC toolbox, we engineered an optimized APOBEC1-containing effector module (^{A1}CRCnu.2, **Figure 11E**).

In addition to improvements to the effector and Cas9 modules, we decided to revisit our gRNA design. Engineered gRNAs containing multiple MS2 sequences reduce signal-to-noise ratio in genomic labeling experiments.[53] Thus, we incorporated these scaffolds into our system. The first vector contains six MS2 sequences in tandem at the 3' end of the gRNA[53] (6xMS2_gRNA, **Figure 11F**). The second gRNA harbors two MS2 scaffolds within the gRNA sequence itself, specifically at the tetraloop and stem loop 2, plus four MS2 sequences at the 3' end[53] (2+4xMS2_gRNA, **Figure 11G**).

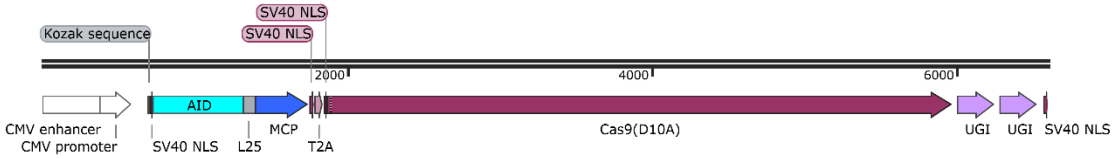
A.

^{A}CRCnu.2



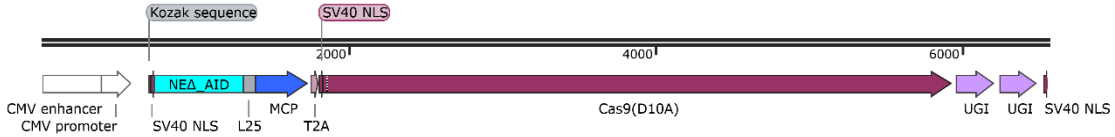
B.

^{2NA}CRCnu.2



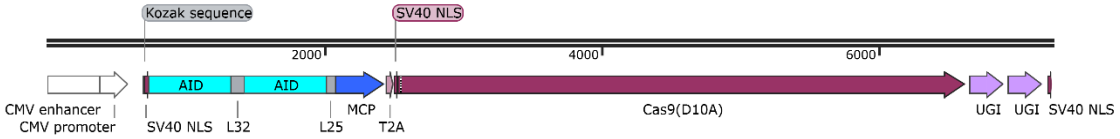
C.

$^{N\Delta A}$CRCnu.2



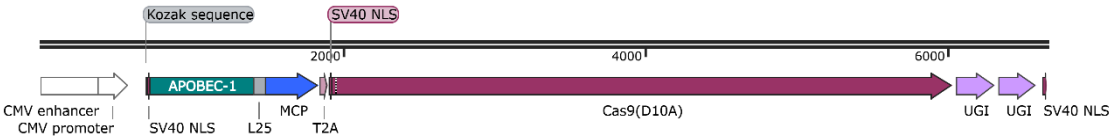
D.

^{2xA}CRCnu.2



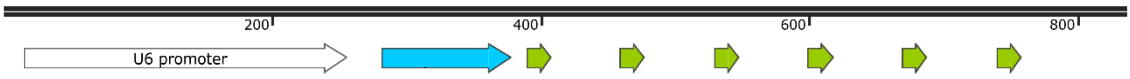
E.

^{A1}CRCnu.2



F.

6xMS2_gRNA



G.

2+4xMS2_gRNA

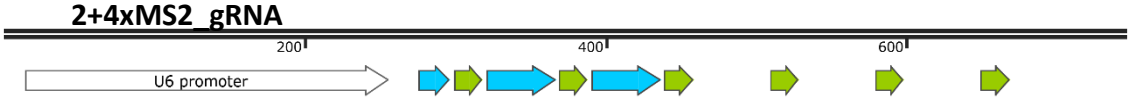


Figure 11. Design of second-generation CRC constructs
A – G. Schematic maps of optimized, second-generation CRC constructs. All protein coding constructs (**A – E**) were codon optimized for efficient expression in human cells; the Cas9 module contains two UGI copies in tandem (purple) and has two NLS at N- and C-terminus. **A.** ^{A}CRCnu.2 expresses standard AID_MCP fusion protein, recapitulating ^{A}CRCnu. **B - C.** To enhance effector module presence and activity in the nucleus, an additional NLS was fused downstream of MCP generating ^{2NA}CRCnu.2 (**B**), or the intrinsic NES sequence was deleted form the C-terminal domain of AID generating $^{N\Delta A}$CRCnu.2 (**C**). **D.** To facilitate the formation of dimers or other oligomers, two copies of AID were fused in tandem to MCP generating ^{2xA}CRCnu.2. **E.** To expand CRC's effector repertoire, an optimized APOBEC1 was fused to MCP generating ^{A1}CRCnu.2. **F - G** Engineered gRNA scaffolds were designed to express 6 MS2 stem loops at the 3' end (6xMS2_gRNA, **F**) or 2 MS2 loops within the gRNA sequence, followed by 4 MS2 sequences at the 3' end (2+4xMS2_gRNA, **G**).

The construction of the second-generation CRC constructs in currently in progress. These optimized systems will be tested on the following targets: nfEGFP, HEK293 site2, site3, site4 and *DMD*, among others. For *DMD* targeting we have obtained an immortalized patient-derived myofibroblast cell line harboring a deletion of exons 45 to 50. This genetic background can be treated by exon 51 skipping, as described earlier, restoring the ORF. To reduce potential cytotoxicity associated with nucleofection of large amounts of DNA in sensitive cell lines, the constructs that show best performance will be packaged in AAV vectors.

REFERENCES

1. Cohen, S.N., Chang, A.C., Boyer, H.W. & Helling, R.B. Construction of biologically functional bacterial plasmids in vitro. *Proceedings of the National Academy of Sciences of the United States of America* **70**, 3240-3244 (1973).
2. Pavletich, N. & Pabo, C. Zinc finger-DNA recognition: crystal structure of a Zif268-DNA complex at 2.1 A. *Science* **252**, 809-817 (1991).
3. Chandrasegaran, S. & Smith, J. Chimeric restriction enzymes: what is next? *Biol Chem* **380**, 841-848 (1999).
4. Kim, Y.G., Cha, J. & Chandrasegaran, S. Hybrid restriction enzymes: Zinc finger fusions to Fok I cleavage domain. *Proceedings of the National Academy of Sciences of the United States of America* **93**, 1156-1160 (1996).
5. Tebas, P.*, et al.* Gene Editing of CCR5 in Autologous CD4 T Cells of Persons Infected with HIV. *New England Journal of Medicine* **370**, 901-910 (2014).
6. Cox, D.B., Platt, R.J. & Zhang, F. Therapeutic genome editing: prospects and challenges. *Nature medicine* **21**, 121-131 (2015).
7. Fujita, T. & Fujii, H. Efficient isolation of specific genomic regions and identification of associated proteins by engineered DNA-binding molecule-mediated chromatin immunoprecipitation (enChIP) using CRISPR. *Biochemical and biophysical research communications* **439**, 132-136 (2013).
8. Maeder, M.L. & Gersbach, C.A. Genome-editing Technologies for Gene and Cell Therapy. *Mol Ther* (2016).
9. Carroll, D. Genome engineering with targetable nucleases. *Annu Rev Biochem* **83**, 409-439 (2014).
10. Corrigan-Curay, J.*, et al.* Genome editing technologies: defining a path to clinic. *Mol Ther* **23**, 796-806 (2015).
11. Carroll, D. A CRISPR approach to gene targeting. *Mol Ther* **20**, 1658-1660 (2012).
12. Yin, H., Kauffman, K.J. & Anderson, D.G. Delivery technologies for genome editing.(Report). *Nature Reviews Drug Discovery* **16**, 387 (2017).
13. Goncz, K.K., Prokopishyn, N.L., Chow, B.L., Davis, B.R. & Gruenert, D.C. Application of SFHR to gene therapy of monogenic disorders. *Gene therapy* **9**, 691-694 (2002).
14. Welter, D.*, et al.* The NHGRI GWAS Catalog, a curated resource of SNP-trait associations. *Nucleic acids research* **42**, D1001-D1006 (2014).
15. Shastry, B.S. SNP alleles in human disease and evolution. *J Hum Genet* **47**, 561-566 (2002).
16. Ochiai, H. Single-Base Pair Genome Editing in Human Cells by Using Site-Specific Endonucleases. *International journal of molecular sciences* **16**, 21128 (2015).
17. Jinek, M.*, et al.* A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science* **337**, 816-821 (2012).
18. Jansen, R., Embden, J.D.A.v., Gaastra, W. & Schouls, L.M. Identification of genes that are associated with DNA repeats in prokaryotes. *Molecular Microbiology* **43**, 1565-1575 (2002).
19. Pennisi, E. The CRISPR Craze. *Science* **341**, 833-836 (2013).
20. Hsu, P.D., Lander, E.S. & Zhang, F. Development and applications of CRISPR-Cas9 for genome engineering. *Cell* **157**, 1262-1278 (2014).
21. Ran, F.A.*, et al.* Genome engineering using the CRISPR-Cas9 system. *Nature protocols* **8**, 2281-2308 (2013).

22. Mali, P., Esvelt, K.M. & Church, G.M. Cas9 as a versatile tool for engineering biology. *Nature methods* **10**, 957-963 (2013).
23. Lin, Y.*, et al.* CRISPR/Cas9 systems have off-target activity with insertions or deletions between target DNA and guide RNA sequences. *Nucleic acids research* (2014).
24. Shen, B.*, et al.* Efficient genome modification by CRISPR-Cas9 nickase with minimal off-target effects. *Nat Meth* **11**, 399-402 (2014).
25. Slaymaker, I.M.*, et al.* Rationally engineered Cas9 nucleases with improved specificity. *Science* **351**, 84-88 (2016).
26. Kleinstiver, B.P.*, et al.* High-fidelity CRISPR–Cas9 nucleases with no detectable genome-wide off-target effects. *Nature* **529**, 490-495 (2016).
27. Guilinger, J.P., Thompson, D.B. & Liu, D.R. Fusion of catalytically inactive Cas9 to FokI nuclease improves the specificity of genome modification. *Nat Biotechnol* **32**, 577-582 (2014).
28. Zhang, H., Yan, Z., Li, M., Peabody, M. & He, T.-C. CRISPR clear? Dimeric Cas9-Fok1 nucleases improve genome-editing specificity. *Genes & Diseases* **1**, 6-7 (2014).
29. Cho, S.W.*, et al.* Analysis of off-target effects of CRISPR/Cas-derived RNA-guided endonucleases and nickases. *Genome research* **24**, 132-141 (2014).
30. Fu, Y.F., Sander, J.D., Reyon, D., Cascio, V.M. & Joung, J.K. Improving CRISPR-Cas nuclease specificity using truncated guide RNAs. *Nature Biotechnology* **32**, 279-284 (2014).
31. Ran, F.A.*, et al.* Double Nicking by RNA-Guided CRISPR Cas9 for Enhanced Genome Editing Specificity. *Cell* **154**, 1380-1389 (2013).
32. Tsai, S.Q. & Joung, J.K. Defining and improving the genome-wide specificities of CRISPR-Cas9 nucleases. *Nature reviews. Genetics* **17**, 300-312 (2016).
33. Tsai, S.Q.*, et al.* Dimeric CRISPR RNA-guided FokI nucleases for highly specific genome editing. *Nat Biotech* **32**, 569-576 (2014).
34. Zhang, X.-H., Tee, L.Y., Wang, X.-G., Huang, Q.-S. & Yang, S.-H. Off-target Effects in CRISPR/Cas9-mediated Genome Engineering. *Mol Ther Nucleic Acids* **4**, e264 (2015).
35. Caldecott, K.W. Mammalian DNA single-strand break repair: an X-ra(y)ted affair. *Bioessays* **23**, 447-455 (2001).
36. Caldecott, K.W. Single-strand break repair and genetic disease. *Nature Reviews Genetics* **9**, 619-631 (2008).
37. Caldecott, K.W. DNA single-strand break repair. *Experimental cell research* **329**, 2-8 (2014).
38. Davis, L. & Maizels, N. Homology-directed repair of DNA nicks via pathways distinct from canonical double-strand break repair. *Proceedings of the National Academy of Sciences of the United States of America* **111**, E924-932 (2014).
39. Perez-Pinera, P.*, et al.* RNA-guided gene activation by CRISPR-Cas9-based transcription factors. *Nat Meth* **10**, 973-976 (2013).
40. Mali, P.*, et al.* CAS9 transcriptional activators for target specificity screening and paired nickases for cooperative genome engineering. *Nat Biotechnol* **31**, 833-838 (2013).
41. Zalatan, J.G.*, et al.* Engineering complex synthetic transcriptional programs with CRISPR RNA scaffolds. *Cell* **160**, 339-350 (2015).
42. Qi, L.S.*, et al.* Repurposing CRISPR as an RNA-guided platform for sequence-specific control of gene expression. *Cell* **152**, 1173-1183 (2013).
43. Larson, M.H.*, et al.* CRISPR interference (CRISPRi) for sequence-specific control of gene expression. *Nature protocols* **8**, 2180-2196 (2013).

44. Hilton, I.B.*, et al.* Epigenome editing by a CRISPR-Cas9-based acetyltransferase activates genes from promoters and enhancers. *Nat Biotech* **33**, 510-517 (2015).
45. Thakore, P.I.*, et al.* Highly specific epigenome editing by CRISPR-Cas9 repressors for silencing of distal regulatory elements. *Nat Meth* **12**, 1143-1149 (2015).
46. Chen, B.*, et al.* Dynamic imaging of genomic loci in living human cells by an optimized CRISPR/Cas system. *Cell* **155**, 1479-1491 (2013).
47. Fu, Y.*, et al.* CRISPR-dCas9 and sgRNA scaffolds enable dual-colour live imaging of satellite sequences and repeat-enriched individual loci. *Nature communications* **7**, 11707 (2016).
48. Komor, A.C., Kim, Y.B., Packer, M.S., Zuris, J.A. & Liu, D.R. Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage. *Nature* **533**, 420-+ (2016).
49. Aartsma-Rus, A.*, et al.* Theoretic applicability of antisense-mediated exon skipping for Duchenne muscular dystrophy mutations. *Human mutation* **30**, 293-299 (2009).
50. Stein, C.A. & Castanotto, D. FDA-Approved Oligonucleotide Therapies in 2017. *Molecular Therapy* **25**, 1069-1075 (2017).
51. Echigoya, Y.*, et al.* Quantitative antisense screening and optimization for exon 51 skipping in Duchenne muscular dystrophy. *Molecular Therapy* **25**, 2561-2572 (2017).
52. Konermann, S.*, et al.* Genome-scale transcriptional activation by an engineered CRISPR-Cas9 complex. *Nature* **517**, 583-588 (2015).
53. Wang, S., Su, J.-H., Zhang, F. & Zhuang, X. An RNA-aptamer-based two-color CRISPR labeling system. *Scientific reports* **6**, 26857 (2016).
54. Liu, X.S.*, et al.* Editing DNA Methylation in the Mammalian Genome. *Cell* **167**, 233-247.e217 (2016).
55. Xiong, T.*, et al.* Targeted DNA methylation in human cells using engineered dCas9-methyltransferases. *Scientific reports* **7**, 6732 (2017).
56. Myhrvold, C.*, et al.* Field-deployable viral diagnostics using CRISPR-Cas13. *Science* **360**, 444-448 (2018).
57. Muramatsu, M.*, et al.* Specific Expression of Activation-induced Cytidine Deaminase (AID), a Novel Member of the RNA-editing Deaminase Family in Germinal Center B Cells. *Journal of Biological Chemistry* **274**, 18470-18476 (1999).
58. Petersen-Mahrt, S.K., Harris, R.S. & Neuberger, M.S. AID mutates E. coli suggesting a DNA deamination mechanism for antibody diversification. *Nature* **418**, 99-104 (2002).
59. Yoshikawa, K.*, et al.* AID enzyme-induced hypermutation in an actively transcribed gene in fibroblasts. *Science* **296**, 2033-2036 (2002).
60. Komor, A.C.*, et al.* Improved base excision repair inhibition and bacteriophage Mu Gam protein yields C:G-to-T:A base editors with higher efficiency and product purity. *Science Advances* **3**(2017).
61. Gaudelli, N.M.*, et al.* Programmable base editing of A•T to G•C in genomic DNA without DNA cleavage. *Nature* (2017).
62. Koblan, L.W.*, et al.* Improving cytidine and adenine base editors by expression optimization and ancestral reconstruction. *Nature Biotechnology* (2018).
63. Di Noia, J.M. & Neuberger, M.S. Molecular Mechanisms of Antibody Somatic Hypermutation. *Annual Review of Biochemistry* **76**, 1-22 (2007).
64. Smith, H.C., Bennett, R.P., Kizilyer, A., McDougall, W.M. & Prohaska, K.M. Functions and regulation of the APOBEC family of proteins. *Semin Cell Dev Biol* **23**, 258-268 (2012).
65. Zan, H. & Casali, P. Regulation of Aicda expression and AID activity. *Autoimmunity* **46**, 83-101 (2013).

66. Campbell, E.A*., et al.* Structural Mechanism for Rifampicin Inhibition of Bacterial RNA Polymerase. *Cell* **104**, 901-912 (2001).
67. Goldstein, B.P. Resistance to rifampicin: a review. *J Antibiot (Tokyo)* **67**, 625-630 (2014).
68. Xu, M., Zhou, Y.N., Goldstein, B.P. & Jin, D.J. Cross-Resistance of Escherichia coli RNA Polymerases Conferring Rifampin Resistance to Different Antibiotics. *Journal of Bacteriology* **187**, 2783-2792 (2005).
69. Conticello, S.G. The AID/APOBEC family of nucleic acid mutators. *Genome biology* **9**, 229 (2008).
70. Nassour, J*., et al.* Defective DNA single-strand break repair is responsible for senescence and neoplastic escape of epithelial cells. *Nature communications* **7**, 10399 (2016).
71. Ma, H*., et al.* Multiplexed labeling of genomic loci with dCas9 and engineered sgRNAs using CRISPRainbow. *Nature Biotechnology* **34**, 528 (2016).
72. Fu, Y.F*., et al.* High-frequency off-target mutagenesis induced by CRISPR-Cas nucleases in human cells. *Nature Biotechnology* **31**, 822-+ (2013).
73. Singh, P., Schimenti, J.C. & Bolcun-Filas, E. A Mouse Geneticist's Practical Guide to CRISPR Applications. *Genetics* (2014).
74. Kosicki, M., Tomberg, K. & Bradley, A. Repair of double-strand breaks induced by CRISPR–Cas9 leads to large deletions and complex rearrangements. *Nature Biotechnology* **36**, 765 (2018).
75. Rivera-Torres, N., Banas, K., Bialk, P., Bloh, K.M. & Kmiec, E.B. Insertional Mutagenesis by CRISPR/Cas9 Ribonucleoprotein Gene Editing in Cells Targeted for Point Mutation Repair Directed by Short Single-Stranded DNA Oligonucleotides. *PloS one* **12**, e0169350 (2017).
76. Iyama, T. & Wilson, D.M. DNA repair mechanisms in dividing and non-dividing cells. *DNA repair* **12**, 620-636 (2013).
77. Kleinstiver, B.P*., et al.* Engineered CRISPR-Cas9 nucleases with altered PAM specificities. *Nature* **523**, 481-U249 (2015).
78. Krokan, H.E. & Bjørås, M. Base excision repair. *Cold Spring Harbor perspectives in biology* **5**, a012583 (2013).
79. Jacobs, A.L. & Schär, P. DNA glycosylases: in DNA repair and beyond. *Chromosoma* **121**, 1-20 (2012).
80. Mol, C.D*., et al.* Crystal structure of human uracil-DNA glycosylase in complex with a protein inhibitor: protein mimicry of DNA. *Cell* **82**, 701-708 (1995).
81. Qin, J.Y*., et al.* Systematic Comparison of Constitutive Promoters and the Doxycycline-Inducible Promoter. *PloS one* **5**, e10611 (2010).
82. Odegard, V.H. & Schatz, D.G. Targeting of somatic hypermutation. *Nat Rev Immunol* **6**, 573-583 (2006).
83. Lau, P.P., Zhu, H.-J., Baldini, A., Charnsangavej, C. & Chan, L. Dimeric structure of a human apolipoprotein B mRNA editing protein and cloning and chromosomal localization of its gene. *Proceedings of the National Academy of Sciences* **91**, 8522-8526 (1994).
84. Shlyakhtenko, L.S*., et al.* Atomic force microscopy studies provide direct evidence for dimerization of the HIV restriction factor APOBEC3G. *Journal of Biological Chemistry* **286**, 3387-3395 (2011).
85. Bohn, M.-F*., et al.* The ssDNA mutator APOBEC3A is regulated by cooperative dimerization. *Structure* **23**, 903-911 (2015).

86. Mondal, S., Begum, N.A., Hu, W. & Honjo, T. Functional requirements of AID's higher order structures and their interaction with RNA-binding proteins. *Proceedings of the National Academy of Sciences* **113**, E1545-E1554 (2016).
87. Qiao, Q.*, et al.* AID Recognizes Structured DNA for Class Switch Recombination. *Molecular cell* **67**, 361-373.e364 (2017).
88. Saraconi, G., Severi, F., Sala, C., Mattiuz, G. & Conticello, S.G. The RNA editing enzyme APOBEC1 induces somatic mutations and a compatible mutational signature is present in esophageal adenocarcinomas. *Genome biology* **15**, 417 (2014).
89. Kabadi, A.M., Ousterout, D.G., Hilton, I.B. & Gersbach, C.A. Multiplex CRISPR/Cas9-based genome engineering from a single lentiviral vector. *Nucleic acids research* **42**, e147-e147 (2014).
90. Cornu, T.I., Mussolino, C. & Cathomen, T. Refining strategies to translate genome editing to the clinic. *Nature medicine* **23**, 415 (2017).
91. Burma, S., Chen, B.P.C. & Chen, D.J. Role of non-homologous end joining (NHEJ) in maintaining genomic integrity. *DNA repair* **5**, 1042-1048 (2006).
92. Komor, A.C., Badran, A.H. & Liu, D.R. Editing the Genome Without Double-Stranded DNA Breaks. *ACS Chemical Biology* (2017).
93. Kuscu, C., Arslan, S., Singh, R., Thorpe, J. & Adli, M. Genome-wide analysis reveals characteristics of off-target sites bound by the Cas9 endonuclease. *Nature Biotechnology* **32**, 677 (2014).
94. Tsai, S.Q.*, et al.* GUIDE-seq enables genome-wide profiling of off-target cleavage by CRISPR-Cas nucleases. *Nature Biotechnology* **33**, 187 (2014).
95. Bowles, D.E.*, et al.* Phase 1 Gene Therapy for Duchenne Muscular Dystrophy Using a Translational Optimized AAV Vector. *Molecular Therapy* **20**, 443-455 (2012).
96. Young, C.S. & Pyle, A.D. Exon Skipping Therapy. *Cell* **167**, 1144 (2016).
97. Amoasii, L.*, et al.* Single-cut genome editing restores dystrophin expression in a new mouse model of muscular dystrophy. *Sci. Transl. Med.* **9**(2017).
98. Amoasii, L.*, et al.* Gene editing restores dystrophin expression in a canine model of Duchenne muscular dystrophy. *Science* **362**, 86-91 (2018).
99. Shapiro, M.B. & Senapathy, P. RNA splice junctions of different classes of eukaryotes: sequence statistics and functional implications in gene expression. *Nucleic acids research* **15**, 7155-7174 (1987).
100. Padgett, R.A. New connections between splicing and human disease. *Trends in Genetics* **28**, 147-154 (2012).
101. Nishimasu, H.*, et al.* Engineered CRISPR-Cas9 nuclease with expanded targeting space. *Science* **361**, 1259-1262 (2018).
102. Fuglsang, A. Codon optimizer: a freeware tool for codon optimization. *Protein expression and purification* **31**, 247-249 (2003).
103. Zafra, M.P.*, et al.* Optimized base editors enable efficient editing in cells, organoids and mice. *Nature biotechnology* **36**, 888 (2018).
104. Chen, X.*, et al.* Probing the impact of chromatin conformation on genome editing tools. *Nucleic acids research* **44**, 6482-6492 (2016).
105. Daer, R.M., Cutts, J.P., Brafman, D.A. & Haynes, K.A. The impact of chromatin dynamics on Cas9-mediated genome editing in human cells. *ACS synthetic biology* **6**, 428-438 (2016).
106. Chen, X., Liu, J., Janssen, J.M. & Gonçalves, M.A. The chromatin structure differentially impacts high-specificity CRISPR-Cas9 nuclease strategies. *Molecular Therapy-Nucleic Acids* **8**, 558-563 (2017).

107. Stemmer, M., Thumberger, T., del Sol Keyer, M., Wittbrodt, J. & Mateo, J.L. CCTop: An Intuitive, Flexible and Reliable CRISPR/Cas9 Target Prediction Tool. *PloS one* **10**, e0124633 (2015).
108. Doi, T.*, et al.* The C-terminal region of activation-induced cytidine deaminase is responsible for a recombination function other than DNA cleavage in class switch recombination. *Proceedings of the National Academy of Sciences of the United States of America* **106**, 2758-2763 (2009).
109. Geisberger, R., Rada, C. & Neuberger, M.S. The stability of AID and its function in class-switching are critically sensitive to the identity of its nuclear-export sequence. *Proceedings of the National Academy of Sciences* **106**, 6736-6741 (2009).

APPENDIX: SEQUENCES OF PLASMIDS DEVELOPED IN THIS STUDY

Bacterial Expression Plasmid Sequences

1. pTet_dCas9-bacteria/p15A

```
LOCUS       Exported                6705 bp ds-DNA     circular SYN 11-APR-2015
DEFINITION  pTet_dCas9/p15A; encodes for nuclease dead Cas9 protein; CRC
       targeting module
ACCESSION   .
VERSION     pTet_dCas9/p15A; CRC targeting module
KEYWORDS    .
SOURCE      synthetic DNA construct
  ORGANISM  synthetic DNA construct
REFERENCE   1  (bases 1 to 6705)
  AUTHORS   .
  TITLE     Direct Submission
  JOURNAL   Exported Sep 29, 2018 from SnapGene Viewer 4.2.5
            http://www.snapgene.com
FEATURES             Location/Qualifiers
     source          1..6705
                     /organism="synthetic DNA construct"
                     /mol_type="other DNA"
     CDS             complement(7..630)
                     /codon_start=1
                     /gene="tetR from transposon Tn10"
                     /product="tetracycline repressor TetR"
                     /label=TetR
                     /note="TetR binds to the tetracycline operator tetO to
                     inhibit transcription. This inhibition can be relieved by
                     adding tetracycline or doxycycline."

/translation="MSRLDKSKVINSALELLNEVGIEGLTTRKLAQKLGVEQPTLYWHV

KNKRALLDALAIEMLDRHHTHFCPLEGESWQDFLRNNAKSFRCALLSHRDGAKVHLGTR

PTEKQYETLENQLAFLCQQGFSLENALYALSAVGHFTLGCVLEDQEHQVAKEERETPTT
                     DSMPPLLRQAIELFDHQGAEPAFLFGLELIICGLEKQLKCESGS"
     promoter        649..704
                     /gene="tetR"
                     /label=tetR/tetA promoters
                     /note="overlapping promoters for bacterial tetR and tetA"
     protein_bind    685..703
                     /gene="tetO"
                     /label=tet operator
                     /bound_moiety="tetracycline repressor TetR"
                     /note="bacterial operator O2 for the tetR and tetA genes"
     CDS             740..4846
                     /codon_start=1
                     /product="catalytically dead mutant of the Cas9
                     endonuclease from the Streptococcus pyogenes Type II
                     CRISPR/Cas system"
                     /label=dCas9
                     /note="RNA-guided DNA-binding protein that lacks
                     endonuclease activity due to the D10 mutation in the
                     RuvC-like domain and the H840A mutation in the HNH domain"
```

```
/translation="MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKK
NLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEES
FLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIK
FRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRL
ENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQ
IGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVR
QQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLL
RKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARG
NSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEY
FTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFD
SVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEER
LKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFM
QLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGR
HKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLY
LYYLQNGRDMYVDQELDINRLSDYDVDAIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVP
SEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKH
VAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYL
NAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKT
EITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSK
ESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGI
TIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNE
LALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILAD
ANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVL
                      DATLIHQSITGLYETRIDLSQLGGD"
     terminator       4870..4941
                      /label=rrnB T1 terminator
                      /note="transcription terminator T1 from the E. coli rrnB
                      gene"
     terminator       4957..4984
                      /label=T7Te terminator
                      /note="phage T7 early transcription terminator"
     rep_origin       complement(5146..5691)
                      /direction=LEFT
                      /label=p15A ori
                      /note="Plasmids containing the medium-copy-number p15A
                      origin of replication can be propagated in E. coli cells
                      that contain a second plasmid with the ColE1 origin."
     terminator       5805..5899
                      /label=lambda t0 terminator
                      /note="transcription terminator from phage lambda"
```

```
     CDS             complement(5920..6579)
                     /codon_start=1
                     /gene="cat"
                     /product="chloramphenicol acetyltransferase"
                     /label=CmR
                     /note="confers resistance to chloramphenicol"

/translation="MEKKITGYTTVDISQWHRKEHFEAFQSVAQCTYNQTVQLDITAFL

KTVKKNKHKFYPAFIHILARLMNAHPEFRMAMKDGELVIWDSVHPCYTVFHEQTETFSS

LWSEYHDDFRQFLHIYSQDVACYGENLAYFPKGFIENMFFVSANPWVSFTSFDLNVANM
                     DNFFAPVFTMGKYYTQGDKVLMPLAIQVHHAVCDGFHVGRMLNELQQYCDEWQGGA"
     promoter        complement(6580..6682)
                     /label=cat promoter
                     /note="promoter of the E. coli cat gene encoding
                     chloramphenicol acetyltransferase"
ORIGIN
        1 gacgtcttaa gacccacttt cacatttaag ttgtttttct aatccgcata tgatcaattc
       61 aaggccgaat aagaaggctg gctctgcacc ttggtgatca aataattcga tagcttgtcg
      121 taataatggc ggcatactat cagtagtagg tgtttccctt tcttctttag cgacttgatg
      181 ctcttgatct tccaatacgc aacctaaagt aaaatgcccc acagcgctga gtgcatataa
      241 tgcattctct agtgaaaaac cttgttggca taaaaaggct aattgatttt cgagagtttc
      301 atactgtttt tctgtaggcc gtgtacctaa atgtactttt gctccatcgc gatgacttag
      361 taaagcacat ctaaaacttt tagcgttatt acgtaaaaaa tcttgccagc tttccccttc
      421 taaagggcaa aagtgagtat ggtgcctatc taacatctca atggctaagg cgtcgagcaa
      481 agcccgctta ttttttacat gccaatacaa tgtaggctgc tctacaccta gcttctgggc
      541 gagtttacgg gttgttaaac cttcgattcc gacctcatta agcagctcta atgcgctgtt
      601 aatcacttta cttttatcta atctagacat cattaattcc taatttttgt tgacactcta
      661 tcgttgatag agttatttta ccactcccta tcagtgatag agaaaagaat tcaaaagatc
      721 taaagaggag aaaggatcta tggataagaa atactcaata ggcttagcta tcggcacaaa
      781 tagcgtcgga tgggcggtga tcactgatga atataaggtt ccgtctaaaa agttcaaggt
      841 tctgggaaat acagaccgcc acagtatcaa aaaaaatctt atagggggctc ttttatttga
      901 cagtggagag acagcggaag cgactcgtct caaacggaca gctcgtagaa ggtatacacg
      961 tcggaagaat cgtatttgtt atctacagga gatttttca aatgagatgg cgaaagtaga
     1021 tgatagtttc tttcatcgac ttgaagagtc ttttttggtg gaagaagaca agaagcatga
     1081 acgtcatcct atttttggaa atatagtaga tgaagttgct tatcatgaga aatatccaac
     1141 tatctatcat ctgcgaaaaa aattggtaga ttctactgat aaagcggatt tgcgcttaat
     1201 ctatttggcc ttagcgcata tgattaagtt tcgtggtcat tttttgattg agggagattt
     1261 aaatcctgat aatagtgatg tggacaaact atttatccag ttggtacaaa cctacaatca
     1321 attatttgaa gaaaacccta ttaacgcaag tggagtagat gctaaagcga ttctttctgc
     1381 acgattgagt aaatcaagac gattagaaaa tctcattgct cagctccccg gtgagaagaa
     1441 aaatggctta tttgggaatc tcattgcttt gtcattgggt ttgaccccta attttaaatc
     1501 aaattttgat ttggcagaag atgctaaatt acagctttca aaagatactt acgatgatga
     1561 tttagataat ttattggcgc aaattggaga tcaatatgct gatttgtttt tggcagctaa
     1621 gaatttatca gatgctattt tactttcaga tatcctaaga gtaaatactg aaataactaa
     1681 ggctccccta tcagcttcaa tgattaaacg ctacgatgaa catcatcaag acttgactct
     1741 tttaaaagct ttagttcgac aacaacttcc agaaaagtat aaagaaatct tttttgatca
     1801 atcaaaaaac ggatatgcag gttatattga tgggggagct agccaagaag aattttataa
     1861 atttatcaaa ccaattttag aaaaaatgga tggtactgag gaattattgg tgaaactaaa
     1921 tcgtgaagat ttgctgcgca agcaacggac ctttgacaac ggctctattc cccatcaaat
     1981 tcacttgggt gagctgcatg ctattttgag aagacaagaa gacttttatc catttttaaa
     2041 agacaatcgt gagaagattg aaaaaatctt gacttttcga attccttatt atgttggtcc
     2101 attggcgcgt ggcaatagtc gttttgcatg gatgactcgg aagtctgaag aaacaattac
     2161 cccatggaat tttgaagaag ttgtcgataa aggtgcttca gctcaatcat ttattgaacg
     2221 catgacaaac tttgataaaa atcttccaaa tgaaaaagta ctaccaaaac atagtttgct
     2281 ttatgagtat tttacggttt ataacgaatt gacaaaggtc aaatatgtta ctgaaggaat
     2341 gcgaaaacca gcatttcttt caggtgaaca gaagaaagcc attgttgatt tactcttcaa
     2401 aacaaatcga aaagtaaccg ttaagcaatt aaaagaagat tatttcaaaa aaatagaatg
     2461 ttttgatagt gttgaaattt caggagttga agatagattt aatgcttcat taggtaccta
     2521 ccatgatttg ctaaaaatta ttaaagataa agatttttg gataatgaag aaaatgaaga
     2581 tatcttagag gatattgttt taacattgac cttatttgaa gatagggaga tgattgagga
     2641 aagacttaaa acatatgctc acctctttga tgataaggtg atgaaacagc ttaaacgtcg
```

2701 ccgttatact ggttggggac gtttgtctcg aaaattgatt aatggtatta gggataagca
2761 atctggcaaa acaatattag atttttgaa atcagatggt tttgccaatc gcaattttat
2821 gcagctgatc catgatgata gtttgacatt taaagaagac attcaaaaag cacaagtgtc
2881 tggacaaggc gatagtttac atgaacatat tgcaaattta gctggtagcc ctgctattaa
2941 aaaaggtatt ttacagactg taaaagttgt tgatgaattg gtcaaagtaa tggggcggca
3001 taagccagaa aatatcgtta ttgaaatggc acgtgaaaat cagacaactc aaaagggcca
3061 gaaaaattcg cgagagcgta tgaaacgaat cgaagaaggt atcaaagaat taggaagtca
3121 gattcttaaa gagcatcctg ttgaaaatac tcaattgcaa aatgaaaagc tctatctcta
3181 ttatctccaa aatggaagag acatgtatgt ggaccaagaa ttagatatta atcgtttaag
3241 tgattatgat gtcgatgcca ttgttccaca aagtttcctt aaagacgatt caatagacaa
3301 taaggtctta acgcgttctg ataaaaatcg tggtaaatcg gataacgttc caagtgaaga
3361 agtagtcaaa aagatgaaaa actattggag acaacttcta aacgccaagt taatcactca
3421 acgtaagttt gataatttaa cgaaagctga acgtggaggt ttgagtgaac ttgataaagc
3481 tggttttatc aaacgccaat tggttgaaac tcgccaaatc actaagcatg tggcacaaat
3541 tttggatagt cgcatgaata ctaaatacga tgaaaatgat aaacttattc gagaggttaa
3601 agtgattacc ttaaaatcta aattagtttc tgacttccga aaagatttcc aattctataa
3661 agtacgtgag attaacaatt accatcatgc ccatgatgcg tatctaaatg ccgtcgttgg
3721 aactgctttg attaagaaat atccaaaact tgaatcggag tttgtctatg gtgattataa
3781 agtttatgat gttcgtaaaa tgattgctaa gtctgagcaa gaaataggca aagcaaccgc
3841 aaaatatttc ttttactcta atatcatgaa cttcttcaaa acagaaatta cacttgcaaa
3901 tggagagatt cgcaaacgcc ctctaatcga aactaatggg gaaactggag aaattgtctg
3961 ggataaaggg cgagattttg ccacagtgcg caaagtattg tccatgcccc aagtcaatat
4021 tgtcaagaaa acagaagtac agacaggcgg attctccaag gagtcaattt taccaaaaag
4081 aaattcggac aagcttattg ctcgtaaaaa agactgggat ccaaaaaaat atggtggttt
4141 tgatagtcca acggtagctt attcagtcct agtggttgct aaggtggaaa aagggaaatc
4201 gaagaagtta aaatccgtta aagagttact agggatcaca attatggaaa gaagttcctt
4261 tgaaaaaaat ccgattgact ttttagaagc taaaggatat aaggaagtta aaaaagactt
4321 aatcattaaa ctacctaaat atagtctttt tgagttagaa aacggtcgta aacggatgct
4381 ggctagtgcc ggagaattac aaaaaggaaa tgagctggct ctgccaagca aatatgtgaa
4441 ttttttatat ttagctagtc attatgaaaa gttgaagggt agtccagaag ataacgaaca
4501 aaaacaattg tttgtggagc agcataagca ttatttagat gagattattg agcaaatcag
4561 tgaattttct aagcgtgtta ttttagcaga tgccaattta gataaagttc ttagtgcata
4621 taacaaacat agagacaaac caatacgtga acaagcagaa aatattattc atttatttac
4681 gttgacgaat cttggagctc ccgctgcttt taaatatttt gatacaacaa ttgatcgtaa
4741 acgatatacg tctacaaaag aagttttaga tgccactctt atccatcaat ccatcactgg
4801 tctttatgaa acacgcattg atttgagtca gctaggaggt gactaactcg agtaaggatc
4861 tccaggcatc aaataaaacg aaaggctcag tcgaaagact gggcctttcg ttttatctgt
4921 tgtttgtcgg tgaacgctct ctactagagt cacactggct caccttcggg tgggcctttc
4981 tgcgtttata cctagggata tattccgctt cctcgctcac tgactcgcta cgctcggtcg
5041 ttcgactgcg gcgagcggaa atggcttacg aacggggcgg agatttcctg gaagatgcca
5101 ggaagatact taacagggaa gtgagagggc cgcggcaaag ccgtttttcc ataggctccg
5161 ccccccctgac aagcatcacg aaatctgacg ctcaaatcag tggtggcgaa acccgacagg
5221 actataaaga taccaggcgt ttccccctgg cggctccctc gtgcgctctc ctgttcctgc
5281 ctttcggttt accggtgtca ttccgctgtt atggccgcgt ttgtctcatt ccacgcctga
5341 cactcagttc cgggtaggca gttcgctcca agctggactg tatgcacgaa ccccccgttc
5401 agtccgaccg ctgcgcctta tccggtaact atcgtcttga gtccaacccg gaaagacatg
5461 caaaagcacc actggcagca gccactggta attgatttag aggagttagt cttgaagtca
5521 tgcgccggtt aaggctaaac tgaaaggaca agttttggtg actgcgctcc tccaagccag
5581 ttacctcggt tcaaagagtt ggtagctcag agaaccttcg aaaaaccgcc ctgcaaggcg
5641 gtttttcgt tttcagagca agagattacg cgcagaccaa aacgatctca agaagatcat
5701 cttattaatc agataaaata tttctagatt tcagtgcaat ttatctcttc aaatgtagca
5761 cctgaagtca gccccatacg atataagttg ttactagtgc ttggattctc accaataaaa
5821 aacgcccggc ggcaaccgag cgttctgaac aaatccagat ggagttctga ggtcattact
5881 ggatctatca acaggagtcc aagcgagctc gatatcaaat tacgccccgc cctgccactc
5941 atcgcagtac tgttgtaatt cattaagcat tctgccgaca tggaagccat cacaaacggc
6001 atgatgaacc tgaatcgcca gcggcatcag caccttgtcg ccttgcgtat aatatttgcc
6061 catggtgaaa acgggggcga agaagttgtc catattggcc acgtttaaat caaaactggt
6121 gaaactcacc cagggattgg ctgagacgaa aaacatattc tcaataaacc ctttagggaa
6181 ataggccagg ttttcaccgt aacacgccac atcttgcgaa tatatgtgta gaaactgccg
6241 gaaatcgtcg tggtattcac tccagagcga tgaaaacgtt tcagtttgct catggaaaac
6301 ggtgtaacaa gggtgaacac tatcccatat caccagctca ccgtctttca ttgccatacg
6361 aaattccgga tgagcattca tcaggcgggc aagaatgtga ataaaggccg gataaaactt
6421 gtgcttattt ttctttacgg tctttaaaaa ggccgtaata tccagctgaa cggtctggtt

```
     6481 ataggtacat tgagcaactg actgaaatgc ctcaaaatgt tctttacgat gccattggga
     6541 tatatcaacg gtggtatatc cagtgatttt tttctccatt ttagcttcct tagctcctga
     6601 aaatctcgat aactcaaaaa atacgcccgg tagtgatctt atttcattat ggtgaaagtt
     6661 ggaacctctt acgtgccgat caacgtctca ttttcgccag atatc
//
```

2. pTet_nCas9$_{H840A}$-bacteria/p15A

```
LOCUS       Exported                6705 bp ds-DNA     circular SYN 29-SEP-2018
DEFINITION  pTet_nCas9H840A/p15A; encodes for nuclease nickase Cas9H840A
       protein; CRC targeting module
ACCESSION   .
VERSION     .
KEYWORDS    pTet_nCas9H840A/p15A; CRC targeting module
SOURCE      synthetic DNA construct
  ORGANISM  synthetic DNA construct
REFERENCE   1  (bases 1 to 6705)
  AUTHORS   .
  TITLE     Direct Submission
  JOURNAL   Exported Sep 29, 2018 from SnapGene Viewer 4.2.5
            http://www.snapgene.com
FEATURES             Location/Qualifiers
     source          1..6705
                     /organism="synthetic DNA construct"
                     /mol_type="other DNA"
     CDS             complement(7..630)
                     /codon_start=1
                     /gene="tetR from transposon Tn10"
                     /product="tetracycline repressor TetR"
                     /label=TetR
                     /note="TetR binds to the tetracycline operator tetO to
                     inhibit transcription. This inhibition can be relieved by
                     adding tetracycline or doxycycline."

/translation="MSRLDKSKVINSALELLNEVGIEGLTTRKLAQKLGVEQPTLYWHV

KNKRALLDALAIEMLDRHHTHFCPLEGESWQDFLRNNAKSFRCALLSHRDGAKVHLGTR

PTEKQYETLENQLAFLCQQGFSLENALYALSAVGHFTLGCVLEDQEHQVAKEERETPTT
                     DSMPPLLRQAIELFDHQGAEPAFLFGLELIICGLEKQLKCESGS"
     promoter        649..704
                     /gene="tetR"
                     /label=tetR/tetA promoters
                     /note="overlapping promoters for bacterial tetR and tetA"
     protein_bind    685..703
                     /gene="tetO"
                     /label=tet operator
                     /bound_moiety="tetracycline repressor TetR"
                     /note="bacterial operator O2 for the tetR and tetA genes"
     RBS             722..733
                     /note="strong bacterial ribosome binding site (Elowitz and
                     Leibler, 2000)"
     CDS             740..4846
                     /codon_start=1
                     /product="nickase mutant of the Cas9 endonuclease from the
                     Streptococcus pyogenes Type II CRISPR/Cas system (H840A)"
                     /label=nCas9_H840A
                     /note="RNA-guided DNA-binding protein that lacks
                     endonuclease activity due to the D10 mutation in the
```

```
                     RuvC-like domain and the H840A mutation in the HNH domain"
/translation="MDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKK
NLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEES
FLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIK
FRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRL
ENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQ
IGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVR
QQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLL
RKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARG
NSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEY
FTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFD
SVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEER
LKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFM
QLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGR
HKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLY
LYYLQNGRDMYVDQELDINRLSDYDVDAIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVP
SEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKH
VAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYL
NAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKT
EITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSK
ESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGI
TIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNE
LALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILAD
ANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVL
                     DATLIHQSITGLYETRIDLSQLGGD"
     terminator      4870..4941
                     /label=rrnB T1 terminator
                     /note="transcription terminator T1 from the E. coli rrnB
                     gene"
     terminator      4957..4984
                     /label=T7Te terminator
                     /note="phage T7 early transcription terminator"
     rep_origin      complement(5146..5691)
                     /direction=LEFT
                     /label=p15A ori
                     /note="Plasmids containing the medium-copy-number p15A
                     origin of replication can be propagated in E. coli cells
                     that contain a second plasmid with the ColE1 origin."
     terminator      5805..5899
                     /label=lambda t0 terminator
```

```
                     /note="transcription terminator from phage lambda"
     CDS             complement(5920..6579)
                     /codon_start=1
                     /gene="cat"
                     /product="chloramphenicol acetyltransferase"
                     /label=CmR
                     /note="confers resistance to chloramphenicol"

/translation="MEKKITGYTTVDISQWHRKEHFEAFQSVAQCTYNQTVQLDITAFL

KTVKKNKHKFYPAFIHILARLMNAHPEFRMAMKDGELVIWDSVHPCYTVFHEQTETFSS

LWSEYHDDFRQFLHIYSQDVACYGENLAYFPKGFIENMFFVSANPWVSFTSFDLNVANM
                     DNFFAPVFTMGKYYTQGDKVLMPLAIQVHHAVCDGFHVGRMLNELQQYCDEWQGGA"
     promoter        complement(6580..6682)
                     /label=cat promoter
                     /note="promoter of the E. coli cat gene encoding
                     chloramphenicol acetyltransferase"
ORIGIN
        1 gacgtcttaa gacccacttt cacatttaag ttgtttttct aatccgcata tgatcaattc
       61 aaggccgaat aagaaggctg gctctgcacc ttggtgatca aataattcga tagcttgtcg
      121 taataatggc ggcatactat cagtagtagg tgtttccctt tcttctttag cgacttgatg
      181 ctcttgatct tccaatacgc aacctaaagt aaaatgcccc acagcgctga gtgcatataa
      241 tgcattctct agtgaaaaac cttgttggca taaaaaggct aattgatttt cgagagtttc
      301 atactgtttt tctgtaggcc gtgtacctaa atgtactttt gctccatcgc gatgacttag
      361 taaagcacat ctaaaacttt tagcgttatt acgtaaaaaa tcttgccagc tttccccttc
      421 taaagggcaa aagtgagtat ggtgcctatc taacatctca atggctaagg cgtcgagcaa
      481 agcccgctta tttttacat gccaatacaa tgtaggctgc tctacaccta gcttctgggc
      541 gagtttacgg gttgttaaac cttcgattcc gacctcatta agcagctcta atgcgctgtt
      601 aatcacttta cttttatcta atctagacat cattaattcc taatttttgt tgacactcta
      661 tcgttgatag agttatttta ccactcccta tcagtgatag agaaaagaat tcaaaagatc
      721 taaagaggag aaaggatcta tggataagaa atactcaata ggcttagata tcggcacaaa
      781 tagcgtcgga tgggcggtga tcactgatga atataaggtt ccgtctaaaa agttcaaggt
      841 tctgggaaat acagaccgcc acagtatcaa aaaaaatctt atagggggctc ttttatttga
      901 cagtggagag acagcggaag cgactcgtct caaacggaca gctcgtagaa ggtatacacg
      961 tcggaagaat cgtatttgtt atctacagga gatttttca aatgagatgg cgaaagtaga
     1021 tgatagtttc tttcatcgac ttgaagagtc ttttttggtg gaagaagaca agaagcatga
     1081 acgtcatcct atttttggaa atatagtaga tgaagttgct tatcatgaga aatatccaac
     1141 tatctatcat ctgcgaaaaa aattggtaga ttctactgat aaagcggatt tgcgcttaat
     1201 ctatttggcc ttagcgcata tgattaagtt tcgtggtcat tttttgattg agggagattt
     1261 aaatcctgat aatagtgatg tggacaaact atttatccag ttggtacaaa cctacaatca
     1321 attatttgaa gaaaacccta ttaacgcaag tggagtagat gctaaagcga ttctttctgc
     1381 acgattgagt aaatcaagac gattagaaaa tctcattgct cagctccccg gtgagaagaa
     1441 aaatggctta tttgggaatc tcattgcttt gtcattgggt ttgaccccta attttaaatc
     1501 aaattttgat ttggcagaag atgctaaatt acagctttca aaagatactt acgatgatga
     1561 tttagataat ttattggcgc aaattggaga tcaatatgct gatttgtttt tggcagctaa
     1621 gaatttatca gatgctattt tactttcaga tatcctaaga gtaaatactg aaataactaa
     1681 ggctccccta tcagcttcaa tgattaaacg ctacgatgaa catcatcaag acttgactct
     1741 tttaaaagct ttagttcgac aacaacttcc agaaaagtat aaagaaatct tttttgatca
     1801 atcaaaaaac ggatatgcag gttatattga tgggggagct agccaagaag aattttataa
     1861 atttatcaaa ccaattttag aaaaaatgga tggtactgag gaattattgg tgaaactaaa
     1921 tcgtgaagat ttgctgcgca agcaacggac ctttgacaac ggctctattc cccatcaaat
     1981 tcacttgggt gagctgcatg ctattttgag aagacaagaa gacttttatc catttttaaa
     2041 agacaatcgt gagaagattg aaaaaatctt gacttttcga attccttatt atgttggtcc
     2101 attggcgcgt ggcaatagtc gttttgcatg gatgactcgg aagtctgaag aaacaattac
     2161 cccatggaat tttgaagaag ttgtcgataa aggtgcttca gctcaatcat ttattgaacg
     2221 catgacaaac tttgataaaa atcttccaaa tgaaaaagta ctaccaaaac atagtttgct
     2281 ttatgagtat tttacggttt ataacgaatt gacaaaggtc aaatatgtta ctgaaggaat
     2341 gcgaaaacca gcatttcttt caggtgaaca gaagaaagcc attgttgatt tactcttcaa
     2401 aacaaatcga aaagtaaccg ttaagcaatt aaaagaagat tatttcaaaa aaatagaatg
     2461 ttttgatagt gttgaaattt caggagttga agatagattt aatgcttcat taggtaccta
     2521 ccatgatttg ctaaaaatta ttaaagataa agatttttg gataatgaag aaaatgaaga
     2581 tatcttagag gatattgttt taacattgac cttatttgaa gatagggaga tgattgagga
```

2641 aagacttaaa acatatgctc acctctttga tgataaggtg atgaaacagc ttaaacgtcg
2701 ccgttatact ggttggggac gtttgtctcg aaaattgatt aatggtatta gggataagca
2761 atctggcaaa acaatattag atttttttgaa atcagatggt tttgccaatc gcaattttat
2821 gcagctgatc catgatgata gtttgacatt taaagaagac attcaaaaag cacaagtgtc
2881 tggacaaggc gatagtttac atgaacatat tgcaaattta gctggtagcc ctgctattaa
2941 aaaaggtatt ttacagactg taaaagttgt tgatgaattg gtcaaagtaa tggggcggca
3001 taagccagaa aatatcgtta ttgaaatggc acgtgaaaat cagacaactc aaaagggcca
3061 gaaaaattcg cgagagcgta tgaaacgaat cgaagaaggt atcaaagaat taggaagtca
3121 gattcttaaa gagcatcctg ttgaaaatac tcaattgcaa aatgaaaagc tctatctcta
3181 ttatctccaa aatggaagag acatgtatgt ggaccaagaa ttagatatta atcgtttaag
3241 tgattatgat gtcgatgcca ttgttccaca aagtttcctt aaagacgatt caatagacaa
3301 taaggtctta acgcgttctg ataaaaatcg tggtaaatcg gataacgttc caagtgaaga
3361 agtagtcaaa aagatgaaaa actattggag acaacttcta aacgccaagt taatcactca
3421 acgtaagttt gataatttaa cgaaagctga acgtggaggt ttgagtgaac ttgataaagc
3481 tggttttatc aaacgccaat tggttgaaac tcgccaaatc actaagcatg tggcacaaat
3541 tttggatagt cgcatgaata ctaaatacga tgaaaatgat aaacttattc gagaggttaa
3601 agtgattacc ttaaaatcta aattagtttc tgacttccga aaagatttcc aattctataa
3661 agtacgtgag attaacaatt accatcatgc ccatgatgcg tatctaaatg ccgtcgttgg
3721 aactgctttg attaagaaat atccaaaact tgaatcggag tttgtctatg gtgattataa
3781 agtttatgat gttcgtaaaa tgattgctaa gtctgagcaa gaaataggca aagcaaccgc
3841 aaaatatttc ttttactcta atatcatgaa cttcttcaaa acagaaatta cacttgcaaa
3901 tggagagatt cgcaaacgcc ctctaatcga aactaatggg gaaactggag aaattgtctg
3961 ggataaaggg cgagattttg ccacagtgcg caaagtattg tccatgcccc aagtcaatat
4021 tgtcaagaaa acagaagtac agacaggcgg attctccaag gagtcaattt taccaaaaag
4081 aaattcggac aagcttattg ctcgtaaaaa agactgggat ccaaaaaaat atggtggttt
4141 tgatagtcca acggtagctt attcagtcct agtggttgct aaggtggaaa aagggaaatc
4201 gaagaagtta aaatccgtta aagagttact agggatcaca attatggaaa gaagttcctt
4261 tgaaaaaaat ccgattgact ttttagaagc taaaggatat aaggaagtta aaaaagactt
4321 aatcattaaa ctacctaaat atagtctttt tgagttagaa aacggtcgta aacggatgct
4381 ggctagtgcc ggagaattac aaaaaggaaa tgagctggct ctgccaagca aatatgtgaa
4441 ttttttatat ttagctagtc attatgaaaa gttgaagggt agtccagaag ataacgaaca
4501 aaaacaattg tttgtggagc agcataagca ttatttagat gagattattg agcaaatcag
4561 tgaattttct aagcgtgtta ttttagcaga tgccaattta gataaagttc ttagtgcata
4621 taacaaacat agagacaaac caatacgtga acaagcagaa aatattattc atttatttac
4681 gttgacgaat cttggagctc ccgctgcttt taaatatttt gatacaacaa ttgatcgtaa
4741 acgatatacg tctacaaaag aagttttaga tgccactctt atccatcaat ccatcactgg
4801 tctttatgaa acacgcattg atttgagtca gctaggaggt gactaactcg agtaaggatc
4861 tccaggcatc aaataaaacg aaaggctcag tcgaaagact gggcctttcg ttttatctgt
4921 tgtttgtcgg tgaacgctct ctactagagt cacactggct caccttcggg tgggcctttc
4981 tgcgtttata cctagggata tattccgctt cctcgctcac tgactcgcta cgctcggtcg
5041 ttcgactgcg gcgagcggaa atggcttacg aacggggcgg agatttcctg gaagatgcca
5101 ggaagatact taacagggaa gtgagagggc cgcggcaaag ccgtttttcc ataggctccg
5161 cccccctgac aagcatcacg aaatctgacg ctcaaatcag tggtggcgaa acccgacagg
5221 actataaaga taccaggcgt ttccccctgg cggctccctc gtgcgctctc ctgttcctgc
5281 ctttcggttt accggtgtca ttccgctgtt atggccgcgt ttgtctcatt ccacgcctga
5341 cactcagttc cgggtaggca gttcgctcca agctggactg tatgcacgaa ccccccgttc
5401 agtccgaccg ctgcgcctta tccggtaact atcgtcttga gtccaacccg gaaagacatg
5461 caaaagcacc actggcagca gccactggta attgatttag aggagttagt cttgaagtca
5521 tgcgccggtt aaggctaaac tgaaaggaca agttttggtg actgcgctcc tccaagccag
5581 ttacctcggt tcaaagagtt ggtagctcag agaaccttcg aaaaaccgcc ctgcaaggcg
5641 gtttttcgt tttcagagca agagattacg cgcagaccaa aacgatctca agaagatcat
5701 cttattaatc agataaaata tttctagatt tcagtgcaat ttatctcttc aaatgtagca
5761 cctgaagtca gccccatacg atataagttg ttactagtgc ttggattctc accaataaaa
5821 aacgcccggc ggcaaccgag cgttctgaac aaatccagat ggagttctga ggtcattact
5881 ggatctatca acaggagtcc aagcgagctc gatatcaaat tacgccccgc cctgccactc
5941 atcgcagtac tgttgtaatt cattaagcat tctgccgaca tggaagccat cacaaacggc
6001 atgatgaacc tgaatcgcca gcggcatcag caccttgtcg ccttgcgtat aatatttgcc
6061 catggtgaaa acgggggcga agaagttgtc catattggcc acgtttaaat caaaactggt
6121 gaaactcacc cagggattgg ctgagacgaa aaacatattc tcaataaacc ctttagggaa
6181 ataggccagg ttttcaccgt aacacgccac atcttgcgaa tatatgtgta gaaactgccg
6241 gaaatcgtcg tggtattcac tccagagcga tgaaaacgtt tcagtttgct catggaaaac
6301 ggtgtaacaa gggtgaacac tatcccatat caccagctca ccgtctttca ttgccatacg
6361 aaattccgga tgagcattca tcaggcgggc aagaatgtga ataaaggccg gataaaactt

```
     6421 gtgcttattt ttctttacgg tctttaaaaa ggccgtaata tccagctgaa cggtctggtt
     6481 ataggtacat tgagcaactg actgaaatgc ctcaaaatgt tctttacgat gccattggga
     6541 tatatcaacg gtggtatatc cagtgatttt tttctccatt ttagcttcct tagctcctga
     6601 aaatctcgat aactcaaaaa atacgcccgg tagtgatctt atttcattat ggtgaaagtt
     6661 ggaacctctt acgtgccgat caacgtctca ttttcgccag atatc
//
```

3. pTet_nCas9D10A-bacteria/p15A

```
LOCUS       Exported                6705 bp ds-DNA     circular SYN 29-SEP-2018
DEFINITION  synthetic circular DNA
ACCESSION   .
VERSION     .
KEYWORDS    .
SOURCE      synthetic DNA construct
  ORGANISM  synthetic DNA construct
REFERENCE   1  (bases 1 to 6705)
  AUTHORS   .
  TITLE     Direct Submission
  JOURNAL   Exported Sep 29, 2018 from SnapGene Viewer 4.2.5
            http://www.snapgene.com
FEATURES             Location/Qualifiers
     source          1..6705
                     /organism="synthetic DNA construct"
                     /mol_type="other DNA"
     CDS             complement(7..630)
                     /codon_start=1
                     /gene="tetR from transposon Tn10"
                     /product="tetracycline repressor TetR"
                     /label=TetR
                     /note="TetR binds to the tetracycline operator tetO to
                     inhibit transcription. This inhibition can be relieved by
                     adding tetracycline or doxycycline."

/translation="MSRLDKSKVINSALELLNEVGIEGLTTRKLAQKLGVEQPTLYWHV

KNKRALLDALAIEMLDRHHTHFCPLEGESWQDFLRNNAKSFRCALLSHRDGAKVHLGTR

PTEKQYETLENQLAFLCQQGFSLENALYALSAVGHFTLGCVLEDQEHQVAKEERETPTT
                     DSMPPLLRQAIELFDHQGAEPAFLFGLELIICGLEKQLKCESGS"
     promoter        649..704
                     /gene="tetR"
                     /label=tetR/tetA promoters
                     /note="overlapping promoters for bacterial tetR and tetA"
     protein_bind    685..703
                     /gene="tetO"
                     /label=tet operator
                     /bound_moiety="tetracycline repressor TetR"
                     /note="bacterial operator O2 for the tetR and tetA genes"
     RBS             722..733
                     /note="strong bacterial ribosome binding site (Elowitz and
                     Leibler, 2000)"
     CDS             740..4846
                     /codon_start=1
                     /product="nickase mutant of the Cas9 endonuclease from the
                     Streptococcus pyogenes Type II CRISPR/Cas system (D10A)"
                     /label=nCas9_D10A
                     /note="generates RNA-guided single strand nicks in DNA due
```

```
                     to the D10 mutation in the RuvC-like catalytic domain"
/translation="MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKK
NLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEES
FLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIK
FRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRL
ENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQ
IGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVR
QQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLL
RKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARG
NSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEY
FTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFD
SVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEER
LKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFM
QLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGR
HKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLY
LYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVP
SEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKH
VAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYL
NAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKT
EITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSK
ESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGI
TIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNE
LALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILAD
ANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVL
                     DATLIHQSITGLYETRIDLSQLGGD"
     terminator      4870..4941
                     /label=rrnB T1 terminator
                     /note="transcription terminator T1 from the E. coli rrnB
                     gene"
     terminator      4957..4984
                     /label=T7Te terminator
                     /note="phage T7 early transcription terminator"
     rep_origin      complement(5146..5691)
                     /direction=LEFT
                     /label=p15A ori
                     /note="Plasmids containing the medium-copy-number p15A
                     origin of replication can be propagated in E. coli cells
                     that contain a second plasmid with the ColE1 origin."
     terminator      5805..5899
                     /label=lambda t0 terminator
```

```
                     /note="transcription terminator from phage lambda"
     CDS             complement(5920..6579)
                     /codon_start=1
                     /gene="cat"
                     /product="chloramphenicol acetyltransferase"
                     /label=CmR
                     /note="confers resistance to chloramphenicol"

/translation="MEKKITGYTTVDISQWHRKEHFEAFQSVAQCTYNQTVQLDITAFL

KTVKKNKHKFYPAFIHILARLMNAHPEFRMAMKDGELVIWDSVHPCYTVFHEQTETFSS

LWSEYHDDFRQFLHIYSQDVACYGENLAYFPKGFIENMFFVSANPWVSFTSFDLNVANM
                     DNFFAPVFTMGKYYTQGDKVLMPLAIQVHHAVCDGFHVGRMLNELQQYCDEWQGGA"
     promoter        complement(6580..6682)
                     /label=cat promoter
                     /note="promoter of the E. coli cat gene encoding
                     chloramphenicol acetyltransferase"
ORIGIN
        1 gacgtcttaa gacccacttt cacatttaag ttgtttttct aatccgcata tgatcaattc
       61 aaggccgaat aagaaggctg gctctgcacc ttggtgatca aataattcga tagcttgtcg
      121 taataatggc ggcatactat cagtagtagg tgtttccctt tcttctttag cgacttgatg
      181 ctcttgatct tccaatacgc aacctaaagt aaaatgcccc acagcgctga gtgcatataa
      241 tgcattctct agtgaaaaac cttgttggca taaaaaggct aattgatttt cgagagtttc
      301 atactgtttt tctgtaggcc gtgtacctaa atgtactttt gctccatcgc gatgacttag
      361 taaagcacat ctaaaacttt tagcgttatt acgtaaaaaa tcttgccagc tttccccttc
      421 taaagggcaa aagtgagtat ggtgcctatc taacatctca atggctaagg cgtcgagcaa
      481 agcccgctta ttttttacat gccaatacaa tgtaggctgc tctacaccta gcttctgggc
      541 gagtttacgg gttgttaaac cttcgattcc gacctcatta agcagctcta atgcgctgtt
      601 aatcacttta cttttatcta atctagacat cattaattcc taatttttgt tgacactcta
      661 tcgttgatag agttatttta ccactcccta tcagtgatag agaaaagaat tcaaaagatc
      721 taaagaggag aaaggatcta tggataagaa atactcaata ggcttagcta tcggcacaaa
      781 tagcgtcgga tgggcggtga tcactgatga atataaggtt ccgtctaaaa agttcaaggt
      841 tctgggaaat acagaccgcc acagtatcaa aaaaaatctt atagggggctc ttttatttga
      901 cagtggagag acagcggaag cgactcgtct caaacggaca gctcgtagaa ggtatacacg
      961 tcggaagaat cgtatttgtt atctacagga gatttttca aatgagatgg cgaaagtaga
     1021 tgatagtttc tttcatcgac ttgaagagtc ttttttggtg gaagaagaca agaagcatga
     1081 acgtcatcct atttttggaa atatagtaga tgaagttgct tatcatgaga aatatccaac
     1141 tatctatcat ctgcgaaaaa aattggtaga ttctactgat aaagcggatt tgcgcttaat
     1201 ctatttggcc ttagcgcata tgattaagtt tcgtggtcat tttttgattg agggagattt
     1261 aaatcctgat aatagtgatg tggacaaact atttatccag ttggtacaaa cctacaatca
     1321 attatttgaa gaaaacccta ttaacgcaag tggagtagat gctaaagcga ttctttctgc
     1381 acgattgagt aaatcaagac gattagaaaa tctcattgct cagctccccg gtgagaagaa
     1441 aaatggctta tttgggaatc tcattgcttt gtcattgggt ttgaccccta attttaaatc
     1501 aaattttgat ttggcagaag atgctaaatt acagctttca aaagatactt acgatgatga
     1561 tttagataat ttattggcgc aaattggaga tcaatatgct gatttgtttt tggcagctaa
     1621 gaatttatca gatgctattt tactttcaga tatcctaaga gtaaatactg aaataactaa
     1681 ggctccccta tcagcttcaa tgattaaacg ctacgatgaa catcatcaag acttgactct
     1741 tttaaaagct ttagttcgac aacaacttcc agaaaagtat aaagaaatct tttttgatca
     1801 atcaaaaaac ggatatgcag gttatattga tgggggagct agccaagaag aattttataa
     1861 atttatcaaa ccaattttag aaaaaatgga tggtactgag gaattattgg tgaaactaaa
     1921 tcgtgaagat ttgctgcgca agcaacggac ctttgacaac ggctctattc cccatcaaat
     1981 tcacttgggt gagctgcatg ctattttgag aagacaagaa gacttttatc catttttaaa
     2041 agacaatcgt gagaagattg aaaaaatctt gacttttcga attccttatt atgttggtcc
     2101 attggcgcgt ggcaatagtc gttttgcatg gatgactcgg aagtctgaag aaacaattac
     2161 cccatggaat tttgaagaag ttgtcgataa aggtgcttca gctcaatcat ttattgaacg
     2221 catgacaaac tttgataaaa atcttccaaa tgaaaaagta ctaccaaaac atagtttgct
     2281 ttatgagtat tttacggttt ataacgaatt gacaaaggtc aaatatgtta ctgaaggaat
     2341 gcgaaaacca gcatttcttt caggtgaaca gaagaaagcc attgttgatt tactcttcaa
     2401 aacaaatcga aaagtaaccg ttaagcaatt aaaagaagat tatttcaaaa aaatagaatg
     2461 ttttgatagt gttgaaattt caggagttga agatagattt aatgcttcat taggtaccta
     2521 ccatgatttg ctaaaaatta ttaaagataa agatttttg gataatgaag aaaatgaaga
     2581 tatcttagag gatattgttt taacattgac cttatttgaa gatagggaga tgattgagga
```

2641 aagacttaaa acatatgctc acctctttga tgataaggtg atgaaacagc ttaaacgtcg
2701 ccgttatact ggttggggac gtttgtctcg aaaattgatt aatggtatta gggataagca
2761 atctggcaaa acaatattag atttttttgaa atcagatggt tttgccaatc gcaattttat
2821 gcagctgatc catgatgata gtttgacatt taaagaagac attcaaaaag cacaagtgtc
2881 tggacaaggc gatagtttac atgaacatat tgcaaattta gctggtagcc ctgctattaa
2941 aaaaggtatt ttacagactg taaaagttgt tgatgaattg gtcaaagtaa tggggcggca
3001 taagccagaa aatatcgtta ttgaaatggc acgtgaaaat cagacaactc aaaagggcca
3061 gaaaaattcg cgagagcgta tgaaacgaat cgaagaaggt atcaaagaat taggaagtca
3121 gattcttaaa gagcatcctg ttgaaaatac tcaattgcaa aatgaaaagc tctatctcta
3181 ttatctccaa aatggaagag acatgtatgt ggaccaagaa ttagatatta atcgtttaag
3241 tgattatgat gtcgatcaca ttgttccaca aagtttcctt aaagacgatt caatagacaa
3301 taaggtctta acgcgttctg ataaaaatcg tggtaaatcg gataacgttc caagtgaaga
3361 agtagtcaaa aagatgaaaa actattggag acaacttcta aacgccaagt taatcactca
3421 acgtaagttt gataatttaa cgaaagctga acgtggaggt ttgagtgaac ttgataaagc
3481 tggttttatc aaacgccaat tggttgaaac tcgccaaatc actaagcatg tggcacaaat
3541 tttggatagt cgcatgaata ctaaatacga tgaaaatgat aaacttattc gagaggttaa
3601 agtgattacc ttaaaatcta aattagtttc tgacttccga aaagatttcc aattctataa
3661 agtacgtgag attaacaatt accatcatgc ccatgatgcg tatctaaatg ccgtcgttgg
3721 aactgctttg attaagaaat atccaaaact tgaatcggag tttgtctatg gtgattataa
3781 agtttatgat gttcgtaaaa tgattgctaa gtctgagcaa gaaataggca aagcaaccgc
3841 aaaatatttc ttttactcta atatcatgaa cttcttcaaa acagaaatta cacttgcaaa
3901 tggagagatt cgcaaacgcc ctctaatcga aactaatggg gaaactggag aaattgtctg
3961 ggataaaggg cgagattttg ccacagtgcg caaagtattg tccatgcccc aagtcaatat
4021 tgtcaagaaa acagaagtac agacaggcgg attctccaag gagtcaattt taccaaaaag
4081 aaattcggac aagcttattg ctcgtaaaaa agactgggat ccaaaaaaat atggtggttt
4141 tgatagtcca acggtagctt attcagtcct agtggttgct aaggtggaaa aagggaaatc
4201 gaagaagtta aaatccgtta aagagttact agggatcaca attatggaaa gaagttcctt
4261 tgaaaaaaat ccgattgact ttttagaagc taaaggatat aaggaagtta aaaaagactt
4321 aatcattaaa ctacctaaat atagtctttt tgagttagaa aacggtcgta aacggatgct
4381 ggctagtgcc ggagaattac aaaaaggaaa tgagctggct ctgccaagca aatatgtgaa
4441 ttttttatat ttagctagtc attatgaaaa gttgaagggt agtccagaag ataacgaaca
4501 aaaacaattg tttgtggagc agcataagca ttatttagat gagattattg agcaaatcag
4561 tgaattttct aagcgtgtta ttttagcaga tgccaattta gataaagttc ttagtgcata
4621 taacaaacat agagacaaac caatacgtga acaagcagaa aatattattc atttatttac
4681 gttgacgaat cttggagctc ccgctgcttt taaatatttt gatacaacaa ttgatcgtaa
4741 acgatatacg tctacaaaag aagttttaga tgccactctt atccatcaat ccatcactgg
4801 tctttatgaa acacgcattg atttgagtca gctaggaggt gactaactcg agtaaggatc
4861 tccaggcatc aaataaaacg aaaggctcag tcgaaagact gggcctttcg ttttatctgt
4921 tgtttgtcgg tgaacgctct ctactagagt cacactggct caccttcggg tgggcctttc
4981 tgcgtttata cctagggata tattccgctt cctcgctcac tgactcgcta cgctcggtcg
5041 ttcgactgcg gcgagcggaa atggcttacg aacggggcgg agatttcctg gaagatgcca
5101 ggaagatact taacagggaa gtgagagggc cgcggcaaag ccgtttttcc ataggctccg
5161 cccccctgac aagcatcacg aaatctgacg ctcaaatcag tggtggcgaa acccgacagg
5221 actataaaga taccaggcgt ttccccctgg cggctccctc gtgcgctctc ctgttcctgc
5281 ctttcggttt accggtgtca ttccgctgtt atggccgcgt ttgtctcatt ccacgcctga
5341 cactcagttc cgggtaggca gttcgctcca agctggactg tatgcacgaa ccccccgttc
5401 agtccgaccg ctgcgcctta tccggtaact atcgtcttga gtccaacccg gaaagacatg
5461 caaaagcacc actggcagca gccactggta attgatttag aggagttagt cttgaagtca
5521 tgcgccggtt aaggctaaac tgaaaggaca agttttggtg actgcgctcc tccaagccag
5581 ttacctcggt tcaaagagtt ggtagctcag agaaccttcg aaaaaccgcc ctgcaaggcg
5641 gtttttcgt tttcagagca agagattacg cgcagaccaa aacgatctca agaagatcat
5701 cttattaatc agataaaata tttctagatt tcagtgcaat ttatctcttc aaatgtagca
5761 cctgaagtca gccccatacg atataagttg ttactagtgc ttggattctc accaataaaa
5821 aacgcccggc ggcaaccgag cgttctgaac aaatccagat ggagttctga ggtcattact
5881 ggatctatca acaggagtcc aagcgagctc gatatcaaat tacgccccgc cctgccactc
5941 atcgcagtac tgttgtaatt cattaagcat tctgccgaca tggaagccat cacaaacggc
6001 atgatgaacc tgaatcgcca gcggcatcag caccttgtcg ccttgcgtat aatatttgcc
6061 catggtgaaa acgggggcga agaagttgtc catattggcc acgtttaaat caaaactggt
6121 gaaactcacc cagggattgg ctgagacgaa aaacatattc tcaataaacc ctttagggaa
6181 ataggccagg ttttcaccgt aacacgccac atcttgcgaa tatatgtgta gaaactgccg
6241 gaaatcgtcg tggtattcac tccagagcga tgaaaacgtt tcagtttgct catggaaaac
6301 ggtgtaacaa gggtgaacac tatcccatat caccagctca ccgtctttca ttgccatacg
6361 aaattccgga tgagcattca tcaggcgggc aagaatgtga ataaaggccg gataaaactt

```
     6421 gtgcttattt ttctttacgg tctttaaaaa ggccgtaata tccagctgaa cggtctggtt
     6481 ataggtacat tgagcaactg actgaaatgc ctcaaaatgt tctttacgat gccattggga
     6541 tatatcaacg gtggtatatc cagtgatttt tttctccatt ttagcttcct tagctcctga
     6601 aaatctcgat aactcaaaaa atacgcccgg tagtgatctt atttcattat ggtgaaagtt
     6661 ggaacctctt acgtgccgat caacgtctca ttttcgccag atatc
//
```

4. pTet_AID_MCP/DF13

```
LOCUS       Exported                5533 bp ds-DNA     circular SYN 29-SEP-2018
DEFINITION  pTet_AID_MCP/DF13, encodes for AID_MCP fusion protein, CRC Effector
            Module
ACCESSION   .
VERSION     .
KEYWORDS    .pTet_AID_MCP/DF13; CRC Effector Module
SOURCE      synthetic DNA construct
  ORGANISM  synthetic DNA construct
REFERENCE   1  (bases 1 to 5533)
  AUTHORS   .Collantes, JC
  TITLE     Direct Submission
  JOURNAL   Exported Sep 29, 2018 from SnapGene Viewer 4.2.5
            http://www.snapgene.com
FEATURES             Location/Qualifiers
     source          1..5533
                     /organism="synthetic DNA construct"
                     /mol_type="other DNA"
     source          2732..2754
                     /organism="Enterobacteria phage T7"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10760"
     promoter        complement(558..635)
                     /gene="lacI"
                     /label=lacI promoter
     CDS             complement(737..1360)
                     /codon_start=1
                     /gene="tetR from transposon Tn10"
                     /product="tetracycline repressor TetR"
                     /label=TetR
                     /note="TetR binds to the tetracycline operator tetO to
                     inhibit transcription. This inhibition can be relieved by
                     adding tetracycline or doxycycline."

/translation="MSRLDKSKVINSALELLNEVGIEGLTTRKLAQKLGVEQPTLYWHV

KNKRALLDALAIEMLDRHHTHFCPLEGESWQDFLRNNAKSFRCALLSHRDGAKVHLGTR

PTEKQYETLENQLAFLCQQGFSLENALYALSAVGHFTLGCVLEDQEHQVAKEERETPTT
                     DSMPPLLRQAIELFDHQGAEPAFLFGLELIICGLEKQLKCESGS"
     promoter        1379..1434
                     /gene="tetR"
                     /label=tetR/tetA promoters
                     /note="overlapping promoters for bacterial tetR and tetA"
     protein_bind    1415..1433
                     /gene="tetO"
                     /label=tet operator
                     /bound_moiety="tetracycline repressor TetR"
                     /note="bacterial operator O2 for the tetR and tetA genes"
     RBS             1452..1463
                     /note="strong bacterial ribosome binding site (Elowitz and
```

```
                     Leibler, 2000)"
     CDS             1470..2063
                     /codon_start=1
                     /gene="AICDA"
                     /product="human activation-induced cytidine deaminase
                     (Dickerson et al., 2003)"
                     /label=AID

/translation="MDSLLMNRRKFLYQFKNVRWAKGRRETYLCYVVKRRDSATSFSLD

FGYLRNKNGCHVELLFLRYISDWDLDPGRCYRVTWFTSWSPCYDCARHVADFLRGNPNL

SLRIFTARLYFCEDRKAEPEGLRRLHRAGVQIAIMTFKDYFYCWNTFVENHERTFKAWE
                     GLHENSVRLSRQLRRILLPLYEVDDLRDAFRTLGL"
     CDS             2064..2138
                     /codon_start=1
                     /label=L25
                     /translation="ELKTPLGDTTHTSPPCPAPELLGGP"
     CDS             2139..2495
                     /codon_start=1
                     /product="MCP"
                     /label=MCP

/translation="MASNFTQFVLVDNGGTGDVTVAPSNFANGIAEWISSNSRSQAYKV

TCSVRQSSAQNRKYTIKVEVPKGAWRSYLNMELTIPIFATNSDCELIVKAMQGLLKDGN
                     PIPSAIAANSGIY*"
     terminator      2519..2590
                     /label=rrnB T1 terminator
                     /note="transcription terminator T1 from the E. coli rrnB
                     gene"
     terminator      2606..2633
                     /label=T7Te terminator
                     /note="phage T7 early transcription terminator"
     promoter        2674..2692
                     /label=T7 promoter
                     /note="promoter for bacteriophage T7 RNA polymerase"
     protein_bind    2693..2717
                     /label=lac operator
                     /bound_moiety="lac repressor encoded by lacI"
                     /note="The lac repressor binds to the lac operator to
                     inhibit transcription in E. coli. This inhibition can be
                     relieved by adding lactose or
                     isopropyl-beta-D-thiogalactopyranoside (IPTG)."
     RBS             2732..2754
                     /note="efficient ribosome binding site from bacteriophage
                     T7 gene 10 (Olins and Rangwala, 1989)"
     CDS             2828..2872
                     /codon_start=1
                     /product="affinity and epitope tag derived from pancreatic
                     ribonuclease A"
                     /label=S-Tag
                     /translation="KETAAAKFERQHMDS"
     terminator      2925..2972
                     /label=T7 terminator
                     /note="transcription terminator for bacteriophage T7 RNA
                     polymerase"
     CDS             complement(3129..3920)
                     /codon_start=1
                     /gene="aadA"
                     /product="aminoglycoside adenylyltransferase (Murphy,
                     1985)"
                     /label=SmR
```

```
                     /note="confers resistance to spectinomycin and
                     streptomycin"

/translation="MREAVIAEVSTQLSEVVGVIERHLEPTLLAVHLYGSAVDGGLKPH

SDIDLLVTVTVRLDETTRRALINDLLETSASPGESEILRAVEVTIVVHDDIIPWRYPAK

RELQFGEWQRNDILAGIFEPATIDIDLAILLTKAREHSVALVGPAAEELFDPVPEQDLF

EALNETLTLWNSPPDWAGDERNVVLTLSRIWYSAVTGKIAPKDVAADWAMERLPAQYQP
                     VILEARQAYLGQEEDRLASRADQLEEFVHYVKGEITKVVGK"
     promoter        complement(3921..4012)
                     /gene="bla"
                     /label=AmpR promoter
     rep_origin      complement(4060..4798)
                     /direction=LEFT
                     /label=CloDF13 ori
                     /note="Plasmids containing the CloDF13 (CDF) origin of
                     replication can be propagated in E. coli cells that
contain
                     additional plasmids with compatible origins."
     CDS             complement(join(5008..5533,1..557))
                     /codon_start=1
                     /gene="lacI"
                     /product="lac repressor"
                     /label=lacI
                     /note="The lac repressor binds to the lac operator to
                     inhibit transcription in E. coli. This inhibition can be
                     relieved by adding lactose or
                     isopropyl-beta-D-thiogalactopyranoside (IPTG)."

/translation="MKPVTLYDVAEYAGVSYQTVSRVVNQASHVSAKTREKVEAAMAEL

NYIPNRVAQQLAGKQSLLIGVATSSLALHAPSQIVAAIKSRADQLGASVVVSMVERSGV

EACKAAVHNLLAQRVSGLIINYPLDDQDAIAVEAACTNVPALFLDVSDQTPINSIIFSH

EDGTRLGVEHLVALGHQQIALLAGPLSSVSARLRLAGWHKYLTRNQIQPIAEREGDWSA

MSGFQQTMQMLNEGIVPTAMLVANDQMALGAMRAITESGLRVGADISVVGYDDTEDSSC

YIPPLTTIKQDFRLLGQTSVDRLLQLSQGQAVKGNQLLPVSLVKRKTTLAPNTQTASPR
                     ALADSLMQLARQVSRLESGQ"
ORIGIN
        1 gctaacagcg cgatttgctg gtgacccaat gcgaccagat gctccacgcc cagtcgcgta
       61 ccgtcttcat gggagaaaat aatactgttg atgggtgtct ggtcagagac atcaagaaat
      121 aacgccggaa cattagtgca ggcagcttcc acagcaatgg catcctggtc atccagcgga
      181 tagttaatga tcagcccact gacgcgttgc gcgagaagat tgtgcaccgc cgctttacag
      241 gcttcgacgc cgcttcgttc taccatcgac accaccacgc tggcacccag ttgatcggcg
      301 cgagatttaa tcgccgcgac aatttgcgac ggcgcgtgca gggccagact ggaggtggca
      361 acgccaatca gcaacgactg tttgcccgcc agttgttgtg ccacgcggtt gggaatgtaa
      421 ttcagctccg ccatcgccgc ttccactttt tcccgcgttt tcgcagaaac gtggctggcc
      481 tggttcacca cgcgggaaac ggtctgataa gagacaccgg catactctgc gacatcgtat
      541 aacgttactg gtttcacatt caccaccctg aattgactct cttccgggcg ctatcatgcc
      601 ataccgcgaa aggttttgcg ccattcgatg gtgtccggga tctcgacgct ctcccttatg
      661 cgactcctgc attaggaagc agcccagtag taggttgagg ccgttgagca ccgccgccgc
      721 aaggaatggt gcatgcttaa gacccacttt cacatttaag ttgtttttct aatccgcata
      781 tgatcaattc aaggccgaat aagaaggctg gctctgcacc ttggtgatca aataattcga
      841 tagcttgtcg taataatggc ggcatactat cagtagtagg tgtttccctt tcttctttag
      901 cgacttgatg ctcttgatct tccaatacgc aacctaaagt aaaatgcccc acagcgctga
      961 gtgcatataa tgcattctct agtgaaaaac cttgttggca taaaaaggct aattgatttt
     1021 cgagagtttc atactgtttt tctgtaggcc gtgtacctaa atgtactttt gctccatcgc
     1081 gatgacttag taaagcacat ctaaaacttt tagcgttatt acgtaaaaaa tcttgccagc
```

1141 tttccccttc taaagggcaa aagtgagtat ggtgcctatc taacatctca atggctaagg
1201 cgtcgagcaa agcccgctta ttttttacat gccaatacaa tgtaggctgc tctacaccta
1261 gcttctgggc gagtttacgg gttgttaaac cttcgattcc gacctcatta agcagctcta
1321 atgcgctgtt aatcacttta cttttatcta atctagacat cattaattcc taatttttgt
1381 tgacactcta tcgttgatag agttatttta ccactcccta tcagtgatag agaaaagaat
1441 tcaaaagatc taaagaggag aaaggatcta tggacagcct cttgatgaac cggaggaagt
1501 ttctttacca attcaaaaat gtccgctggg ctaagggtcg gcgtgagacc tacctgtgct
1561 acgtagtgaa gaggcgtgac agtgctacat ccttttcact ggactttggt tatcttcgca
1621 ataagaacgg ctgccacgtg gaattgctct tcctccgcta catctcggac tgggacctag
1681 accctggccg ctgctaccgc gtcacctggt tcacctcctg gagcccctgc tacgactgtg
1741 cccgacatgt ggccgacttt ctgcgaggga accccaacct cagtctgagg atcttcaccg
1801 cgcgcctcta cttctgtgag gaccgcaagg ctgagcccga ggggctgcgg cggctgcacc
1861 gcgccggggt gcaaatagcc atcatgacct tcaaagatta tttttactgc tggaatactt
1921 ttgtagaaaa ccatgaaaga actttcaaag cctgggaagg gctgcatgaa aattcagttc
1981 gtctctccag acagcttcgg cgcatccttt tgcccctgta tgaggttgat gacttacgag
2041 acgcatttcg tactttggga cttgaactga aaactccact tggagatact acacatacct
2101 caccaccttg cccagctcct gaactgctgg gtggacctat ggcttctaac tttactcagt
2161 tcgttctcgt cgacaatggc ggaactggcg acgtgactgt cgccccaagc aacttcgcta
2221 acgggatcgc tgaatggatc agctctaact cgcgttcaca ggcttacaaa gtaacctgta
2281 gcgttcgtca gagctctgcg cagaatcgca aatacaccat caaagtcgag gtgcctaaag
2341 gcgcctggcg ttcgtactta aatatggaac taaccattcc aattttcgcc acgaattccg
2401 actgcgagct tattgttaag gcaatgcaag gtctcctaaa agatggaaac ccgattccct
2461 cagcaatcgc agcaaactcc ggcatctact gatgactcga gtaaggatct ccaggcatca
2521 aataaaacga aaggctcagt cgaaagactg ggcctttcgt tttatctgtt gtttgtcggt
2581 gaacgctctc tactagagtc acactggctc accttcgggt gggcctttct gcgtttatac
2641 ctaggatcga gatcgatctc gatcccgcga aattaatacg actcactata ggggaattgt
2701 gagcggataa caattcccct ctagaaataa ttttgtttaa ctttaagaag gagatataca
2761 tatggcagat ctcaattgga tatcggccgg ccacgcgatc gctgacgtcg gtaccctcga
2821 gtctggtaaa gaaaccgctg ctgcgaaatt tgaacgccag cacatggact cgtctactag
2881 tcgcagctta attaacctaa actgctgcca ccgctgagca ataactagca taaccccttg
2941 gggcctctaa acgggtcttg aggggttttt tgctagcgaa aggaggagtc gacactgctt
3001 ccggtagtca ataaaccggt aaaccagcaa tagacataag cggctattta acgaccctgc
3061 cctgaaccga cgaccgggtc atcgtggccg gatcttgcgg cccctcggct tgaacgaatt
3121 gttagacatt atttgccgac taccttggtg atctcgcctt tcacgtagtg gacaaattct
3181 tccaactgat ctgcgcgcga ggccaagcga tcttcttctt gtccaagata agcctgtcta
3241 gcttcaagta tgacgggctg atactgggcc ggcaggcgct ccattgccca gtcggcagcg
3301 acatccttcg gcgcgatttt gccggttact gcgctgtacc aaatgcggga caacgtaagc
3361 actacatttc gctcatcgcc agcccagtcg ggcggcgagt tccatagcgt taaggtttca
3421 tttagcgcct caaatagatc ctgttcagga accggatcaa agagttcctc cgccgctgga
3481 cctaccaagg caacgctatg ttctcttgct tttgtcagca agatagccag atcaatgtcg
3541 atcgtggctg gctcgaagat acctgcaaga atgtcattgc gctgccattc tccaaattgc
3601 agttcgcgct tagctggata acgccacgga atgatgtcgt cgtgcacaac aatggtgact
3661 tctacagcgc ggagaatctc gctctctcca ggggaagccg aagtttccaa aaggtcgttg
3721 atcaaagctc gccgcgttgt ttcatcaagc cttacggtca ccgtaaccag caaatcaata
3781 tcactgtgtg gcttcaggcc gccatccact gcggagccgt acaaatgtac ggccagcaac
3841 gtcggttcga gatggcgctc gatgacgcca actacctctg atagttgagt cgatacttcg
3901 gcgatcaccg cttccctcat actcttcctt tttcaatatt attgaagcat ttatcagggt
3961 tattgtctca tgagcggata catatttgaa tgtatttaga aaaataaaca aatagccagc
4021 tcactcggtc gctacgctcc gggcgtgaga ctgcggcggg cgctgcggac acatacaaag
4081 ttacccacag attccgtgga taagcagggg actaacatgt gaggcaaaac agcagggccg
4141 cgccggtggc gtttttccat aggctccgcc ctcctgccag agttcacata aacagacgct
4201 tttccggtgc atctgtggga gccgtgaggc tcaaccatga atctgacagt acgggcgaaa
4261 cccgacagga cttaaagatc cccaccgttt ccggcgggtc gctccctctt gcgctctcct
4321 gttccgaccc tgccgtttac cggatacctg ttccgccttt ctcccttacg ggaagtgtgg
4381 cgctttctca tagctcacac actggtatct cggctcggtg taggtcgttc gctccaagct
4441 gggctgtaag caagaactcc ccgttcagcc cgactgctgc gccttatccg gtaactgttc
4501 acttgagtcc aacccggaaa agcacggtaa aacgccactg gcagcagcca ttggtaactg
4561 ggagttcgca gaggatttgt ttagctaaac acgcggttgc tcttgaagtg tgcgccaaag
4621 tccggctaca ctggaaggac agatttggtt gctgtgctct gcgaaagcca gttaccacgg
4681 ttaagcagtt ccccaactga cttaaccttc gatcaaacca cctccccagg tggtttttc
4741 gtttacaggg caaaagatta cgcgcagaaa aaaaggatct caagaagatc ctttgatctt
4801 ttctactgaa ccgctctaga tttcagtgca atttatctct tcaaatgtag cacctgaagt
4861 cagccccata cgatataagt tgtaattctc atgttagtca tgccccgcgc ccaccggaag

```
  4921 gagctgactg ggttgaaggc tctcaagggc atcggtcgag atcccggtgc ctaatgagtg
  4981 agctaactta cattaattgc gttgcgctca ctgcccgctt tccagtcggg aaacctgtcg
  5041 tgccagctgc attaatgaat cggccaacgc gcggggagag gcggtttgcg tattgggcgc
  5101 cagggtggtt tttcttttca ccagtgagac gggcaacagc tgattgccct tcaccgcctg
  5161 gccctgagag agttgcagca agcggtccac gctggtttgc cccagcaggc gaaaatcctg
  5221 tttgatggtg gttaacggcg ggatataaca tgagctgtct tcggtatcgt cgtatcccac
  5281 taccgagatg tccgcaccaa cgcgcagccc ggactcggta atggcgcgca ttgcgcccag
  5341 cgccatctga tcgttggcaa ccagcatcgc agtgggaacg atgccctcat tcagcatttg
  5401 catggtttgt tgaaaaccgg acatggcact ccagtcgcct tcccgttccg ctatcggctg
  5461 aatttgattg cgagtgagat atttatgcca gccagccaga cgcagacgcg ccgagacaga
  5521 acttaatggg ccc
//
```

5. sgRNA_MS2-bacteria/pUC19

```
LOCUS       Exported                2606 bp ds-DNA     circular SYN 29-SEP-
2018
DEFINITION  pgRNA_MS2-bacteria/pUC19, synthetic circular DNA
ACCESSION   .
VERSION     .
KEYWORDS    .
SOURCE      synthetic DNA construct
  ORGANISM  synthetic DNA construct
REFERENCE   1  (bases 1 to 2606)
  AUTHORS   .
  TITLE     Direct Submission
  JOURNAL   Exported Sep 29, 2018 from SnapGene Viewer 4.2.5
            http://www.snapgene.com
FEATURES             Location/Qualifiers
     source          1..2606
                     /organism="synthetic DNA construct"
                     /mol_type="other DNA"
     promoter        16..50
                     /label=J23119(SpeI) promoter
                     /note="bacterial promoter (Registry of Standard
Biological
                     Parts BBa_J23119) modified to end with an SpeI site"
     misc_RNA        68..143
                     /label=gRNA scaffold
                     /note="guide RNA scaffold for the CRISPR/Cas9 system"
     misc_RNA        148..166
                     /label=MS2 stem loop
                     /note="stem loop that binds the bacteriophage MS2 coat
                     protein"
     CDS             189..218
                     /codon_start=1
                     /product="Myc (human c-Myc oncogene) epitope tag"
                     /label=Myc
                     /translation="EQKLISEEDL"
     CDS             234..251
                     /codon_start=1
                     /product="6xHis affinity tag"
                     /label=6xHis
                     /translation="HHHHHH"
     terminator      477..523
                     /label=rrnB T1 terminator
                     /note="transcription terminator T1 from the E. coli
rrnB
                     gene"
```

```
     terminator      477..522
                     /label=rrnB T1 terminator
                     /note="transcription terminator T1 from the E. coli
rrnB
                     gene"
     terminator      480..523
                     /gene="Escherichia coli rrnB"
                     /label=rrnB T1 terminator
                     /note="transcription terminator T1 from the E. coli
rrnB
                     gene"
     rep_origin      complement(747..1335)
                     /direction=LEFT
                     /label=ori
                     /note="high-copy-number ColE1/pMB1/pBR322/pUC origin of
                     replication"
     CDS             complement(1506..2366)
                     /codon_start=1
                     /gene="bla"
                     /product="beta-lactamase"
                     /label=AmpR
                     /note="confers resistance to ampicillin, carbenicillin,
and
                     related antibiotics"

/translation="MSIQHFRVALIPFFAAFCLPVFAHPETLVKVKDAEDQLGARVGYI

ELDLNSGKILESFRPEERFPMMSTFKVLLCGAVLSRIDAGQEQLGRRIHYSQNDLVEYS

PVTEKHLTDGMTVRELCSAAITMSDNTAANLLLTTIGGPKELTAFLHNMGDHVTRLDRW

EPELNEAIPNDERDTTMPVAMATTLRKLLTGELLTLASRQQLIDWMEADKVAGPLLRSA

LPAGWFIADKSGAGERGSRGIIAALGPDGKPSRIVVIYTTGSQATMDERNRQIAEIGAS
                     LIKHW"
     promoter        complement(2367..2471)
                     /gene="bla"
                     /label=AmpR promoter
     terminator      2561..2604
                     /label=bacterial terminator
                     /note="putative bacterial transcription terminator"
ORIGIN
        1 gaattctaaa gatctttgac agctagctca gtcctaggta taatactagt ggtcttcgag
       61 aagacctgtt ttagagctag aaatagcaag ttaaaataag gctagtccgt tatcaacttg
      121 aaaaagtggc accgagtcgg tgcgcgcaca tgaggatcac ccatgtgctt tttttgaagc
      181 ttgggcccga acaaaaactc atctcagaag aggatctgaa tagcgccgtc gaccatcatc
      241 atcatcatca ttgagtttaa acggtctcca gcttggctgt tttggcggat gagagaagat
      301 tttcagcctg atacagatta aatcagaacg cagaagcggt ctgataaaac agaatttgcc
      361 tggcggcagt agcgcggtgg tcccacctga ccccatgccg aactcagaag tgaaacgccg
      421 tagcgccgat ggtagtgtgg ggtctcccca tgcgagagta gggaactgcc aggcatcaaa
      481 taaaacgaaa ggctcagtcg aaagactggg cctttcgttt tatctgttgt ttgtcggtga
      541 actggatcct tactcgagtc tagactgcag gcttcctcgc tcactgactc gctgcgctcg
      601 gtcgttcggc tgcggcgagc ggtatcagct cactcaaagg cggtaatacg gttatccaca
      661 gaatcagggg ataacgcagg aaagaacatg tgagcaaaag gccagcaaaa ggccaggaac
      721 cgtaaaaagg ccgcgttgct ggcgtttttc cataggctcc gcccccctga cgagcatcac
      781 aaaaatcgac gctcaagtca gaggtggcga aacccgacag gactataaag ataccaggcg
      841 tttccccctg gaagctccct cgtgcgctct cctgttccga ccctgccgct taccggatac
      901 ctgtccgcct ttctcccttc gggaagcgtg gcgctttctc atagctcacg ctgtaggtat
      961 ctcagttcgg tgtaggtcgt tcgctccaag ctgggctgtg tgcacgaacc ccccgttcag
     1021 cccgaccgct gcgccttatc cggtaactat cgtcttgagt ccaacccggt aagacacgac
     1081 ttatcgccac tggcagcagc cactggtaac aggattagca gagcgaggta tgtaggcggt
     1141 gctacagagt tcttgaagtg gtggcctaac tacggctaca ctagaaggac agtatttggt
     1201 atctgcgctc tgctgaagcc agttaccttc ggaaaaagag ttggtagctc ttgatccggc
```

```
        1261 aaacaaacca ccgctggtag cggtggtttt tttgtttgca agcagcagat tacgcgcaga
        1321 aaaaaaggat ctcaagaaga tcctttgatc ttttctacgg ggtctgacgc tcagtggaac
        1381 gaaaactcac gttaagggat tttggtcatg agattatcaa aaaggatctt cacctagatc
        1441 cttttaaatt aaaaatgaag ttttaaatca atctaaagta tatatgagta aacttggtct
        1501 gacagttacc aatgcttaat cagtgaggca cctatctcag cgatctgtct atttcgttca
        1561 tccatagttg cctgactccc cgtcgtgtag ataactacga tacgggaggg cttaccatct
        1621 ggccccagtg ctgcaatgat accgcgagac ccacgctcac cggctccaga tttatcagca
        1681 ataaaccagc cagccggaag ggccgagcgc agaagtggtc ctgcaacttt atccgcctcc
        1741 atccagtcta ttaattgttg ccgggaagct agagtaagta gttcgccagt taatagtttg
        1801 cgcaacgttg ttgccattgc tacaggcatc gtggtgtcac gctcgtcgtt tggtatggct
        1861 tcattcagct ccggttccca acgatcaagg cgagttacat gatcccccat gttgtgcaaa
        1921 aaagcggtta gctccttcgg tcctccgatc gttgtcagaa gtaagttggc cgcagtgtta
        1981 tcactcatgg ttatggcagc actgcataat tctcttactg tcatgccatc cgtaagatgc
        2041 ttttctgtga ctggtgagta ctcaaccaag tcattctgag aatagtgtat gcggcgaccg
        2101 agttgctctt gcccggcgtc aatacgggat aataccgcgc cacatagcag aactttaaaa
        2161 gtgctcatca ttggaaaacg ttcttcgggg cgaaaactct caaggatctt accgctgttg
        2221 agatccagtt cgatgtaacc cactcgtgca cccaactgat cttcagcatc ttttactttc
        2281 accagcgttt ctgggtgagc aaaaacagga aggcaaaatg ccgcaaaaaa gggaataagg
        2341 gcgacacgga aatgttgaat actcatactc ttcctttttc aatattattg aagcatttat
        2401 cagggttatt gtctcatgag cggatacata tttgaatgta tttagaaaaa taaacaaata
        2461 ggggttccgc gcacatttcc ccgaaaagtg ccacctgacg tctaagaaac cattattatc
        2521 atgacattaa cctataaaaa taggcgtatc acgaggcaga atttcagata aaaaaaatcc
        2581 ttagctttcg ctaaggatga tttctg
   //
```

6. sgRNA_2xMS2-bacteria/pUC19

```
LOCUS       Exported                2648 bp ds-DNA     circular SYN 08-AUG-2016
DEFINITION  pgRNA_2xMS2-bacteria/pUC19, synthetic circular DNA
ACCESSION   .
VERSION     .
KEYWORDS    .
SOURCE      synthetic DNA construct
  ORGANISM  synthetic DNA construct
REFERENCE   1  (bases 1 to 2648)
  AUTHORS   .
  TITLE     Direct Submission
  JOURNAL   Exported Sep 29, 2018 from SnapGene Viewer 4.2.5
            http://www.snapgene.com
FEATURES             Location/Qualifiers
     source          1..2648
                     /organism="synthetic DNA construct"
                     /mol_type="other DNA"
     promoter        16..50
                     /label=J23119(SpeI) promoter
                     /note="bacterial promoter (Registry of Standard Biological
                     Parts BBa_J23119) modified to end with an SpeI site"
     misc_RNA        68..143
                     /label=gRNA scaffold
                     /note="guide RNA scaffold for the CRISPR/Cas9 system"
     misc_RNA        150..168
                     /label=MS2 stem loop
                     /note="stem loop that binds the bacteriophage MS2 coat
                     protein"
     misc_RNA        179..197
                     /label=MS2 stem loop
                     /note="stem loop that binds the bacteriophage MS2 coat
                     protein"
     CDS             231..260
```

```
                     /codon_start=1
                     /product="Myc (human c-Myc oncogene) epitope tag"
                     /label=Myc
                     /translation="EQKLISEEDL"
     CDS             231..260
                     /codon_start=1
                     /product="Myc (human c-Myc oncogene) epitope tag"
                     /label=Myc
                     /translation="EQKLISEEDL"
     CDS             276..293
                     /codon_start=1
                     /product="6xHis affinity tag"
                     /label=6xHis
                     /translation="HHHHHH"
     CDS             276..293
                     /codon_start=1
                     /product="6xHis affinity tag"
                     /label=6xHis
                     /translation="HHHHHH"
     terminator      519..565
                     /label=rrnB T1 terminator
                     /note="transcription terminator T1 from the E. coli rrnB
                     gene"
     terminator      519..564
                     /label=rrnB T1 terminator
                     /note="transcription terminator T1 from the E. coli rrnB
                     gene"
     terminator      522..565
                     /gene="Escherichia coli rrnB"
                     /label=rrnB T1 terminator
                     /note="transcription terminator T1 from the E. coli rrnB
                     gene"
     primer_bind     complement(566..586)
                     /label=R
     rep_origin      complement(789..1377)
                     /direction=LEFT
                     /label=ori
                     /note="high-copy-number ColE1/pMB1/pBR322/pUC origin of
                     replication"
     CDS             complement(1548..2408)
                     /codon_start=1
                     /gene="bla"
                     /product="beta-lactamase"
                     /label=AmpR
                     /note="confers resistance to ampicillin, carbenicillin,
and
                     related antibiotics"

/translation="MSIQHFRVALIPFFAAFCLPVFAHPETLVKVKDAEDQLGARVGYI

ELDLNSGKILESFRPEERFPMMSTFKVLLCGAVLSRIDAGQEQLGRRIHYSQNDLVEYS

PVTEKHLTDGMTVRELCSAAITMSDNTAANLLLTTIGGPKELTAFLHNMGDHVTRLDRW

EPELNEAIPNDERDTTMPVAMATTLRKLLTGELLTLASRQQLIDWMEADKVAGPLLRSA

LPAGWFIADKSGAGERGSRGIIAALGPDGKPSRIVVIYTTGSQATMDERNRQIAEIGAS
                     LIKHW"
     promoter        complement(2409..2513)
                     /gene="bla"
                     /label=AmpR promoter
     terminator      2603..2646
                     /label=bacterial terminator
```

```
                     /note="putative bacterial transcription terminator"
     primer_bind     join(2648,1..22)
                     /label=F
ORIGIN
        1 gaattctaaa gatctttgac agctagctca gtcctaggta taatactagt ggtcttcgag
       61 aagacctgtt ttagagctag aaatagcaag ttaaaataag gctagtccgt tatcaacttg
      121 aaaaagtggc accgagtcgg tgcgggagca catgaggatc acccatgtgc cacgagcgac
      181 atgaggatca cccatgtcgc tcgtgttccc tttttttgaa gcttgggccc gaacaaaaac
      241 tcatctcaga agaggatctg aatagcgccg tcgaccatca tcatcatcat cattgagttt
      301 aaacggtctc cagcttggct gttttggcgg atgagagaag attttcagcc tgatacagat
      361 taaatcagaa cgcagaagcg gtctgataaa acagaatttg cctggcggca gtagcgcggt
      421 ggtcccacct gaccccatgc cgaactcaga agtgaaacgc cgtagcgccg atggtagtgt
      481 ggggtctccc catgcgagag tagggaactg ccaggcatca aataaaacga aaggctcagt
      541 cgaaagactg ggcctttcgt tttatctgtt gtttgtcggt gaactggatc cttactcgag
      601 tctagactgc aggcttcctc gctcactgac tcgctgcgct cggtcgttcg gctgcggcga
      661 gcggtatcag ctcactcaaa ggcggtaata cggttatcca cagaatcagg ggataacgca
      721 ggaaagaaca tgtgagcaaa aggccagcaa aaggccagga accgtaaaaa ggccgcgttg
      781 ctggcgtttt tccataggct ccgcccccct gacgagcatc acaaaaatcg acgctcaagt
      841 cagaggtggc gaaacccgac aggactataa agataccagg cgtttccccc tggaagctcc
      901 ctcgtgcgct ctcctgttcc gaccctgccg cttaccggat acctgtccgc ctttctccct
      961 tcgggaagcg tggcgctttc tcatagctca cgctgtaggt atctcagttc ggtgtaggtc
     1021 gttcgctcca agctgggctg tgtgcacgaa ccccccgttc agcccgaccg ctgcgcctta
     1081 tccggtaact atcgtcttga gtccaacccg gtaagacacg acttatcgcc actggcagca
     1141 gccactggta acaggattag cagagcgagg tatgtaggcg gtgctacaga gttcttgaag
     1201 tggtggccta actacggcta cactagaagg acagtatttg gtatctgcgc tctgctgaag
     1261 ccagttacct tcggaaaaag agttggtagc tcttgatccg gcaaacaaac caccgctggt
     1321 agcggtggtt tttttgtttg caagcagcag attacgcgca gaaaaaaagg atctcaagaa
     1381 gatcctttga tcttttctac ggggtctgac gctcagtgga acgaaaactc acgttaaggg
     1441 attttggtca tgagattatc aaaaaggatc ttcacctaga tccttttaaa ttaaaaatga
     1501 agttttaaat caatctaaag tatatatgag taaacttggt ctgacagtta ccaatgctta
     1561 atcagtgagg cacctatctc agcgatctgt ctatttcgtt catccatagt tgcctgactc
     1621 cccgtcgtgt agataactac gatacgggag ggcttaccat ctggccccag tgctgcaatg
     1681 ataccgcgag acccacgctc accggctcca gatttatcag caataaacca gccagccgga
     1741 agggccgagc gcagaagtgg tcctgcaact ttatccgcct ccatccagtc tattaattgt
     1801 tgccgggaag ctagagtaag tagttcgcca gttaatagtt tgcgcaacgt tgttgccatt
     1861 gctacaggca tcgtggtgtc acgctcgtcg tttggtatgg cttcattcag ctccggttcc
     1921 caacgatcaa ggcgagttac atgatccccc atgttgtgca aaaaagcggt tagctccttc
     1981 ggtcctccga tcgttgtcag aagtaagttg gccgcagtgt tatcactcat ggttatggca
     2041 gcactgcata attctcttac tgtcatgcca tccgtaagat gcttttctgt gactggtgag
     2101 tactcaacca agtcattctg agaatagtgt atgcggcgac cgagttgctc ttgcccggcg
     2161 tcaatacggg ataataccgc gccacatagc agaactttaa aagtgctcat cattggaaaa
     2221 cgttcttcgg ggcgaaaact ctcaaggatc ttaccgctgt tgagatccag ttcgatgtaa
     2281 cccactcgtg cacccaactg atcttcagca tcttttactt tcaccagcgt ttctgggtga
     2341 gcaaaaacag gaaggcaaaa tgccgcaaaa aagggaataa gggcgacacg gaaatgttga
     2401 atactcatac tcttccttt tcaatattat tgaagcattt atcagggtta ttgtctcatg
     2461 agcggataca tatttgaatg tatttagaaa aataaacaaa tagggggttcc gcgcacattt
     2521 ccccgaaaag tgccacctga cgtctaagaa accattatta tcatgacatt aacctataaa
     2581 aataggcgta tcacgaggca gaatttcaga taaaaaaaat ccttagcttt cgctaaggat
     2641 gatttctg
//
```

Mammalian Expression Plasmid Sequences

1. pCMV_nfEGFP

```
LOCUS       Exported                5112 bp ds-DNA     circular SYN 09-OCT-2018
DEFINITION  Non-fluorescent EGFP synthetic circular DNA
ACCESSION   .
VERSION     .
KEYWORDS    .
SOURCE      synthetic DNA construct
  ORGANISM  synthetic DNA construct
REFERENCE   1  (bases 1 to 5112)
  AUTHORS   Juan-Carlos Collantes
  TITLE     Direct Submission
  JOURNAL   Exported Oct 9, 2018 from SnapGene Viewer 4.2.6
            http://www.snapgene.com
FEATURES             Location/Qualifiers
     source          1..5112
                     /organism="synthetic DNA construct"
                     /mol_type="other DNA"
     polyA_signal    417..551
                     /label=SV40 poly(A) signal
                     /note="SV40 polyadenylation signal"
     rep_origin      677..1132
                     /direction=RIGHT
                     /label=f1 ori
                     /note="f1 bacteriophage origin of replication; arrow
                     indicates direction of (+) strand synthesis"
     CDS             1420..2280
                     /codon_start=1
                     /gene="bla"
                     /product="beta-lactamase"
                     /label=AmpR
                     /note="confers resistance to ampicillin, carbenicillin,
and
                     related antibiotics"

/translation="MSIQHFRVALIPFFAAFCLPVFAHPETLVKVKDAEDQLGARVGYI

ELDLNSGKILESFRPEERFPMMSTFKVLLCGAVLSRIDAGQEQLGRRIHYSQNDLVEYS

PVTEKHLTDGMTVRELCSAAITMSDNTAANLLLTTIGGPKELTAFLHNMGDHVTRLDRW

EPELNEAIPNDERDTTMPVAMATTLRKLLTGELLTLASRQQLIDWMEADKVAGPLLRSA

LPAGWFIADKSGAGERGSRGIIAALGPDGKPSRIVVIYTTGSQATMDERNRQIAEIGAS
                     LIKHW"
     protein_bind    complement(2382..2415)
                     /label=loxP
                     /bound_moiety="Cre recombinase"
                     /note="Cre-mediated recombination occurs in the 8-bp core
                     sequence (GCATACAT)."
     rep_origin      2493..3081
                     /direction=RIGHT
                     /label=ori
```

```
                     /note="high-copy-number ColE1/pMB1/pBR322/pUC origin of
                     replication"
     enhancer        3541..3844
                     /label=CMV enhancer
                     /note="human cytomegalovirus immediate early enhancer"
     promoter        3845..4048
                     /label=CMV promoter
                     /note="human cytomegalovirus (CMV) immediate early
                     promoter"
     primer_bind     4174..4190
                     /label=M13 rev
                     /note="common sequencing primer, one of multiple similar
                     variants"
     promoter        4201..4219
                     /label=SP6 promoter
                     /note="promoter for bacteriophage SP6 RNA polymerase"
     protein_bind    4220..4244
                     /gene="mutant version of attB"
                     /label=attB1
                     /bound_moiety="BP Clonase(TM)"
                     /note="recombination site for the Gateway(R) BP reaction"
     regulatory      4280..4289
                     /regulatory_class="other"
                     /note="vertebrate consensus sequence for strong initiation
                     of translation (Kozak, 1987)"
     CDS             4286..5005
                     /codon_start=1
                     /product="non-fluorescent enhanced GFP"
                     /label=nfEGFP
                     /note="mammalian codon-optimized"

/translation="MVSKGEELFTGVVPILVELDGDVNGHKFSVSGEGEGDATYGKLTL

KFICTTGKLPVPWPTLVTTLTCGVQCFSRYPDHMKQHDFFKSAMPEGYVQERTIFFKDD

GNYKTRAEVKFEGDTLVNRIELKGIDFKEDGNILGHKLEYNYNSHNVYIMADKQKNGIK

VNFKIRHNIEDGSVQLADHYQQNTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHMVLL
                     EFVTAAGITLGMDELYK"
     protein_bind    complement(5053..5077)
                     /gene="mutant version of attB"
                     /label=attB2
                     /bound_moiety="BP Clonase(TM)"
                     /note="recombination site for the Gateway(R) BP reaction"
     promoter        complement(5079..5097)
                     /label=T7 promoter
                     /note="promoter for bacteriophage T7 RNA polymerase"
     primer_bind     complement(join(5108..5112,1..12))
                     /label=M13 fwd
                     /note="common sequencing primer, one of multiple similar
                     variants"
ORIGIN
        1 ccgtcgtttt acaacgtcgt gactgggaaa actgctagct tgggatcttt gtgaaggaac
       61 cttacttctg tggtgtgaca taattggaca aactacctac agagatttaa agctctaagg
      121 taaatataaa atttttaagt gtataatgtg ttaaactagc tgcatatgct tgctgcttga
      181 gagttttgct tactgagtat gatttatgaa aatattatac acaggagcta gtgattctaa
      241 ttgtttgtgt attttagatt cacagtccca aggctcattt caggcccctc agtcctcaca
      301 gtctgttcat gatcataatc agccatacca catttgtaga ggttttactt gctttaaaaa
      361 acctcccaca cctccccctg aacctgaaac ataaaatgaa tgcaattgtt gttgttaact
      421 tgtttattgc agcttataat ggttacaaat aaagcaatag catcacaaat ttcacaaata
      481 aagcattttt ttcactgcat tctagttgtg gtttgtccaa actcatcaat gtatcttatc
      541 atgtctggat cgatcctgca ttaatgaatc ggccaacgcg cggggagagg cggtttgcgt
      601 attggctggc gtaatagcga agaggcccgc accgatcgcc cttcccaaca gttgcgcagc
```

```
 661 ctgaatggcg aatgggacgc gccctgtagc ggcgcattaa gcgcggcggg tgtggtggtt
 721 acgcgcagcg tgaccgctac acttgccagc gccctagcgc ccgctccttt cgctttcttc
 781 ccttcctttc tcgccacgtt cgccggcttt ccccgtcaag ctctaaatcg ggggctccct
 841 ttagggttcc gatttagtgc tttacggcac ctcgacccca aaaaacttga ttagggtgat
 901 ggttcacgta gtgggccatc gccctgatag acggtttttc gccctttgac gttggagtcc
 961 acgttcttta atagtggact cttgttccaa actggaacaa cactcaaccc tatctcggtc
1021 tattcttttg atttataagg gattttgccg atttcggcct attggttaaa aaatgagctg
1081 atttaacaaa aatttaacgc gaattttaac aaaatattaa cgtttacaat ttcaggtggc
1141 acttttcggg gaaatgtgcg cggaacccct atttgtttat ttttctaaat acattcaaat
1201 atgtatccgc tcatgccagg tcttggactg gtgagaacgg cttgctcggc agcttcgatg
1261 tgtgctggag ggagaataaa ggtctaagat gtgcgataga gggaagtcgc attgaattat
1321 gtgctgtgta gggatcgctg gtatcaaata tgtgtgccca cccctggcat gagacaataa
1381 ccctgataaa tgcttcaata atattgaaaa aggaagagta tgagtattca acatttccgt
1441 gtcgccctta ttccctttt tgcggcattt tgccttcctg tttttgctca cccagaaacg
1501 ctggtgaaag taaaagatgc tgaagatcag ttgggtgcac gagtgggtta catcgaactg
1561 gatctcaaca gcggtaagat ccttgagagt tttcgccccg aagaacgttt tccaatgatg
1621 agcactttta aagttctgct atgtggcgcg gtattatccc gtattgacgc cgggcaagag
1681 caactcggtc gccgcataca ctattctcag aatgacttgg ttgagtactc accagtcaca
1741 gaaaagcatc ttacggatgg catgacagta agagaattat gcagtgctgc cataaccatg
1801 agtgataaca ctgcggccaa cttacttctg acaacgatcg gaggaccgaa ggagctaacc
1861 gctttttgc acaacatggg ggatcatgta actcgccttg atcgttggga accggagctg
1921 aatgaagcca taccaaacga cgagcgtgac accacgatgc ctgtagcaat ggcaacaacg
1981 ttgcgcaaac tattaactgg cgaactactt actctagctt cccggcaaca attaatagac
2041 tggatggagg cggataaagt tgcaggacca cttctgcgct cggcccttcc ggctggctgg
2101 tttattgctg ataaatctgg agccggtgag cgtgggtctc gcggtatcat tgcagcactg
2161 gggccagatg gtaagccctc ccgtatcgta gttatctaca cgacggggag tcaggcaact
2221 atggatgaac gaaatagaca gatcgctgag ataggtgcct cactgattaa gcattggtaa
2281 ctgtcagacc aagtttactc atatatactt tagattgatt taaaacttca tttttaattt
2341 aaaaggatct aggtgaagat cctttttgat aatctcatgc cataacttcg tataatgtat
2401 gctatacgaa gttatggcat gaccaaaatc ccttaacgtg agttttcgtt ccactgagcg
2461 tcagaccccg tagaaaagat caaaggatct tcttgagatc ctttttttct gcgcgtaatc
2521 tgctgcttgc aaacaaaaaa accaccgcta ccagcggtgg tttgtttgcc ggatcaagag
2581 ctaccaactc tttttccgaa ggtaactggc ttcagcagag cgcagatacc aaatactgtc
2641 cttctagtgt agccgtagtt aggccaccac ttcaagaact ctgtagcacc gcctacatac
2701 ctcgctctgc taatcctgtt accagtggct gctgccagtg gcgataagtc gtgtcttacc
2761 gggttggact caagacgata gttaccggat aaggcgcagc ggtcgggctg aacggggggt
2821 tcgtgcacac agcccagctt ggagcgaacg acctacaccg aactgagata cctacagcgt
2881 gagcattgag aaagcgccac gcttcccgaa gggagaaagg cggacaggta tccggtaagc
2941 ggcagggtcg gaacaggaga gcgcacgagg gagcttccag ggggaaacgc ctggtatctt
3001 tatagtcctg tcgggtttcg ccacctctga cttgagcgtc gatttttgtg atgctcgtca
3061 gggggggcgga gcctatggaa aaacgccagc aacgcggcct ttttacggtt cctggccttt
3121 tgctggcctt ttgctcacat gttctttcct gcgttatccc ctgattctgt ggataaccgt
3181 attaccgcct ttgagtgagc tgataccgct cgccgcagcc gaacgaccga gcgcagcgag
3241 tcagtgagcg aggaagcgga agagcgccca atacgcaaac cgcctctccc cgcgcgttgg
3301 ccgattcatt aatgcagagc ttgcaattcg cgcgtttttc aatattattg aagcatttat
3361 cagggttatt gtctcatgag cggatacata tttgaatgta tttagaaaaa taaacaaata
3421 ggggttccgc gcacatttcc ccgaaaagtg ccacctgacg tctaagaaac cattattatc
3481 atgacattaa cctataaaaa taggcgtagt acgaggccct ttcactcatt agatgcatgt
3541 cgttacataa cttacggtaa atggcccgcc tggctgaccg cccaacgacc cccgcccatt
3601 gacgtcaata atgacgtatg ttcccatagt aacgccaata gggactttcc attgacgtca
3661 atgggtggag tatttacggt aaactgccca cttggcagta catcaagtgt atcatatgcc
3721 aagtacgccc cctattgacg tcaatgacgg taaatggccc gcctggcatt atgcccagta
3781 catgacctta tgggactttc ctacttggca gtacatctac gtattagtca tcgctattac
3841 catggtgatg cggttttggc agtacatcaa tgggcgtgga tagcggtttg actcacgggg
3901 atttccaagt ctccacccca ttgacgtcaa tgggagtttg ttttggcacc aaaatcaacg
3961 ggactttcca aaatgtcgta acaactccgc cccattgacg caaatgggcg gtaggcgtgt
4021 acggtgggag gtctatataa gcagagctcg tttagtgaac cgtcagatcg cctggagacg
4081 ccatccacgc tgttttgacc tccatagaag acaccgggac cgatccagcc tccggactct
4141 agcctaggcc gcggagcgga taacaatttc acacaggaaa cagctatgac cattaggcct
4201 atttaggtga cactatagaa caagtttgta caaaaaagca ggctggtacc ggtccggaat
4261 tcccgggata tcgtcgacag ccaccatggt gagcaagggc gaggagctgt tcaccggggt
4321 ggtgcccatc ctggtcgagc tggacggcga cgtaaacggc cacaagttca gcgtgtccgg
4381 cgagggcgag ggcgatgcca cctacggcaa gctgaccctg aagttcatct gcaccaccgg
```

```
     4441 caagctgccc gtgccctggc ccaccctcgt gaccaccctg acctgcggcg tgcagtgctt
     4501 cagccgctac cccgaccaca tgaagcagca cgacttcttc aagtccgcca tgcccgaagg
     4561 ctacgtccag gagcgcacca tcttcttcaa ggacgacggc aactacaaga cccgcgccga
     4621 ggtgaagttc gagggcgaca ccctggtgaa ccgcatcgag ctgaagggca tcgacttcaa
     4681 ggaggacggc aacatcctgg ggcacaagct ggagtacaac tacaacagcc acaacgtcta
     4741 tatcatggcc gacaagcaga agaacggcat caaggtgaac ttcaagatcc gccacaacat
     4801 cgaggacggc agcgtgcagc tcgccgacca ctaccagcag aacacccccca tcggcgacgg
     4861 ccccgtgctg ctgcccgaca accactacct gagcacccag tccgccctga gcaaagaccc
     4921 caacgagaag cgcgatcaca tggtcctgct ggagttcgtg accgccgccg ggatcactct
     4981 cggcatggac gagctgtaca agtgataagc ggccgctcta gagtatccct cgaggggccc
     5041 aagcttacgc gtacccagct ttcttgtaca aagtggtccc tatagtgagt cgtattataa
     5101 gctaggcact gg
//
```

2. pUbC_ACRCn

```
LOCUS       Exported               15129 bp ds-DNA     circular SYN 08-OCT-2018
DEFINITION  pUbC_ACRCn synthetic circular DNA
ACCESSION   .
VERSION     .
KEYWORDS    .
SOURCE      synthetic DNA construct
  ORGANISM  synthetic DNA construct
REFERENCE   1  (bases 1 to 15129)
  AUTHORS   Juan-Carlos Collantes
  TITLE     Direct Submission
  JOURNAL   Exported Oct 8, 2018 from SnapGene Viewer 4.2.6
            http://www.snapgene.com
FEATURES             Location/Qualifiers
     source          1..15129
                     /organism="synthetic DNA construct"
                     /mol_type="other DNA"
     enhancer        238..617
                     /label=CMV enhancer
                     /note="human cytomegalovirus immediate early enhancer"
     promoter        618..821
                     /label=CMV promoter
                     /note="human cytomegalovirus (CMV) immediate early
                     promoter"
     LTR             835..1015
                     /label=5' LTR (truncated)
                     /note="truncated 5' long terminal repeat (LTR) from HIV-1"
     misc_feature    1062..1187
                     /label=HIV-1 Psi
                     /note="packaging signal of human immunodeficiency virus
                     type 1"
     misc_feature    1680..1913
                     /label=RRE
                     /note="The Rev response element (RRE) of HIV-1 allows for
                     Rev-dependent mRNA export from the nucleus to the
                     cytoplasm."
     misc_feature    2440..2557
                     /label=cPPT/CTS
                     /note="central polypurine tract and central termination
                     sequence of HIV-1"
     protein_bind    2602..2635
                     /label=loxP
                     /bound_moiety="Cre recombinase"
                     /note="Cre-mediated recombination occurs in the 8-bp core
                     sequence (GCATACAT)."
     CDS             complement(2622..2948)
```

```
                     /codon_start=1
                     /gene="lacZ fragment"
                     /product="LacZ-alpha fragment of beta-galactosidase"
                     /label=lacZ-alpha

/translation="MTMITPSAQLTLTKGNKSWRTSRGGPVPNSPYSESYYARSLAVVL

QRRDWENPGVTQLNRLAAHPPFASWRNSEEARTDRPSQQLRSLNGEWDAPCSGATTVSY
                     HNFV"
     primer_bind     2813..2829
                     /label=M13 fwd
                     /note="common sequencing primer, one of multiple similar
                     variants"
     promoter        2839..2857
                     /label=T7 promoter
                     /note="promoter for bacteriophage T7 RNA polymerase"
     promoter        complement(2905..2923)
                     /label=T3 promoter
                     /note="promoter for bacteriophage T3 RNA polymerase"
     primer_bind     complement(2944..2960)
                     /label=M13 rev
                     /note="common sequencing primer, one of multiple similar
                     variants"
     protein_bind    2968..2984
                     /label=lac operator
                     /bound_moiety="lac repressor encoded by lacI"
                     /note="The lac repressor binds to the lac operator to
                     inhibit transcription in E. coli. This inhibition can be
                     relieved by adding lactose or
                     isopropyl-beta-D-thiogalactopyranoside (IPTG)."
     promoter        complement(2992..3022)
                     /label=lac promoter
                     /note="promoter for the E. coli lac operon"
     protein_bind    3037..3058
                     /label=CAP binding site
                     /bound_moiety="E. coli catabolite activator protein"
                     /note="CAP binding activates transcription in the presence
                     of cAMP."
     promoter        3124..4334
                     /label=UbC promoter
                     /note="human ubiquitin C promoter"
     regulatory      4363..4372
                     /regulatory_class="other"
                     /note="vertebrate consensus sequence for strong initiation
                     of translation (Kozak, 1987)"
     CDS             4372..4437
                     /codon_start=1
                     /product="three tandem FLAG(R) epitope tags, followed by
an
                     enterokinase cleavage site"
                     /label=3xFLAG
                     /translation="DYKDHDGDYKDHDIDYKDDDDK"
     CDS             4444..4464
                     /codon_start=1
                     /product="nuclear localization signal of SV40 (simian
virus
                     40) large T antigen"
                     /label=SV40 NLS
                     /translation="PKKKRKV"
     CDS             4474..8577
                     /codon_start=1
                     /product="nickase mutant of the Cas9 endonuclease from the
                     Streptococcus pyogenes Type II CRISPR/Cas system"
```

```
                     /label=Cas9(D10A)
                     /note="generates RNA-guided single strand nicks in DNA due
                     to the D10A mutation in the RuvC catalytic domain"

/translation="MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKK

NLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEES

FLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIK

FRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRL

ENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQ

IGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVR

QQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLL

RKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARG

NSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEY

FTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFD

SVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEER

LKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFM

QLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGR

HKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLY

LYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVP

SEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKH

VAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYL

NAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKT

EITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSK

ESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGI

TIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNE

LALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILAD

ANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVL
                     DATLIHQSITGLYETRIDLSQLGGD"
     CDS             8590..8610
                     /codon_start=1
                     /product="nuclear localization signal of SV40 (simian
virus
                     40) large T antigen"
                     /label=SV40 NLS
                     /translation="PKKKRKV"
     CDS             8620..8676
                     /codon_start=1
                     /product="2A peptide from porcine teschovirus-1
                     polyprotein"
                     /label=P2A
                     /note="Eukaryotic ribosomes fail to insert a peptide bond
```

```
                     between the Gly and Pro residues, yielding separate
                     polypeptides."
                     /translation="ATNFSLLKQAGDVEENPGP"
     CDS             8677..8697
                     /codon_start=1
                     /product="nuclear localization signal of SV40 (simian
virus
                     40) large T antigen"
                     /label=SV40 NLS
                     /translation="PKKKRKV"
     CDS             8698..9291
                     /codon_start=1
                     /gene="AICDA"
                     /product="human activation-induced cytidine deaminase
                     (Dickerson et al., 2003)"
                     /label=AID

/translation="MDSLLMNRRKFLYQFKNVRWAKGRRETYLCYVVKRRDSATSFSLD

FGYLRNKNGCHVELLFLRYISDWDLDPGRCYRVTWFTSWSPCYDCARHVADFLRGNPNL

SLRIFTARLYFCEDRKAEPEGLRRLHRAGVQIAIMTFKDYFYCWNTFVENHERTFKAWE
                     GLHENSVRLSRQLRRILLPLYEVDDLRDAFRTLGL"
     misc_feature    9292..9366
                     /label=L25
     CDS             9367..9717
                     /codon_start=1
                     /label=MCP

/translation="MASNFTQFVLVDNGGTGDVTVAPSNFANGIAEWISSNSRSQAYKV

TCSVRQSSAQNRKYTIKVEVPKGAWRSYLNMELTIPIFATNSDCELIVKAMQGLLKDGN
                     PIPSAIAANSGIY"
     protein_bind    9763..9796
                     /label=loxP
                     /bound_moiety="Cre recombinase"
                     /note="Cre-mediated recombination occurs in the 8-bp core
                     sequence (GCATACAT)."
     misc_feature    9821..10409
                     /label=WPRE
                     /note="woodchuck hepatitis virus posttranscriptional
                     regulatory element"
     CDS             complement(10292..10303)
                     /codon_start=1
                     /product="Factor Xa recognition and cleavage site"
                     /label=Factor Xa site
                     /translation="IEGR"
     primer_bind     complement(10412..10428)
                     /label=KS primer
                     /note="common sequencing primer, one of multiple similar
                     variants"
     LTR             10934..11114
                     /label=5' LTR (truncated)
                     /note="truncated 5' long terminal repeat (LTR) from HIV-1"
     polyA_signal    11146..11370
                     /label=bGH poly(A) signal
                     /note="bovine growth hormone polyadenylation signal"
     rep_origin      11416..11844
                     /direction=RIGHT
                     /label=f1 ori
                     /note="f1 bacteriophage origin of replication; arrow
                     indicates direction of (+) strand synthesis"
     promoter        11858..12187
```

```
                     /label=SV40 promoter
                     /note="SV40 enhancer and early promoter"
     rep_origin      12038..12173
                     /label=SV40 ori
                     /note="SV40 origin of replication"
     promoter        12235..12282
                     /label=EM7 promoter
                     /note="synthetic bacterial promoter "
     CDS             12301..12675
                     /codon_start=1
                     /gene="Sh ble from Streptoalloteichus hindustanus"
                     /product="antibiotic-binding protein"
                     /label=BleoR
                     /note="confers resistance to bleomycin, phleomycin, and
                     Zeocin(TM)"

/translation="MAKLTSAVPVLTARDVAGAVEFWTDRLGFSRDFVEDDFAGVVRDD

VTLFISAVQDQVVPDNTLAWVWVRGLDELYAEWSEVVSTNFRDASGPAMTEIGEQPWGR
                     EFALRDPAGNCVHFVAEEQD"
     polyA_signal    12805..12926
                     /label=SV40 poly(A) signal
                     /note="SV40 polyadenylation signal"
     primer_bind     complement(12975..12991)
                     /label=M13 rev
                     /note="common sequencing primer, one of multiple similar
                     variants"
     protein_bind    12999..13015
                     /label=lac operator
                     /bound_moiety="lac repressor encoded by lacI"
                     /note="The lac repressor binds to the lac operator to
                     inhibit transcription in E. coli. This inhibition can be
                     relieved by adding lactose or
                     isopropyl-beta-D-thiogalactopyranoside (IPTG)."
     promoter        complement(13023..13053)
                     /label=lac promoter
                     /note="promoter for the E. coli lac operon"
     protein_bind    13068..13089
                     /label=CAP binding site
                     /bound_moiety="E. coli catabolite activator protein"
                     /note="CAP binding activates transcription in the presence
                     of cAMP."
     rep_origin      complement(13377..13965)
                     /direction=LEFT
                     /label=ori
                     /note="high-copy-number ColE1/pMB1/pBR322/pUC origin of
                     replication"
     CDS             complement(14136..14996)
                     /codon_start=1
                     /gene="bla"
                     /product="beta-lactamase"
                     /label=AmpR
                     /note="confers resistance to ampicillin, carbenicillin,
and
                     related antibiotics"

/translation="MSIQHFRVALIPFFAAFCLPVFAHPETLVKVKDAEDQLGARVGYI

ELDLNSGKILESFRPEERFPMMSTFKVLLCGAVLSRIDAGQEQLGRRIHYSQNDLVEYS

PVTEKHLTDGMTVRELCSAAITMSDNTAANLLLTTIGGPKELTAFLHNMGDHVTRLDRW

EPELNEAIPNDERDTTMPVAMATTLRKLLTGELLTLASRQQLIDWMEADKVAGPLLRSA
```

```
LPAGWFIADKSGAGERGSRGIIAALGPDGKPSRIVVIYTTGSQATMDERNRQIAEIGAS
                     LIKHW"
     promoter        complement(14997..15101)
                     /gene="bla"
                     /label=AmpR promoter
ORIGIN
        1 gtcgacggat cgggagatct cccgatcccc tatggtgcac tctcagtaca atctgctctg
       61 atgccgcata gttaagccag tatctgctcc ctgcttgtgt gttggaggtc gctgagtagt
      121 gcgcgagcaa aatttaagct acaacaaggc aaggcttgac cgacaattgc atgaagaatc
      181 tgcttagggt taggcgtttt gcgctgcttc gcgatgtacg ggccagatat acgcgttgac
      241 attgattatt gactagttat taatagtaat caattacggg gtcattagtt catagcccat
      301 atatggagtt ccgcgttaca taacttacgg taaatggccc gcctggctga ccgcccaacg
      361 acccccgccc attgacgtca ataatgacgt atgttcccat agtaacgcca atagggactt
      421 tccattgacg tcaatgggtg gagtatttac ggtaaactgc ccacttggca gtacatcaag
      481 tgtatcatat gccaagtacg ccccctattg acgtcaatga cggtaaatgg cccgcctggc
      541 attatgccca gtacatgacc ttatgggact ttcctacttg gcagtacatc tacgtattag
      601 tcatcgctat taccatggtg atgcggtttt ggcagtacat caatgggcgt ggatagcggt
      661 ttgactcacg gggatttcca agtctccacc ccattgacgt caatgggagt ttgttttggc
      721 accaaaatca acgggacttt ccaaaatgtc gtaacaactc cgccccattg acgcaaatgg
      781 gcggtaggcg tgtacggtgg gaggtctata taagcagcgc gttttgcctg tactgggtct
      841 ctctggttag accagatctg agcctgggag ctctctggct aactagggaa cccactgctt
      901 aagcctcaat aaagcttgcc ttgagtgctt caagtagtgt gtgcccgtct gttgtgtgac
      961 tctggtaact agagatccct cagacccttt tagtcagtgt ggaaaatctc tagcagtggc
     1021 gcccgaacag ggacttgaaa gcgaaaggga aaccagagga gctctctcga cgcaggactc
     1081 ggcttgctga agcgcgcacg gcaagaggcg aggggcggcg actggtgagt acgccaaaaa
     1141 ttttgactag cggaggctag aaggagagag atgggtgcga gagcgtcagt attaagcggg
     1201 ggagaattag atcgcgatgg gaaaaaattc ggttaaggcc agggggaaag aaaaaatata
     1261 aattaaaaca tatagtatgg gcaagcaggg agctagaacg attcgcagtt aatcctggcc
     1321 tgttagaaac atcagaaggc tgtagacaaa tactgggaca gctacaacca tcccttcaga
     1381 caggatcaga agaacttaga tcattatata atacagtagc aaccctctat tgtgtgcatc
     1441 aaaggataga gataaaagac accaaggaag ctttagacaa gatagaggaa gagcaaaaca
     1501 aaagtaagac caccgcacag caagcggccg ctgatcttca gacctggagg aggagatatg
     1561 agggacaatt ggagaagtga attatataaa tataaagtag taaaaattga accattagga
     1621 gtagcaccca ccaaggcaaa gagaagagtg gtgcagagag aaaaaagagc agtgggaata
     1681 ggagctttgt tccttgggtt cttgggagca gcaggaagca ctatgggcgc agcgtcaatg
     1741 acgctgacgg tacaggccag acaattattg tctggtatag tgcagcagca gaacaatttg
     1801 ctgagggcta ttgaggcgca acagcatctg ttgcaactca cagtctgggg catcaagcag
     1861 ctccaggcaa gaatcctggc tgtggaaaga tacctaaagg atcaacagct cctggggatt
     1921 tggggttgct ctggaaaact catttgcacc actgctgtgc cttggaatgc tagttggagt
     1981 aataaatctc tggaacagat ttggaatcac acgacctgga tggagtggga cagagaaatt
     2041 aacaattaca caagcttaat acactcctta attgaagaat cgcaaaacca gcaagaaaag
     2101 aatgaacaag aattattgga attagataaa tgggcaagtt tgtggaattg gtttaacata
     2161 acaaattggc tgtggtatat aaaattattc ataatgatag taggaggctt ggtaggttta
     2221 agaatagttt ttgctgtact ttctatagtg aatagagtta ggcagggata ttcaccatta
     2281 tcgtttcaga cccacctccc aaccccgagg ggacccgaca ggcccgaagg aatagaagaa
     2341 gaaggtggag agagagacag agacagatcc attcgattag tgaacggatc ggcactgcgt
     2401 gcgccaattc tgcagacaaa tggcagtatt catccacaat tttaaaagaa aagggggat
     2461 tgggggtac agtgcagggg aaagaatagt agacataata gcaacagaca tacaaactaa
     2521 agaattacaa aaacaaatta caaaaattca aaattttcgg gtttattaca gggacagcag
     2581 agatccagtt tggttaatta aataacttcg tatagcatac attatacgaa gttatgataa
     2641 gagacggtgg tggcgccgct acagggcgcg tcccattcgc cattcaggct gcgcaactgt
     2701 tgggaagggc gatcggtgcg ggcctcttcg ctattacgcc agctggcgaa agggggatgt
     2761 gctgcaaggc gattaagttg ggtaacgcca gggttttccc agtcacgacg ttgtaaaacg
     2821 acggccagtg agcgcgcgta atacgactca ctatagggcg aattgggtac cgggcccccc
     2881 ctcgaggtcc tccagctttt gttccttta gtgagggtta attgcgcgct tggcgtaatc
     2941 atggtcatag ctgtttcctg tgtgaaattg ttatccgctc acaattccac acaacatacg
     3001 agccggaagc ataaagtgta aagcctgggg tgcctaatga gtgagctaac tcacattaat
     3061 tgcgttgcgc tcactgcccg ctttccactg catgacgtct ccacaattaa ttaagggtgc
     3121 agcggcctcc gcgccgggtt ttggcgcctc ccgcgggcgc ccccctcctc acggcgagcg
     3181 ctgccacgtc agacgaaggg cgcaggagcg ttcctgatcc ttccgcccgg acgctcagga
     3241 cagcggcccg ctgctcataa gactcggcct tagaacccca gtatcagcag aaggacattt
     3301 taggacggga cttgggtgac tctagggcac tggttttctt tccagagagc ggaacaggcg
```

3361 aggaaaagta gtcccttctc ggcgattctg cggagggatc tccgtggggc ggtgaacgcc
3421 gatgattata taaggacgcg ccgggtgtgg cacagctagt tccgtcgcag ccgggatttg
3481 ggtcgcggtt cttgtttgtg gatcgctgtg atcgtcactt ggtgagttgc gggctgctgg
3541 gctggccggg gctttcgtgg ccgccgggcc gctcggtggg acggaagcgt gtggagagac
3601 cgccaagggc tgtagtctgg gtccgcgagc aaggttgccc tgaactgggg gttgggggga
3661 gcgcacaaaa tggcggctgt tcccgagtct tgaatggaag acgcttgtaa ggcgggctgt
3721 gaggtcgttg aaacaaggtg gggggcatgg tgggcggcaa gaacccaagg tcttgaggcc
3781 ttcgctaatg cgggaaagct cttattcggg tgagatgggc tggggcacca tctggggacc
3841 ctgacgtgaa gtttgtcact gactggagaa ctcgggtttg tcgtctggtt gcgggggcgg
3901 cagttatgcg gtgccgttgg gcagtgcacc cgtacctttg ggagcgcgcg cctcgtcgtg
3961 tcgtgacgtc acccgttctg ttggcttata atgcagggtg gggccacctg ccggtaggtg
4021 tgcggtaggc ttttctccgt cgcaggacgc agggttcggg cctagggtag gctctcctga
4081 atcgacaggc gccggacctc tggtgagggg agggataagt gaggcgtcag tttctttggt
4141 cggttttatg tacctatctt cttaagtagc tgaagctccg gttttgaact atgcgctcgg
4201 ggttggcgag tgtgttttgt gaagtttttt aggcaccttt tgaaatgtaa tcatttgggt
4261 caatatgtaa ttttcagtgt tagactagta aattgtccgc taaattctgg ccgtttttgg
4321 ctttttgtt agacgaagct tgggctgcag gtcgactcta gagccaccat ggactacaaa
4381 gaccatgacg gtgattataa agatcatgac atcgattaca aggatgacga tgacaagatg
4441 gcccccaaga agaagaggaa ggtgggccgc ggaatggaca agaagtactc cattgggctc
4501 gccatcggca caaacagcgt cggctgggcc gtcattacgg acgagtacaa ggtgccgagc
4561 aaaaaattca aagttctggg caataccgat cgccacagca taaagaagaa cctcattggc
4621 gccctcctgt tcgactccgg ggaaaccgcc gaagccacgc ggctcaaaag aacagcacgg
4681 cgcagatata cccgcagaaa gaatcggatc tgctacctgc aggagatctt tagtaatgag
4741 atggctaagg tggatgactc tttcttccat aggctggagg agtccttttt ggtggaggag
4801 gataaaaagc acgagcgcca cccaatcttt ggcaatatcg tggacgaggt ggcgtaccat
4861 gaaaagtacc caaccatata tcatctgagg aagaagcttg tagacagtac tgataaggct
4921 gacttgcggt tgatctatct cgcgctggcg catatgatca aatttcgggg acacttcctc
4981 atcgaggggg acctgaaccc agacaacagc gatgtcgaca aactctttat ccaactggtt
5041 cagacttaca atcagctttt cgaagagaac ccgatcaacg catccggagt tgacgccaaa
5101 gcaatcctga gcgctaggct gtccaaatcc cggcggctcg aaaacctcat cgcacagctc
5161 cctggggaga agaagaacgg cctgtttggt aatcttatcg ccctgtcact cgggctgacc
5221 cccaacttta aatctaactt cgacctggcc gaagatgcca agcttcaact gagcaaagac
5281 acctacgatg atgatctcga caatctgctg gcccagatcg gcgaccagta cgcagacctt
5341 tttttggcgg caaagaacct gtcagacgcc attctgctga gtgatattct gcgagtgaac
5401 acggagatca ccaaagctcc gctgagcgct agtatgatca agcgctatga tgagcaccac
5461 caagacttga ctttgctgaa ggcccttgtc agacagcaac tgcctgagaa gtacaaggaa
5521 attttcttcg atcagtctaa aaatggctac gccggataca ttgacggcgg agcaagccag
5581 gaggaatttt acaaatttat taagcccatc ttggaaaaaa tggacggcac cgaggagctg
5641 ctggtaaagc ttaacagaga agatctgttg cgcaaacagc gcactttcga caatggaagc
5701 atccccccacc agattcacct gggcgaactg cacgctatcc tcaggcggca agaggatttc
5761 taccccttttt tgaaagataa cagggaaaag attgagaaaa tcctcacatt tcggataccc
5821 tactatgtag gcccccctcgc ccggggaaat tccagattcg cgtggatgac tcgcaaatca
5881 gaagagacca tcactccctg gaacttcgag gaagtcgtgg ataagggggc ctctgcccag
5941 tccttcatcg aaaggatgac taactttgat aaaaatctgc ctaacgaaaa ggtgcttcct
6001 aaacactctc tgctgtacga gtacttcaca gtttataacg agctcaccaa ggtcaaatac
6061 gtcacagaag ggatgagaaa gccagcattc ctgtctggag agcagaagaa agctatcgtg
6121 gacctcctct tcaagacgaa ccggaaagtt accgtgaaac agctcaaaga agactatttc
6181 aaaaagattg aatgtttcga ctctgttgaa atcagcggag tggaggatcg cttcaacgca
6241 tccctgggaa cgtatcacga tctcctgaaa atcattaaag acaaggactt cctggacaat
6301 gaggagaacg aggacattct tgaggacatt gtcctcaccc ttacgttgtt tgaagatagg
6361 gagatgattg aagaacgctt gaaaacttac gctcatctct tcgacgacaa agtcatgaaa
6421 cagctcaaga ggcgccgata tacaggatgg gggcggctgt caagaaaact gatcaatggg
6481 atccgagaca agcagagtgg aaagacaatc ctggattttc ttaagtccga tggatttgcc
6541 aaccggaact tcatgcagtt gatccatgat gactctctca cctttaagga ggacatccag
6601 aaagcacaag tttctggcca gggggacagt cttcacgagc acatcgctaa tcttgcaggt
6661 agcccagcta tcaaaaaggg aatactgcag accgttaagg tcgtggatga actcgtcaaa
6721 gtaatgggaa ggcataagcc cgagaatatc gttatcgaga tggcccgaga gaaccaaact
6781 acccagaagg gacagaagaa cagtagggaa aggatgaaga ggattgaaga gggtataaaa
6841 gaactggggt cccaaatcct taaggaacac ccagttgaaa acacccagct tcagaatgag
6901 aagctctacc tgtactacct gcagaacggc agggacatgt acgtggatca ggaactggac
6961 atcaatcggc tctccgacta cgacgtggat catatcgtgc cccagtcttt tctcaaagat
7021 gattctattg ataataaagt gttgacaaga tccgataaaa atagagggaa gagtgataac
7081 gtcccctcag aagaagttgt caagaaaatg aaaaattatt ggcggcagct gctgaacgcc

7141 aaactgatca cacaacggaa gttcgataat ctgactaagg ctgaacgagg tggcctgtct
7201 gagttggata aagccggctt catcaaaagg cagcttgttg agacacgcca gatcaccaag
7261 cacgtggccc aaattctcga ttcacgcatg aacaccaagt acgatgaaaa tgacaaactg
7321 attcgagagg tgaaagttat tactctgaag tctaagctgg tctcagattt cagaaaggac
7381 tttcagtttt ataaggtgag agagatcaac aattaccacc atgcgcatga tgcctacctg
7441 aatgcagtgg taggcactgc acttatcaaa aaatatccca agcttgaatc tgaatttgtt
7501 tacggagact ataaagtgta cgatgttagg aaaatgatcg caaagtctga gcaggaaata
7561 ggcaaggcca ccgctaagta cttcttttac agcaatatta tgaatttttt caagaccgag
7621 attacactgg ccaatggaga gattcggaag cgaccactta tcgaaacaaa cggagaaaca
7681 ggagaaatcg tgtgggacaa gggtagggat ttcgcgacag tccggaaggt cctgtccatg
7741 ccgcaggtga acatcgttaa aaagaccgaa gtacagaccg gaggcttctc caaggaaagt
7801 atcctcccga aaaggaacag cgacaagctg atcgcacgca aaaaagattg ggaccccaag
7861 aaatacggcg gattcgattc tcctacagtc gcttacagtg tactggttgt ggccaaagtg
7921 gagaaaggga agtctaaaaa actcaaaagc gtcaaggaac tgctgggcat cacaatcatg
7981 gagcgatcaa gcttcgaaaa aaaccccatc gactttctcg aggcgaaagg atataaagag
8041 gtcaaaaaag acctcatcat taagcttccc aagtactctc tctttgagct tgaaaacggc
8101 cggaaacgaa tgctcgctag tgcgggcgag ctgcagaaag gtaacgagct ggcactgccc
8161 tctaaatacg ttaatttctt gtatctggcc agccactatg aaaagctcaa agggtctccc
8221 gaagataatg agcagaagca gctgttcgtg gaacaacaca aacactacct tgatgagatc
8281 atcgagcaaa taagcgaatt ctccaaaaga gtgatcctcg ccgacgctaa cctcgataag
8341 gtgctttctg cttacaataa gcacagggat aagcccatca gggagcaggc agaaaacatt
8401 atccacttgt ttactctgac caacttgggc gcgcctgcag ccttcaagta cttcgacacc
8461 accatagaca gaaagcggta cacctctaca aaggaggtcc tggacgccac actgattcat
8521 cagtcaatta cggggctcta tgaaacaaga atcgacctct ctcagctcgg tggagacagc
8581 agggctgacc ccaagaagaa gaggaaggtg ggatctggag caacaaactt ctcactactc
8641 aaacaagcag gtgacgtgga ggagaatccc gggcctccca agaagaagag gaaggtgatg
8701 gacagcctct tgatgaaccg gaggaagttt ctttaccaat tcaaaaatgt ccgctgggct
8761 aagggtcggc gtgagaccta cctgtgctac gtagtgaaga ggcgtgacag tgctacatcc
8821 ttttcactgg actttggtta tcttcgcaat aagaacggct gccacgtgga attgctcttc
8881 ctccgctaca tctcggactg ggacctagac cctggccgct gctaccgcgt cacctggttc
8941 acctcctgga gcccctgcta cgactgtgcc cgacatgtgg ccgactttct gcgagggaac
9001 cccaacctca gtctgaggat cttcaccgcg cgcctctact tctgtgagga ccgcaaggct
9061 gagcccgagg ggctgcggcg gctgcaccgc gccggggtgc aaatagccat catgaccttc
9121 aaagattatt tttactgctg gaatactttt gtagaaaacc atgaaagaac tttcaaagcc
9181 tgggaagggc tgcatgaaaa ttcagttcgt ctctccagac agcttcggcg catccttttg
9241 cccctgtatg aggttgatga cttacgagac gcatttcgta ctttgggact tgaactgaaa
9301 actccacttg gagatactac acatacctca ccaccttgcc cagctcctga actgctgggt
9361 ggacctatgg cttctaactt tactcagttc gttctcgtcg acaatggcgg aactggcgac
9421 gtgactgtcg ccccaagcaa cttcgctaac gggatcgctg aatggatcag ctctaactcg
9481 cgttcacagg cttacaaagt aacctgtagc gttcgtcaga gctctgcgca gaatcgcaaa
9541 tacaccatca aagtcgaggt gcctaaaggc gcctggcgtt cgtacttaaa tatggaacta
9601 accattccaa ttttcgccac gaattccgac tgcgagctta ttgttaaggc aatgcaaggt
9661 ctcctaaaag atggaaaccc gattccctca gcaatcgcag caaactccgg catctacgct
9721 aggtgataaa gtgcggccgc gggagacacc ggttgataat agataacttc gtatagcata
9781 cattatacga agttatgaat tcgatatcaa gcttatcgat aatcaacctc tggattacaa
9841 aatttgtgaa agattgactg gtattcttaa ctatgttgct ccttttacgc tatgtggata
9901 cgctgcttta atgcctttgt atcatgctat tgcttcccgt atggctttca ttttctcctc
9961 cttgtataaa tcctggttgc tgtctcttta tgaggagttg tggcccgttg tcaggcaacg
10021 tggcgtggtg tgcactgtgt ttgctgacgc aacccccact ggttggggca ttgccaccac
10081 ctgtcagctc ctttccggga ctttcgcttt ccccctccct attgccacgg cggaactcat
10141 cgccgcctgc cttgcccgct gctggacagg ggctcggctg ttgggcactg acaattccgt
10201 ggtgttgtcg gggaaatcat cgtcctttcc ttggctgctc gcctgtgttg ccacctggat
10261 tctgcgcggg acgtccttct gctacgtccc ttcggccctc aatccagcgg accttccttc
10321 ccgcggcctg ctgccggctc tgcggcctct tccgcgtctt cgccttcgcc ctcagacgag
10381 tcggatctcc ctttgggccg cctccccgca tcgataccgt cgacctcgag acctagaaaa
10441 acatggagca atcacaagta gcaatacagc agctaccaat gctgattgtg cctggctaga
10501 agcacaagag gaggaggagg tgggttttcc agtcacacct caggtacctt taagaccaat
10561 gacttacaag gcagctgtag atcttagcca ctttttaaaa gaaaaggggg gactggaagg
10621 gctaattcac tcccaacgaa gacaagatat ccttgatctg tggatctacc acacacaagg
10681 ctacttccct gattggcaga actacacacc agggccaggg atcagatatc cactgacctt
10741 tggatggtgc tacaagctag taccagttga gcaagagaag gtagaagaag ccaatgaagg
10801 agagaacacc cgcttgttac accctgtgag cctgcatggg atggatgacc cggagagaga
10861 agtattagag tggaggtttg acagccgcct agcatttcat cacatggccc gagagctgca

10921 tccggactgt actgggtctc tctggttaga ccagatctga gcctgggagc tctctggcta
10981 actagggaac ccactgctta agcctcaata aagcttgcct tgagtgcttc aagtagtgtg
11041 tgcccgtctg ttgtgtgact ctggtaacta gagatccctc agaccctttt agtcagtgtg
11101 gaaaatctct agcagggccc gtttaaaccc gctgatcagc ctcgactgtg ccttctagtt
11161 gccagccatc tgttgtttgc ccctcccccg tgccttcctt gaccctggaa ggtgccactc
11221 ccactgtcct ttcctaataa aatgaggaaa ttgcatcgca ttgtctgagt aggtgtcatt
11281 ctattctggg gggtggggtg gggcaggaca gcaagggggа ggattgggaa gacaatagca
11341 ggcatgctgg ggatgcggtg ggctctatgg cttctgaggc ggaaagaacc agctggggct
11401 ctagggggta tccccacgcg ccctgtagcg gcgcattaag cgcggcgggt gtggtggtta
11461 cgcgcagcgt gaccgctaca cttgccagcg ccctagcgcc cgctcctttc gctttcttcc
11521 cttcctttct cgccacgttc gccggctttc cccgtcaagc tctaaatcgg gggctccctt
11581 tagggttccg atttagtgct ttacggcacc tcgaccccaa aaaacttgat tagggtgatg
11641 gttcacgtag tgggccatcg ccctgataga cggtttttcg ccctttgacg ttggagtcca
11701 cgttctttaa tagtggactc ttgttccaaa ctggaacaac actcaaccct atctcggtct
11761 attcttttga tttataaggg attttgccga tttcggccta ttggttaaaa aatgagctga
11821 tttaacaaaa atttaacgcg aattaattct gtggaatgtg tgtcagttag ggtgtggaaa
11881 gtccccaggc tccccagcag gcagaagtat gcaaagcatg catctcaatt agtcagcaac
11941 caggtgtgga aagtccccag gctccccagc aggcagaagt atgcaaagca tgcatctcaa
12001 ttagtcagca accatagtcc cgcccctaac tccgcccatc ccgcccctaa ctccgcccag
12061 ttccgcccat tctccgcccc atggctgact aatttttttt atttatgcag aggccgaggc
12121 cgcctctgcc tctgagctat tccagaagta gtgaggaggc ttttttggag gcctaggctt
12181 ttgcaaaaag ctcccgggag cttgtatatc cattttcgga tctgatcagc acgtgttgac
12241 aattaatcat cggcatagta tatcggcata gtataatacg acaaggtgag gaactaaacc
12301 atggccaagt tgaccagtgc cgttccggtg ctcaccgcgc gcgacgtcgc cggagcggtc
12361 gagttctgga ccgaccggct cgggttctcc cgggacttcg tggaggacga cttcgccggt
12421 gtggtccggg acgacgtgac cctgttcatc agcgcggtcc aggaccaggt ggtgccggac
12481 aacaccctgg cctgggtgtg ggtgcgcggc ctggacgagc tgtacgccga gtggtcggag
12541 gtcgtgtcca cgaacttccg ggacgcctcc gggccggcca tgaccgagat cggcgagcag
12601 ccgtgggggc gggagttcgc cctgcgcgac ccggccggca actgcgtgca cttcgtggcc
12661 gaggagcagg actgacacgt gctacgagat ttcgattcca ccgccgcctt ctatgaaagg
12721 ttgggcttcg gaatcgtttt ccgggacgcc ggctggatga tcctccagcg cggggatctc
12781 atgctggagt tcttcgccca ccccaacttg tttattgcag cttataatgg ttacaaataa
12841 agcaatagca tcacaaattt cacaaataaa gcatttttt cactgcattc tagttgtggt
12901 ttgtccaaac tcatcaatgt atcttatcat gtctgtatac cgtcgacctc tagctagagc
12961 ttggcgtaat catggtcata gctgtttcct gtgtgaaatt gttatccgct cacaattcca
13021 cacaacatac gagccggaag cataaagtgt aaagcctggg gtgcctaatg agtgagctaa
13081 ctcacattaa ttgcgttgcg ctcactgccc gctttccagt cgggaaacct gtcgtgccag
13141 ctgcattaat gaatcggcca acgcgcgggg agaggcggtt tgcgtattgg gcgctcttcc
13201 gcttcctcgc tcactgactc gctgcgctcg gtcgttcggc tgcggcgagc ggtatcagct
13261 cactcaaagg cggtaatacg gttatccaca gaatcagggg ataacgcagg aaagaacatg
13321 tgagcaaaag gccagcaaaa ggccaggaac cgtaaaaagg ccgcgttgct ggcgtttttc
13381 cataggctcc gcccccctga cgagcatcac aaaaatcgac gctcaagtca gaggtggcga
13441 aacccgacag gactataaag ataccaggcg tttccccctg gaagctccct cgtgcgctct
13501 cctgttccga ccctgccgct taccggatac ctgtccgcct ttctcccttc gggaagcgtg
13561 gcgctttctc atagctcacg ctgtaggtat ctcagttcgg tgtaggtcgt tcgctccaag
13621 ctgggctgtg tgcacgaacc ccccgttcag cccgaccgct gcgccttatc cggtaactat
13681 cgtcttgagt ccaacccggt aagacacgac ttatcgccac tggcagcagc cactggtaac
13741 aggattagca gagcgaggta tgtaggcggt gctacagagt tcttgaagtg gtggcctaac
13801 tacggctaca ctagaagaac agtatttggt atctgcgctc tgctgaagcc agttaccttc
13861 ggaaaaagag ttggtagctc ttgatccggc aaacaaacca ccgctggtag cggtggtttt
13921 tttgtttgca agcagcagat tacgcgcaga aaaaaaggat ctcaagaaga tcctttgatc
13981 ttttctacgg ggtctgacgc tcagtggaac gaaaactcac gttaagggat tttggtcatg
14041 agattatcaa aaaggatctt cacctagatc cttttaaatt aaaaatgaag ttttaaatca
14101 atctaaagta tatatgagta aacttggtct gacagttacc aatgcttaat cagtgaggca
14161 cctatctcag cgatctgtct atttcgttca tccatagttg cctgactccc cgtcgtgtag
14221 ataactacga tacgggaggg cttaccatct ggccccagtg ctgcaatgat accgcgagac
14281 ccacgctcac cggctccaga tttatcagca ataaaccagc cagccggaag ggccgagcgc
14341 agaagtggtc ctgcaacttt atccgcctcc atccagtcta ttaattgttg ccgggaagct
14401 agagtaagta gttcgccagt taatagtttg cgcaacgttg ttgccattgc tacaggcatc
14461 gtggtgtcac gctcgtcgtt tggtatggct tcattcagct ccggttccca acgatcaagg
14521 cgagttacat gatcccccat gttgtgcaaa aaagcggtta gctccttcgg tcctccgatc
14581 gttgtcagaa gtaagttggc cgcagtgtta tcactcatgg ttatggcagc actgcataat
14641 tctcttactg tcatgccatc cgtaagatgc ttttctgtga ctggtgagta ctcaaccaag

```
    14701 tcattctgag aatagtgtat gcggcgaccg agttgctctt gcccggcgtc aatacgggat
    14761 aataccgcgc cacatagcag aactttaaaa gtgctcatca ttggaaaacg ttcttcgggg
    14821 cgaaaactct caaggatctt accgctgttg agatccagtt cgatgtaacc cactcgtgca
    14881 cccaactgat cttcagcatc ttttactttc accagcgttt ctgggtgagc aaaaacagga
    14941 aggcaaaatg ccgcaaaaaa gggaataagg gcgacacgga aatgttgaat actcatactc
    15001 ttcctttttc aatattattg aagcatttat cagggttatt gtctcatgag cggatacata
    15061 tttgaatgta tttagaaaaa taaacaaata ggggttccgc gcacatttcc ccgaaaagtg
    15121 ccacctgac
//
```

3. pCMV_ACRCnu

```
LOCUS       Exported                9890 bp ds-DNA     circular SYN 24-AUG-2018
DEFINITION  synthetic circular DNA
ACCESSION   .
VERSION     .
KEYWORDS    .
SOURCE      synthetic DNA construct
  ORGANISM  synthetic DNA construct
REFERENCE   1  (bases 1 to 9890)
  AUTHORS   Juan-Carlos Collantes
  TITLE     Direct Submission
  JOURNAL   Exported Oct 8, 2018 from SnapGene Viewer 4.2.6
            http://www.snapgene.com
FEATURES             Location/Qualifiers
     source          1..9890
                     /organism="synthetic DNA construct"
                     /mol_type="other DNA"
     polyA_signal    417..551
                     /label=SV40 poly(A) signal
                     /note="SV40 polyadenylation signal"
     rep_origin      677..1132
                     /direction=RIGHT
                     /label=f1 ori
                     /note="f1 bacteriophage origin of replication; arrow
                     indicates direction of (+) strand synthesis"
     CDS             1420..2280
                     /codon_start=1
                     /gene="bla"
                     /product="beta-lactamase"
                     /label=AmpR
                     /note="confers resistance to ampicillin, carbenicillin,
and
                     related antibiotics"

/translation="MSIQHFRVALIPFFAAFCLPVFAHPETLVKVKDAEDQLGARVGYI

ELDLNSGKILESFRPEERFPMMSTFKVLLCGAVLSRIDAGQEQLGRRIHYSQNDLVEYS

PVTEKHLTDGMTVRELCSAAITMSDNTAANLLLTTIGGPKELTAFLHNMGDHVTRLDRW

EPELNEAIPNDERDTTMPVAMATTLRKLLTGELLTLASRQQLIDWMEADKVAGPLLRSA

LPAGWFIADKSGAGERGSRGIIAALGPDGKPSRIVVIYTTGSQATMDERNRQIAEIGAS
                     LIKHW"
     protein_bind    complement(2382..2415)
                     /label=loxP
                     /bound_moiety="Cre recombinase"
                     /note="Cre-mediated recombination occurs in the 8-bp core
```

```
                     sequence (GCATACAT)."
     rep_origin      2493..3081
                     /direction=RIGHT
                     /label=ori
                     /note="high-copy-number ColE1/pMB1/pBR322/pUC origin of
                     replication"
     enhancer        3541..3844
                     /label=CMV enhancer
                     /note="human cytomegalovirus immediate early enhancer"
     promoter        3845..4048
                     /label=CMV promoter
                     /note="human cytomegalovirus (CMV) immediate early
                     promoter"
     misc_feature    4059..5069
                     /label=Seq
     primer_bind     4174..4190
                     /label=M13 rev
                     /note="common sequencing primer, one of multiple similar
                     variants"
     promoter        4201..4219
                     /label=SP6 promoter
                     /note="promoter for bacteriophage SP6 RNA polymerase"
     protein_bind    4220..4244
                     /gene="mutant version of attB"
                     /label=attB1
                     /bound_moiety="BP Clonase(TM)"
                     /note="recombination site for the Gateway(R) BP reaction"
     primer_bind     4280..4326
                     /label=XbaI_AID F
     primer_bind     4280..4298
                     /label=XbaI_AID short F
     regulatory      4290..4299
                     /regulatory_class="other"
                     /note="vertebrate consensus sequence for strong initiation
                     of translation (Kozak, 1987)"
     CDS             4302..4322
                     /codon_start=1
                     /product="nuclear localization signal of SV40 large T
                     antigen"
                     /label=SV40 NLS
                     /translation="PKKKRKV"
     primer_bind     complement(4303..4321)
                     /label=PspOMI_SNL-UGI R
     CDS             4323..4916
                     /codon_start=1
                     /label=AID

/translation="MDSLLMNRRKFLYQFKNVRWAKGRRETYLCYVVKRRDSATSFSLD

FGYLRNKNGCHVELLFLRYISDWDLDPGRCYRVTWFTSWSPCYDCARHVADFLRGNPNL

SLRIFTARLYFCEDRKAEPEGLRRLHRAGVQIAIMTFKDYFYCWNTFVENHERTFKAWE
                     GLHENSVRLSRQLRRILLPLYEVDDLRDAFRTLGL"
     CDS             4917..4991
                     /codon_start=1
                     /label=L25
                     /translation="ELKTPLGDTTHTSPPCPAPELLGGP"
     CDS             4992..5342
                     /codon_start=1
                     /label=MCP

/translation="MASNFTQFVLVDNGGTGDVTVAPSNFANGIAEWISSNSRSQAYKV
```

```
TCSVRQSSAQNRKYTIKVEVPKGAWRSYLNMELTIPIFATNSDCELIVKAMQGLLKDGN
                     PIPSAIAANSGIY"
     CDS             5352..5405
                     /codon_start=1
                     /product="2A peptide from Thosea asigna virus capsid
                     protein"
                     /label=T2A
                     /note="Eukaryotic ribosomes fail to insert a peptide bond
                     between the Gly and Pro residues, yielding separate
                     polypeptides."
                     /translation="EGRGSLLTCGDVEENPGP"
     primer_bind     5365..5436
                     /label=T2A_Cas9-UGI F
     primer_bind     complement(5370..5404)
                     /label=T2A R
     CDS             5412..9512
                     /codon_start=1
                     /product="nickase mutant of the Cas9 endonuclease from the
                     Streptococcus pyogenes Type II CRISPR/Cas system"
                     /label=Cas9(D10A)
                     /note="generates RNA-guided single strand nicks in DNA due
                     to the D10A mutation in the RuvC catalytic domain"

/translation="DKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKN

LIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESF

LVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKF

RGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLE

NLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQI

GDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQ

QLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLR

KQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGN

SRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYF

TVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDS

VEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERL

KTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQ

LIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRH

KPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYL

YYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPS

EEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHV

AQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLN

AVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTE

ITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKE

SILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGIT
```

```
IMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNEL

ALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADA

NLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLD
                     ATLIHQSITGLYETRIDLSQLGGD"
     primer_bind     complement(5842..5880)
                     /label=Bsu36I_Cas9-UGI R
     primer_bind     complement(5855..5880)
                     /label=Bsu36I_Cas9-UGI Short R
     misc_feature    8762..9809
                     /label=Seq
     CDS             9525..9773
                     /codon_start=1
                     /label=UGI

/translation="TNLSDIIEKETGKQLVIQESILMLPEEVEEVIGNKPESDILVHTA
                     YDESTDENVMLLTSDAPEYKPWALVIQDSNGENKIKML"
     CDS             9786..9806
                     /codon_start=1
                     /product="nuclear localization signal of SV40 large T
                     antigen"
                     /label=SV40 NLS
                     /translation="PKKKRKV"
     primer_bind     complement(9787..9826)
                     /label=PspOMI_SNL-UGI R
     protein_bind    complement(9831..9855)
                     /gene="mutant version of attB"
                     /label=attB2
                     /bound_moiety="BP Clonase(TM)"
                     /note="recombination site for the Gateway(R) BP reaction"
     promoter        complement(9857..9875)
                     /label=T7 promoter
                     /note="promoter for bacteriophage T7 RNA polymerase"
     primer_bind     complement(join(9886..9890,1..12))
                     /label=M13 fwd
                     /note="common sequencing primer, one of multiple similar
                     variants"
ORIGIN
        1 ccgtcgtttt acaacgtcgt gactgggaaa actgctagct tgggatcttt gtgaaggaac
       61 cttacttctg tggtgtgaca taattggaca aactacctac agagatttaa agctctaagg
      121 taaatataaa atttttaagt gtataatgtg ttaaactagc tgcatatgct tgctgcttga
      181 gagttttgct tactgagtat gatttatgaa aatattatac acaggagcta gtgattctaa
      241 ttgtttgtgt attttagatt cacagtccca aggctcattt caggcccctc agtcctcaca
      301 gtctgttcat gatcataatc agccatacca catttgtaga ggttttactt gctttaaaaa
      361 acctcccaca cctccccctg aacctgaaac ataaaatgaa tgcaattgtt gttgttaact
      421 tgtttattgc agcttataat ggttacaaat aaagcaatag catcacaaat ttcacaaata
      481 aagcattttt ttcactgcat tctagttgtg gtttgtccaa actcatcaat gtatcttatc
      541 atgtctggat cgatcctgca ttaatgaatc ggccaacgcg cggggagagg cggtttgcgt
      601 attggctggc gtaatagcga agaggcccgc accgatcgcc cttcccaaca gttgcgcagc
      661 ctgaatggcg aatgggacgc gccctgtagc ggcgcattaa gcgcggcggg tgtggtggtt
      721 acgcgcagcg tgaccgctac acttgccagc gccctagcgc ccgctccttt cgctttcttc
      781 ccttcctttc tcgccacgtt cgccggcttt ccccgtcaag ctctaaatcg ggggctccct
      841 ttagggttcc gatttagtgc tttacggcac ctcgacccca aaaaacttga ttagggtgat
      901 ggttcacgta gtgggccatc gccctgatag acggtttttc gccctttgac gttggagtcc
      961 acgttcttta atagtggact cttgttccaa actggaacaa cactcaaccc tatctcggtc
     1021 tattcttttg atttataagg gattttgccg atttcggcct attggttaaa aaatgagctg
     1081 atttaacaaa aatttaacgc gaattttaac aaaatattaa cgtttacaat ttcaggtggc
     1141 acttttcggg gaaatgtgcg cggaacccct atttgtttat ttttctaaat acattcaaat
     1201 atgtatccgc tcatgccagg tcttggactg gtgagaacgg cttgctcggc agcttcgatg
     1261 tgtgctggag ggagaataaa ggtctaagat gtgcgataga gggaagtcgc attgaattat
     1321 gtgctgtgta gggatcgctg gtatcaaata tgtgtgccca cccctggcat gagacaataa
```

1381 ccctgataaa tgcttcaata atattgaaaa aggaagagta tgagtattca acatttccgt
1441 gtcgccctta ttccctttt tgcggcattt tgccttcctg tttttgctca cccagaaacg
1501 ctggtgaaag taaaagatgc tgaagatcag ttgggtgcac gagtgggtta catcgaactg
1561 gatctcaaca gcggtaagat ccttgagagt tttcgccccg aagaacgttt tccaatgatg
1621 agcactttta aagttctgct atgtggcgcg gtattatccc gtattgacgc cgggcaagag
1681 caactcggtc gccgcataca ctattctcag aatgacttgg ttgagtactc accagtcaca
1741 gaaaagcatc ttacggatgg catgacagta agagaattat gcagtgctgc cataaccatg
1801 agtgataaca ctgcggccaa cttacttctg acaacgatcg gaggaccgaa ggagctaacc
1861 gcttttttgc acaacatggg ggatcatgta actcgccttg atcgttggga accggagctg
1921 aatgaagcca taccaaacga cgagcgtgac accacgatgc ctgtagcaat ggcaacaacg
1981 ttgcgcaaac tattaactgg cgaactactt actctagctt cccggcaaca attaatagac
2041 tggatggagg cggataaagt tgcaggacca cttctgcgct cggcccttcc ggctggctgg
2101 tttattgctg ataaatctgg agccggtgag cgtgggtctc gcggtatcat tgcagcactg
2161 gggccagatg gtaagccctc ccgtatcgta gttatctaca cgacggggag tcaggcaact
2221 atggatgaac gaaatagaca gatcgctgag ataggtgcct cactgattaa gcattggtaa
2281 ctgtcagacc aagtttactc atatatactt tagattgatt taaaacttca tttttaattt
2341 aaaaggatct aggtgaagat cctttttgat aatctcatgc cataacttcg tataatgtat
2401 gctatacgaa gttatggcat gaccaaaatc ccttaacgtg agttttcgtt ccactgagcg
2461 tcagaccccg tagaaaagat caaaggatct tcttgagatc ctttttttct gcgcgtaatc
2521 tgctgcttgc aaacaaaaaa accaccgcta ccagcggtgg tttgtttgcc ggatcaagag
2581 ctaccaactc tttttccgaa ggtaactggc ttcagcagag cgcagatacc aaatactgtc
2641 cttctagtgt agccgtagtt aggccaccac ttcaagaact ctgtagcacc gcctacatac
2701 ctcgctctgc taatcctgtt accagtggct gctgccagtg gcgataagtc gtgtcttacc
2761 gggttggact caagacgata gttaccggat aaggcgcagc ggtcgggctg aacggggggt
2821 tcgtgcacac agcccagctt ggagcgaacg acctacaccg aactgagata cctacagcgt
2881 gagcattgag aaagcgccac gcttcccgaa gggagaaagg cggacaggta tccggtaagc
2941 ggcagggtcg gaacaggaga gcgcacgagg gagcttccag ggggaaacgc ctggtatctt
3001 tatagtcctg tcgggtttcg ccacctctga cttgagcgtc gatttttgtg atgctcgtca
3061 gggggggcgga gcctatggaa aaacgccagc aacgcggcct ttttacggtt cctggccttt
3121 tgctggcctt ttgctcacat gttctttcct gcgttatccc ctgattctgt ggataaccgt
3181 attaccgcct ttgagtgagc tgataccgct cgccgcagcc gaacgaccga gcgcagcgag
3241 tcagtgagcg aggaagcgga agagcgccca atacgcaaac cgcctctccc cgcgcgttgg
3301 ccgattcatt aatgcagagc ttgcaattcg cgcgtttttc aatattattg aagcatttat
3361 cagggttatt gtctcatgag cggatacata tttgaatgta tttagaaaaa taaacaaata
3421 ggggttccgc gcacatttcc ccgaaaagtg ccacctgacg tctaagaaac cattattatc
3481 atgacattaa cctataaaaa taggcgtagt acgaggccct ttcactcatt agatgcatgt
3541 cgttacataa cttacggtaa atggcccgcc tggctgaccg cccaacgacc cccgcccatt
3601 gacgtcaata atgacgtatg ttcccatagt aacgccaata gggactttcc attgacgtca
3661 atgggtggag tatttacggt aaactgccca cttggcagta catcaagtgt atcatatgcc
3721 aagtacgccc cctattgacg tcaatgacgg taaatggccc gcctggcatt atgcccagta
3781 catgacctta tgggactttc ctacttggca gtacatctac gtattagtca tcgctattac
3841 catggtgatg cggttttggc agtacatcaa tgggcgtgga tagcggtttg actcacgggg
3901 atttccaagt ctccacccca ttgacgtcaa tgggagtttg ttttggcacc aaaatcaacg
3961 ggactttcca aaatgtcgta acaactccgc cccattgacg caaatgggcg gtaggcgtgt
4021 acggtgggag gtctatataa gcagagctcg tttagtgaac cgtcagatcg cctggagacg
4081 ccatccacgc tgttttgacc tccatagaag acaccgggac cgatccagcc tccggactct
4141 agcctaggcc gcggagcgga taacaatttc acacaggaaa cagctatgac cattaggcct
4201 atttaggtga cactatagaa caagtttgta caaaaaagca ggctggtacc ggtccggaat
4261 tcccgggata tcgtcgaggc cgctctagag ccaccatggc ccccaagaag aagaggaagg
4321 tgatggacag cctcttgatg aaccggagga agtttcttta ccaattcaaa aatgtccgct
4381 gggctaaggg tcggcgtgag acctacctgt gctacgtagt gaagaggcgt gacagtgcta
4441 catccttttc actggacttt ggttatcttc gcaataagaa cggctgccac gtggaattgc
4501 tcttcctccg ctacatctcg gactgggacc tagaccctgg ccgctgctac cgcgtcacct
4561 ggttcacctc ctggagcccc tgctacgact gtgcccgaca tgtggccgac tttctgcgag
4621 ggaaccccaa cctcagtctg aggatcttca ccgcgcgcct ctacttctgt gaggaccgca
4681 aggctgagcc cgaggggctg cggcggctgc accgcgccgg ggtgcaaata gccatcatga
4741 ccttcaaaga ttatttttac tgctggaata cttttgtaga aaaccatgaa agaactttca
4801 aagcctggga agggctgcat gaaaattcag ttcgtctctc cagacagctt cggcgcatcc
4861 ttttgcccct gtatgaggtt gatgacttac gagacgcatt tcgtactttg ggacttgaac
4921 tgaaaactcc acttggagat actacacata cctcaccacc ttgcccagct cctgaactgc
4981 tgggtggacc tatggcttct aactttactc agttcgttct cgtcgacaat ggcggaactg
5041 gcgacgtgac tgtcgcccca agcaacttcg ctaacgggat cgctgaatgg atcagctcta
5101 actcgcgttc acaggcttac aaagtaacct gtagcgttcg tcagagctct gcgcagaatc

5161 gcaaatacac catcaaagtc gaggtgccta aaggcgcctg gcgttcgtac ttaaatatgg
5221 aactaaccat tccaattttc gccacgaatt ccgactgcga gcttattgtt aaggcaatgc
5281 aaggtctcct aaaagatgga aacccgattc cctcagcaat cgcagcaaac tccggcatct
5341 acggatctgg agagggcaga ggaagtcttc taacatgcgg tgacgtggag gagaatcccg
5401 gccctggtac cgataaaaag tattctattg gtttagccat cggcactaat tccgttggat
5461 gggctgtcat aaccgatgaa tacaaagtac cttcaaagaa atttaaggtg ttggggaaca
5521 cagaccgtca ttcgattaaa aagaatctta tcggtgccct cctattcgat agtggcgaaa
5581 cggcagaggc gactcgcctg aaacgaaccg ctcggagaag gtatacacgt cgcaagaacc
5641 gaatatgtta cttacaagaa atttttagca atgagatggc caaagttgac gattctttct
5701 ttcaccgttt ggaagagtcc ttccttgtcg aagaggacaa gaaacatgaa cggcacccca
5761 tctttggaaa catagtagat gaggtggcat atcatgaaaa gtacccaacg atttatcacc
5821 tcagaaaaaa gctagttgac tcaactgata aagcggacct gaggttaatc tacttggctc
5881 ttgcccatat gataaagttc cgtgggcact ttctcattga gggtgatcta aatccggaca
5941 actcggatgt cgacaaactg ttcatccagt tagtacaaac ctataatcag ttgtttgaag
6001 agaaccctat aaatgcaagt ggcgtggatg cgaaggctat tcttagcgcc cgcctctcta
6061 aatcccgacg gctagaaaac ctgatcgcac aattacccgg agagaagaaa aatgggttgt
6121 tcggtaacct tatagcgctc tcactaggcc tgacaccaaa ttttaagtcg aacttcgact
6181 tagctgaaga tgccaaattg cagcttagta aggacacgta cgatgacgat ctcgacaatc
6241 tactggcaca aattggagat cagtatgcgg acttattttt ggctgccaaa aaccttagcg
6301 atgcaatcct cctatctgac atactgagag ttaatactga gattaccaag gcgccgttat
6361 ccgcttcaat gatcaaaagg tacgatgaac atcaccaaga cttgacactt ctcaaggccc
6421 tagtccgtca gcaactgcct gagaaatata aggaaatatt ctttgatcag tcgaaaaacg
6481 ggtacgcagg ttatattgac ggcggagcga gtcaagagga attctacaag tttatcaaac
6541 ccatattaga gaagatggat gggacggaag agttgcttgt aaaactcaat cgcgaagatc
6601 tactgcgaaa gcagcggact ttcgacaacg gtagcattcc acatcaaatc cacttaggcg
6661 aattgcatgc tatacttaga aggcaggagg atttttatcc gttcctcaaa gacaatcgtg
6721 aaaagattga gaaaatccta acctttcgca taccttacta tgtgggaccc ctggcccgag
6781 ggaactctcg gttcgcatgg atgacaagaa agtccgaaga aacgattact ccatggaatt
6841 ttgaggaagt tgtcgataaa ggtgcgtcag ctcaatcgtt catcgagagg atgaccaact
6901 ttgacaagaa tttaccgaac gaaaaagtat tgcctaagca cagtttactt tacgagtatt
6961 tcacagtgta caatgaactc acgaaagtta agtatgtcac tgagggcatg cgtaaacccg
7021 cctttctaag cggagaacag aagaaagcaa tagtagatct gttattcaag accaaccgca
7081 aagtgacagt taagcaattg aaagaggact actttaagaa aattgaatgc ttcgattctg
7141 tcgagatctc cggggtagaa gatcgattta atgcgtcact tggtacgtat catgacctcc
7201 taaagataat taaagataag gacttcctgg ataacgaaga gaatgaagat atcttagaag
7261 atatagtgtt gactcttacc ctctttgaag atcgggaaat gattgaggaa agactaaaaa
7321 catacgctca cctgttcgac gataaggtta tgaaacagtt aaagaggcgt cgctatacgg
7381 gctggggacg attgtcgcgg aaacttatca acgggataag agacaagcaa agtggtaaaa
7441 ctattctcga ttttctaaag agcgacggct tcgccaatag gaactttatg cagctgatcc
7501 atgatgactc tttaaccttc aaagaggata tacaaaaggc acaggtttcc ggacaagggg
7561 actcattgca cgaacatatt gcgaatcttg ctggttcgcc agccatcaaa aagggcatac
7621 tccagacagt caaagtagtg gatgagctag ttaaggtcat gggacgtcac aaaccggaaa
7681 acattgtaat cgagatggca cgcgaaaatc aaacgactca gaaggggcaa aaaaacagtc
7741 gagagcggat gaagagaata gaagagggta ttaaagaact gggcagccag atcttaaagg
7801 agcatcctgt ggaaaatacc caattgcaga acgagaaact ttacctctat tacctacaaa
7861 atggaaggga catgtatgtt gatcaggaac tggacataaa ccgtttatct gattacgacg
7921 tcgatcacat tgtaccccaa tccttttga aggacgattc aatcgacaat aaagtgctta
7981 cacgctcgga taagaaccga gggaaaagtg acaatgttcc aagcgaggaa gtcgtaaaga
8041 aaatgaagaa ctattggcgg cagctcctaa atgcgaaact gataacgcaa agaaagttcg
8101 ataacttaac taaagctgag aggggtggct tgtctgaact tgacaaggcc ggatttatta
8161 aacgtcagct cgtggaaacc cgccaaatca caaagcatgt tgcacagata ctagattccc
8221 gaatgaatac gaaatacgac gagaacgata agctgattcg ggaagtcaaa gtaatcactt
8281 taaagtcaaa attggtgtcg gacttcagaa aggattttca attctataaa gttagggaga
8341 taaataacta ccaccatgcg cacgacgctt atcttaatgc cgtcgtaggg accgcactca
8401 ttaagaaata cccgaagcta gaaagtgagt ttgtgtatgg tgattacaaa gtttatgacg
8461 tccgtaagat gatcgcgaaa agcgaacagg agataggcaa ggctacagcc aaatacttct
8521 tttattctaa cattatgaat ttctttaaga cggaaatcac tctggcaaac ggagagatac
8581 gcaaacgacc tttaattgaa accaatgggg agacaggtga aatcgtatgg gataagggcc
8641 gggacttcgc gacggtgaga aaagttttgt ccatgcccca agtcaacata gtaaagaaaa
8701 ctgaggtgca gaccggaggg ttttcaaagg aatcgattct tccaaaaagg aatagtgata
8761 agctcatcgc tcgtaaaaag gactgggacc cgaaaaagta cggtggcttc gatagcccta
8821 cagttgccta ttctgtccta gtagtggcaa aagttgagaa gggaaaatcc aagaaactga
8881 agtcagtcaa agaattattg gggataacga ttatggagcg ctcgtctttt gaaaagaacc

```
     8941 ccatcgactt ccttgaggcg aaaggttaca aggaagtaaa aaaggatctc ataattaaac
     9001 taccaaagta tagtctgttt gagttagaaa atggccgaaa acggatgttg gctagcgccg
     9061 gagagcttca aaaggggaac gaactcgcac taccgtctaa atacgtgaat ttcctgtatt
     9121 tagcgtccca ttacgagaag ttgaaaggtt cacctgaaga taacgaacag aagcaacttt
     9181 ttgttgagca gcacaaacat tatctcgacg aaatcataga gcaaatttcg gaattcagta
     9241 agagagtcat cctagctgat gccaatctgg acaaagtatt aagcgcatac aacaagcaca
     9301 gggataaacc catacgtgag caggcggaaa atattatcca tttgtttact cttaccaacc
     9361 tcggcgctcc agccgcattc aagtattttg acacaacgat agatcgcaaa cgatacactt
     9421 ctaccaagga ggtgctagac gcgacactga ttcaccaatc catcacggga ttatatgaaa
     9481 ctcggataga tttgtcacag cttgggggtg actctggtgg ttctactaat ctgtcagata
     9541 ttattgaaaa ggagaccggt aagcaactgg ttatccagga atccatcctc atgctcccag
     9601 aggaggtgga agaagtcatt gggaacaagc cggaaagcga tatactcgtg cacaccgcct
     9661 acgacgagag caccgacgag aatgtcatgc ttctgactag cgacgcccct gaatacaagc
     9721 cttgggctct ggtcatacag gatagcaacg gtgagaacaa gattaagatg ctctctggtg
     9781 gttctcccaa gaagaagagg aaagtctaat gagggcccaa gcttacgcgt acccagcttt
     9841 cttgtacaaa gtggtcccta tagtgagtcg tattataagc taggcactgg
//
```

4. pCMV_ACRCvnu

```
LOCUS       Exported                9890 bp ds-DNA     circular SYN 02-OCT-2018
DEFINITION  pCMV_ACRCvnu synthetic circular DNA
ACCESSION   .
VERSION     .
KEYWORDS    .
SOURCE      synthetic DNA construct
  ORGANISM  synthetic DNA construct
REFERENCE   1  (bases 1 to 9890)
  AUTHORS   Juan-Carlos Collantes
  TITLE     Direct Submission
  JOURNAL   Exported Oct 8, 2018 from SnapGene Viewer 4.2.6
            http://www.snapgene.com
FEATURES             Location/Qualifiers
     source          1..9890
                     /organism="synthetic DNA construct"
                     /mol_type="other DNA"
     polyA_signal    417..551
                     /label=SV40 poly(A) signal
                     /note="SV40 polyadenylation signal"
     rep_origin      677..1132
                     /direction=RIGHT
                     /label=f1 ori
                     /note="f1 bacteriophage origin of replication; arrow
                     indicates direction of (+) strand synthesis"
     CDS             1420..2280
                     /codon_start=1
                     /gene="bla"
                     /product="beta-lactamase"
                     /label=AmpR
                     /note="confers resistance to ampicillin, carbenicillin,
and
                     related antibiotics"

/translation="MSIQHFRVALIPFFAAFCLPVFAHPETLVKVKDAEDQLGARVGYI

ELDLNSGKILESFRPEERFPMMSTFKVLLCGAVLSRIDAGQEQLGRRIHYSQNDLVEYS

PVTEKHLTDGMTVRELCSAAITMSDNTAANLLLTTIGGPKELTAFLHNMGDHVTRLDRW
```

```
EPELNEAIPNDERDTTMPVAMATTLRKLLTGELLTLASRQQLIDWMEADKVAGPLLRSA

LPAGWFIADKSGAGERGSRGIIAALGPDGKPSRIVVIYTTGSQATMDERNRQIAEIGAS
                     LIKHW"
     protein_bind    complement(2382..2415)
                     /label=loxP
                     /bound_moiety="Cre recombinase"
                     /note="Cre-mediated recombination occurs in the 8-bp core
                     sequence (GCATACAT)."
     rep_origin      2493..3081
                     /direction=RIGHT
                     /label=ori
                     /note="high-copy-number ColE1/pMB1/pBR322/pUC origin of
                     replication"
     enhancer        3541..3844
                     /label=CMV enhancer
                     /note="human cytomegalovirus immediate early enhancer"
     promoter        3845..4048
                     /label=CMV promoter
                     /note="human cytomegalovirus (CMV) immediate early
                     promoter"
     primer_bind     4174..4190
                     /label=M13 rev
                     /note="common sequencing primer, one of multiple similar
                     variants"
     promoter        4201..4219
                     /label=SP6 promoter
                     /note="promoter for bacteriophage SP6 RNA polymerase"
     protein_bind    4220..4244
                     /gene="mutant version of attB"
                     /label=attB1
                     /bound_moiety="BP Clonase(TM)"
                     /note="recombination site for the Gateway(R) BP reaction"
     regulatory      4290..4299
                     /regulatory_class="other"
                     /note="vertebrate consensus sequence for strong initiation
                     of translation (Kozak, 1987)"
     CDS             4302..4322
                     /codon_start=1
                     /product="nuclear localization signal of SV40 (simian
virus
                     40) large T antigen"
                     /label=SV40 NLS
                     /translation="PKKKRKV"
     CDS             4323..4916
                     /codon_start=1
                     /gene="AICDA"
                     /product="human activation-induced cytidine deaminase
                     (Dickerson et al., 2003)"
                     /label=AID

/translation="MDSLLMNRRKFLYQFKNVRWAKGRRETYLCYVVKRRDSATSFSLD

FGYLRNKNGCHVELLFLRYISDWDLDPGRCYRVTWFTSWSPCYDCARHVADFLRGNPNL

SLRIFTARLYFCEDRKAEPEGLRRLHRAGVQIAIMTFKDYFYCWNTFVENHERTFKAWE
                     GLHENSVRLSRQLRRILLPLYEVDDLRDAFRTLGL"
     CDS             4917..4991
                     /codon_start=1
                     /label=L25
                     /translation="ELKTPLGDTTHTSPPCPAPELLGGP"
     CDS             4992..5342
```

```
                     /codon_start=1
                     /label=MCP

/translation="MASNFTQFVLVDNGGTGDVTVAPSNFANGIAEWISSNSRSQAYKV

TCSVRQSSAQNRKYTIKVEVPKGAWRSYLNMELTIPIFATNSDCELIVKAMQGLLKDGN
                     PIPSAIAANSGIY"
     CDS             5352..5405
                     /codon_start=1
                     /product="2A peptide from Thosea asigna virus capsid
                     protein"
                     /label=T2A
                     /note="Eukaryotic ribosomes fail to insert a peptide bond
                     between the Gly and Pro residues, yielding separate
                     polypeptides."
                     /translation="EGRGSLLTCGDVEENPGP"
     CDS             5412..9512
                     /codon_start=1
                     /product="Streptococcus pyogenes Cas9 endonuclease with
the
                     D1135V/R1335Q/T1337R mutations"
                     /label=nCas9 VQR
                     /note="has an altered recognition specificity for the
                     protospacer adjacent motif (PAM) (Kleinstiver et al.,
                     2015)"

/translation="DKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKN

LIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESF

LVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKF

RGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLE

NLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQI

GDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQ

QLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLR

KQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGN

SRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYF

TVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDS

VEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERL

KTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQ

LIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRH

KPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYL

YYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPS

EEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHV

AQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLN

AVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTE

ITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKE
```

```
SILPKRNSDKLIARKKDWDPKKYGGFVSPTVAYSVLVVAKVEKGKSKKLKSVKELLGIT

IMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNEL

ALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADA

NLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKQYRSTKEVLD
                     ATLIHQSITGLYETRIDLSQLGGD"
     misc_feature    8407..9782
                     /label=seq clone 2
     CDS             9525..9773
                     /codon_start=1
                     /product="uracil-DNA glycosylase inhibitor from a Bacillus
                     subtilis bacteriophage (Mol et al., 1995)"
                     /label=UGI

/translation="TNLSDIIEKETGKQLVIQESILMLPEEVEEVIGNKPESDILVHTA
                     YDESTDENVMLLTSDAPEYKPWALVIQDSNGENKIKML"
     CDS             9786..9806
                     /codon_start=1
                     /product="nuclear localization signal of SV40 (simian
virus
                     40) large T antigen"
                     /label=SV40 NLS
                     /translation="PKKKRKV"
     protein_bind    complement(9831..9855)
                     /gene="mutant version of attB"
                     /label=attB2
                     /bound_moiety="BP Clonase(TM)"
                     /note="recombination site for the Gateway(R) BP reaction"
     promoter        complement(9857..9875)
                     /label=T7 promoter
                     /note="promoter for bacteriophage T7 RNA polymerase"
ORIGIN
        1 ccgtcgtttt acaacgtcgt gactgggaaa actgctagct tgggatcttt gtgaaggaac
       61 cttacttctg tggtgtgaca taattggaca aactacctac agagatttaa agctctaagg
      121 taaatataaa atttttaagt gtataatgtg ttaaactagc tgcatatgct tgctgcttga
      181 gagttttgct tactgagtat gatttatgaa aatattatac acaggagcta gtgattctaa
      241 ttgtttgtgt attttagatt cacagtccca aggctcattt caggcccctc agtcctcaca
      301 gtctgttcat gatcataatc agccatacca catttgtaga ggttttactt gctttaaaaa
      361 acctcccaca cctccccctg aacctgaaac ataaaatgaa tgcaattgtt gttgttaact
      421 tgtttattgc agcttataat ggttacaaat aaagcaatag catcacaaat ttcacaaata
      481 aagcattttt ttcactgcat tctagttgtg gtttgtccaa actcatcaat gtatcttatc
      541 atgtctggat cgatcctgca ttaatgaatc ggccaacgcg cggggagagg cggtttgcgt
      601 attggctggc gtaatagcga agaggcccgc accgatcgcc cttcccaaca gttgcgcagc
      661 ctgaatggcg aatgggacgc gccctgtagc ggcgcattaa gcgcggcggg tgtggtggtt
      721 acgcgcagcg tgaccgctac acttgccagc gccctagcgc ccgctccttt cgctttcttc
      781 ccttcctttc tcgccacgtt cgccggcttt ccccgtcaag ctctaaatcg ggggctccct
      841 ttagggttcc gatttagtgc tttacggcac ctcgacccca aaaaacttga ttagggtgat
      901 ggttcacgta gtgggccatc gccctgatag acggtttttc gccctttgac gttggagtcc
      961 acgttcttta atagtggact cttgttccaa actggaacaa cactcaaccc tatctcggtc
     1021 tattcttttg atttataagg gattttgccg atttcggcct attggttaaa aaatgagctg
     1081 atttaacaaa aatttaacgc gaattttaac aaaatattaa cgtttacaat ttcaggtggc
     1141 acttttcggg gaaatgtgcg cggaacccct atttgtttat ttttctaaat acattcaaat
     1201 atgtatccgc tcatgccagg tcttggactg gtgagaacgg cttgctcggc agcttcgatg
     1261 tgtgctggag ggagaataaa ggtctaagat gtgcgataga gggaagtcgc attgaattat
     1321 gtgctgtgta gggatcgctg gtatcaaata tgtgtgccca cccctggcat gagacaataa
     1381 ccctgataaa tgcttcaata atattgaaaa aggaagagta tgagtattca acatttccgt
     1441 gtcgccctta ttccctttt tgcggcattt tgccttcctg tttttgctca cccagaaacg
     1501 ctggtgaaag taaaagatgc tgaagatcag ttgggtgcac gagtgggtta catcgaactg
     1561 gatctcaaca gcggtaagat ccttgagagt tttcgccccg aagaacgttt tccaatgatg
     1621 agcactttta aagttctgct atgtggcgcg gtattatccc gtattgacgc cgggcaagag
```

1681 caactcggtc gccgcataca ctattctcag aatgacttgg ttgagtactc accagtcaca
1741 gaaaagcatc ttacggatgg catgacagta agagaattat gcagtgctgc cataaccatg
1801 agtgataaca ctgcggccaa cttacttctg acaacgatcg gaggaccgaa ggagctaacc
1861 gctttttgc acaacatggg ggatcatgta actcgccttg atcgttggga accggagctg
1921 aatgaagcca taccaaacga cgagcgtgac accacgatgc ctgtagcaat ggcaacaacg
1981 ttgcgcaaac tattaactgg cgaactactt actctagctt cccggcaaca attaatagac
2041 tggatggagg cggataaagt tgcaggacca cttctgcgct cggcccttcc ggctggctgg
2101 tttattgctg ataaatctgg agccggtgag cgtgggtctc gcggtatcat tgcagcactg
2161 gggccagatg gtaagccctc ccgtatcgta gttatctaca cgacggggag tcaggcaact
2221 atggatgaac gaaatagaca gatcgctgag ataggtgcct cactgattaa gcattggtaa
2281 ctgtcagacc aagtttactc atatatactt tagattgatt taaaacttca tttttaattt
2341 aaaaggatct aggtgaagat cctttttgat aatctcatgc cataacttcg tataatgtat
2401 gctatacgaa gttatggcat gaccaaaatc ccttaacgtg agttttcgtt ccactgagcg
2461 tcagaccccg tagaaaagat caaaggatct tcttgagatc ctttttttct gcgcgtaatc
2521 tgctgcttgc aaacaaaaaa accaccgcta ccagcggtgg tttgtttgcc ggatcaagag
2581 ctaccaactc tttttccgaa ggtaactggc ttcagcagag cgcagatacc aaatactgtc
2641 cttctagtgt agccgtagtt aggccaccac ttcaagaact ctgtagcacc gcctacatac
2701 ctcgctctgc taatcctgtt accagtggct gctgccagtg gcgataagtc gtgtcttacc
2761 gggttggact caagacgata gttaccggat aaggcgcagc ggtcgggctg aacggggggt
2821 tcgtgcacac agcccagctt ggagcgaacg acctacaccg aactgagata cctacagcgt
2881 gagcattgag aaagcgccac gcttcccgaa gggagaaagg cggacaggta tccggtaagc
2941 ggcagggtcg gaacaggaga gcgcacgagg gagcttccag ggggaaacgc ctggtatctt
3001 tatagtcctg tcgggtttcg ccacctctga cttgagcgtc gatttttgtg atgctcgtca
3061 gggggggcgga gcctatggaa aaacgccagc aacgcggcct ttttacggtt cctggccttt
3121 tgctggcctt ttgctcacat gttctttcct gcgttatccc ctgattctgt ggataaccgt
3181 attaccgcct ttgagtgagc tgataccgct cgccgcagcc gaacgaccga gcgcagcgag
3241 tcagtgagcg aggaagcgga agagcgccca atacgcaaac cgcctctccc cgcgcgttgg
3301 ccgattcatt aatgcagagc ttgcaattcg cgcgtttttc aatattattg aagcatttat
3361 cagggttatt gtctcatgag cggatacata tttgaatgta tttagaaaaa taaacaaata
3421 ggggttccgc gcacatttcc ccgaaaagtg ccacctgacg tctaagaaac cattattatc
3481 atgacattaa cctataaaaa taggcgtagt acgaggccct ttcactcatt agatgcatgt
3541 cgttacataa cttacggtaa atggcccgcc tggctgaccg cccaacgacc cccgcccatt
3601 gacgtcaata atgacgtatg ttcccatagt aacgccaata gggactttcc attgacgtca
3661 atgggtggag tatttacggt aaactgccca cttggcagta catcaagtgt atcatatgcc
3721 aagtacgccc cctattgacg tcaatgacgg taaatggccc gcctggcatt atgcccagta
3781 catgacctta tgggactttc ctacttggca gtacatctac gtattagtca tcgctattac
3841 catggtgatg cggttttggc agtacatcaa tgggcgtgga tagcggtttg actcacgggg
3901 atttccaagt ctccacccca ttgacgtcaa tgggagtttg ttttggcacc aaaatcaacg
3961 ggactttcca aaatgtcgta acaactccgc cccattgacg caaatgggcg gtaggcgtgt
4021 acggtgggag gtctatataa gcagagctcg tttagtgaac cgtcagatcg cctggagacg
4081 ccatccacgc tgttttgacc tccatagaag acaccgggac cgatccagcc tccggactct
4141 agcctaggcc gcggagcgga taacaatttc acacaggaaa cagctatgac cattaggcct
4201 atttaggtga cactatagaa caagtttgta caaaaaagca ggctggtacc ggtccggaat
4261 tcccgggata tcgtcgaggc cgctctagag ccaccatggc ccccaagaag aagaggaagg
4321 tgatggacag cctcttgatg aaccggagga agtttcttta ccaattcaaa aatgtccgct
4381 gggctaaggg tcggcgtgag acctacctgt gctacgtagt gaagaggcgt gacagtgcta
4441 catccttttc actggacttt ggttatcttc gcaataagaa cggctgccac gtggaattgc
4501 tcttcctccg ctacatctcg gactgggacc tagaccctgg ccgctgctac cgcgtcacct
4561 ggttcacctc ctggagcccc tgctacgact gtgcccgaca tgtggccgac tttctgcgag
4621 ggaaccccaa cctcagtctg aggatcttca ccgcgcgcct ctacttctgt gaggaccgca
4681 aggctgagcc cgaggggctg cggcggctgc accgcgccgg ggtgcaaata gccatcatga
4741 ccttcaaaga ttatttttac tgctggaata cttttgtaga aaaccatgaa agaactttca
4801 aagcctggga agggctgcat gaaaattcag ttcgtctctc cagacagctt cggcgcatcc
4861 ttttgcccct gtatgaggtt gatgacttac gagacgcatt tcgtactttg ggacttgaac
4921 tgaaaactcc acttggagat actacacata cctcaccacc ttgcccagct cctgaactgc
4981 tgggtggacc tatggcttct aactttactc agttcgttct cgtcgacaat ggcggaactg
5041 gcgacgtgac tgtcgcccca agcaacttcg ctaacgggat cgctgaatgg atcagctcta
5101 actcgcgttc acaggcttac aaagtaacct gtagcgttcg tcagagctct gcgcagaatc
5161 gcaaatacac catcaaagtc gaggtgccta aaggcgcctg gcgttcgtac ttaaatatgg
5221 aactaaccat tccaattttc gccacgaatt ccgactgcga gcttattgtt aaggcaatgc
5281 aaggtctcct aaaagatgga aacccgattc cctcagcaat cgcagcaaac tccggcatct
5341 acggatctgg agagggcaga ggaagtcttc taacatgcgg tgacgtggag gagaatcccg
5401 gccctggtac cgataaaaag tattctattg gtttagccat cggcactaat tccgttggat

5461 gggctgtcat aaccgatgaa tacaaagtac cttcaaagaa atttaaggtg ttggggaaca
5521 cagaccgtca ttcgattaaa aagaatctta tcggtgccct cctattcgat agtggcgaaa
5581 cggcagaggc gactcgcctg aaacgaaccg ctcggagaag gtatacacgt cgcaagaacc
5641 gaatatgtta cttacaagaa atttttagca atgagatggc caaagttgac gattctttct
5701 ttcaccgttt ggaagagtcc ttccttgtcg aagaggacaa gaaacatgaa cggcacccca
5761 tctttggaaa catagtagat gaggtggcat atcatgaaaa gtacccaacg atttatcacc
5821 tcagaaaaaa gctagttgac tcaactgata aagcggacct gaggttaatc tacttggctc
5881 ttgcccatat gataaagttc cgtgggcact ttctcattga gggtgatcta aatccggaca
5941 actcggatgt cgacaaactg ttcatccagt tagtacaaac ctataatcag ttgtttgaag
6001 agaaccctat aaatgcaagt ggcgtggatg cgaaggctat tcttagcgcc cgcctctcta
6061 aatcccgacg gctagaaaac ctgatcgcac aattacccgg agagaagaaa aatgggttgt
6121 tcggtaacct tatagcgctc tcactaggcc tgacaccaaa ttttaagtcg aacttcgact
6181 tagctgaaga tgccaaattg cagcttagta aggacacgta cgatgacgat ctcgacaatc
6241 tactggcaca aattggagat cagtatgcgg acttattttt ggctgccaaa aaccttagcg
6301 atgcaatcct cctatctgac atactgagag ttaatactga gattaccaag gcgccgttat
6361 ccgcttcaat gatcaaaagg tacgatgaac atcaccaaga cttgacactt ctcaaggccc
6421 tagtccgtca gcaactgcct gagaaatata aggaaatatt ctttgatcag tcgaaaaacg
6481 ggtacgcagg ttatattgac ggcggagcga gtcaagagga attctacaag tttatcaaac
6541 ccatattaga gaagatggat gggacggaag agttgcttgt aaaactcaat cgcgaagatc
6601 tactgcgaaa gcagcggact ttcgacaacg gtagcattcc acatcaaatc cacttaggcg
6661 aattgcatgc tatacttaga aggcaggagg atttttatcc gttcctcaaa gacaatcgtg
6721 aaaagattga gaaaatccta acctttcgca taccttacta tgtgggaccc ctggcccgag
6781 ggaactctcg gttcgcatgg atgacaagaa agtccgaaga aacgattact ccatggaatt
6841 ttgaggaagt tgtcgataaa ggtgcgtcag ctcaatcgtt catcgagagg atgaccaact
6901 ttgacaagaa tttaccgaac gaaaaagtat tgcctaagca cagtttactt tacgagtatt
6961 tcacagtgta caatgaactc acgaaagtta agtatgtcac tgagggcatg cgtaaacccg
7021 cctttctaag cggagaacag aagaaagcaa tagtagatct gttattcaag accaaccgca
7081 aagtgacagt taagcaattg aaagaggact actttaagaa aattgaatgc ttcgattctg
7141 tcgagatctc cggggtagaa gatcgattta atgcgtcact tggtacgtat catgacctcc
7201 taaagataat taaagataag gacttcctgg ataacgaaga gaatgaagat atcttagaag
7261 atatagtgtt gactcttacc ctctttgaag atcgggaaat gattgaggaa agactaaaaa
7321 catacgctca cctgttcgac gataaggtta tgaaacagtt aaagaggcgt cgctatacgg
7381 gctggggacg attgtcgcgg aaacttatca acgggataag agacaagcaa agtggtaaaa
7441 ctattctcga ttttctaaag agcgacggct tcgccaatag gaactttatg cagctgatcc
7501 atgatgactc tttaaccttc aaagaggata tacaaaaggc acaggtttcc ggacaagggg
7561 actcattgca cgaacatatt gcgaatcttg ctggttcgcc agccatcaaa aagggcatac
7621 tccagacagt caaagtagtg gatgagctag ttaaggtcat gggacgtcac aaaccggaaa
7681 acattgtaat cgagatggca cgcgaaaatc aaacgactca gaaggggcaa aaaaacagtc
7741 gagagcggat gaagagaata gaagagggta ttaaagaact gggcagccag atcttaaagg
7801 agcatcctgt ggaaaatacc caattgcaga acgagaaact ttacctctat tacctacaaa
7861 atggaaggga catgtatgtt gatcaggaac tggacataaa ccgtttatct gattacgacg
7921 tcgatcacat tgtaccccaa tccttttga aggacgattc aatcgacaat aaagtgctta
7981 cacgctcgga taagaaccga gggaaaagtg acaatgttcc aagcgaggaa gtcgtaaaga
8041 aaatgaagaa ctattggcgg cagctcctaa atgcgaaact gataacgcaa agaaagttcg
8101 ataacttaac taaagctgag aggggtggct tgtctgaact tgacaaggcc ggatttatta
8161 aacgtcagct cgtggaaacc cgccaaatca caaagcatgt tgcacagata ctagattccc
8221 gaatgaatac gaaatacgac gagaacgata agctgattcg ggaagtcaaa gtaatcactt
8281 taaagtcaaa attggtgtcg gacttcagaa aggattttca attctataaa gttagggaga
8341 taaataacta ccaccatgcg cacgacgctt atcttaatgc cgtcgtaggg accgcactca
8401 ttaagaaata cccgaagcta gaaagtgagt ttgtgtatgg tgattacaaa gtttatgacg
8461 tccgtaagat gatcgcgaaa agcgaacagg agataggcaa ggctacagcc aaatacttct
8521 tttattctaa cattatgaat ttctttaaga cggaaatcac tctggcaaac ggagagatac
8581 gcaaacgacc tttaattgaa accaatgggg agacaggtga aatcgtatgg gataagggcc
8641 gggacttcgc gacggtgaga aaagttttgt ccatgcccca agtcaacata gtaaagaaaa
8701 ctgaggtgca gaccggaggg ttttcaaagg aatcgattct tccaaaaagg aatagtgata
8761 agctcatcgc tcgtaaaaag gactgggacc cgaaaaagta cggtggcttc gtgagcccta
8821 cagttgccta ttctgtccta gtagtggcaa aagttgagaa gggaaaatcc aagaaactga
8881 agtcagtcaa agaattattg gggataacga ttatggagcg ctcgtctttt gaaaagaacc
8941 ccatcgactt ccttgaggcg aaaggttaca aggaagtaaa aaaggatctc ataattaaac
9001 taccaaagta tagtctgttt gagttagaaa atggccgaaa acggatgttg gctagcgccg
9061 gagagcttca aaaggggaac gaactcgcac taccgtctaa atacgtgaat ttcctgtatt
9121 tagcgtccca ttacgagaag ttgaaaggtt cacctgaaga taacgaacag aagcaacttt
9181 ttgttgagca gcacaaacat tatctcgacg aaatcataga gcaaatttcg gaattcagta

```
     9241 agagagtcat cctagctgat gccaatctgg acaaagtatt aagcgcatac aacaagcaca
     9301 gggataaacc catacgtgag caggcggaaa atattatcca tttgtttact cttaccaacc
     9361 tcggcgctcc agccgcattc aagtattttg acacaacgat agatcgcaaa cagtacagat
     9421 ctaccaagga ggtgctagac gcgacactga ttcaccaatc catcacggga ttatatgaaa
     9481 ctcggataga tttgtcacag cttgggggtg actctggtgg ttctactaat ctgtcagata
     9541 ttattgaaaa ggagaccggt aagcaactgg ttatccagga atccatcctc atgctcccag
     9601 aggaggtgga agaagtcatt gggaacaagc cggaaagcga tatactcgtg cacaccgcct
     9661 acgacgagag caccgacgag aatgtcatgc ttctgactag cgacgcccct gaatacaagc
     9721 cttgggctct ggtcatacag gatagcaacg gtgagaacaa gattaagatg ctctctggtg
     9781 gttctcccaa gaagaagagg aaagtctaat gagggcccaa gcttacgcgt acccagcttt
     9841 cttgtacaaa gtggtcccta tagtgagtcg tattataagc taggcactgg
//
```

5. phU6_gRNA_2xMS2

```
LOCUS       Exported                3582 bp ds-DNA     circular SYN 24-MAR-2016
DEFINITION  synthetic circular DNA
ACCESSION   .
VERSION     .
KEYWORDS    .
SOURCE      synthetic DNA construct
  ORGANISM  synthetic DNA construct
REFERENCE   1  (bases 1 to 3582)
  AUTHORS   .
  TITLE     Direct Submission
  JOURNAL   Exported Oct 9, 2018 from SnapGene Viewer 4.2.6
            http://www.snapgene.com
FEATURES             Location/Qualifiers
     source          1..3582
                     /organism="synthetic DNA construct"
                     /mol_type="other DNA"
     rep_origin      complement(3..458)
                     /direction=LEFT
                     /label=f1 ori
                     /note="f1 bacteriophage origin of replication; arrow
                     indicates direction of (+) strand synthesis"
     primer_bind     600..616
                     /label=M13 fwd
                     /note="common sequencing primer, one of multiple similar
                     variants"
     promoter        626..644
                     /label=T7 promoter
                     /note="promoter for bacteriophage T7 RNA polymerase"
     promoter        696..936
                     /label=U6 promoter
                     /note="RNA polymerase III promoter for human U6 snRNA"
     misc_RNA        963..1038
                     /label=gRNA scaffold
                     /note="guide RNA scaffold for the Streptococcus pyogenes
                     CRISPR/Cas9 system"
     misc_RNA        1045..1063
                     /label=MS2 stem loop
                     /note="stem loop that binds the bacteriophage MS2 coat
                     protein"
     misc_RNA        1074..1092
                     /label=MS2 stem loop
                     /note="stem loop that binds the bacteriophage MS2 coat
                     protein"
     promoter        complement(1161..1179)
```

```
                     /label=T3 promoter
                     /note="promoter for bacteriophage T3 RNA polymerase"
     primer_bind     complement(1200..1216)
                     /label=M13 rev
                     /note="common sequencing primer, one of multiple similar
                     variants"
     protein_bind    1224..1240
                     /label=lac operator
                     /bound_moiety="lac repressor encoded by lacI"
                     /note="The lac repressor binds to the lac operator to
                     inhibit transcription in E. coli. This inhibition can be
                     relieved by adding lactose or
                     isopropyl-beta-D-thiogalactopyranoside (IPTG)."
     promoter        complement(1248..1278)
                     /label=lac promoter
                     /note="promoter for the E. coli lac operon"
     protein_bind    1293..1314
                     /label=CAP binding site
                     /bound_moiety="E. coli catabolite activator protein"
                     /note="CAP binding activates transcription in the presence
                     of cAMP."
     rep_origin      complement(1602..2190)
                     /direction=LEFT
                     /label=ori
                     /note="high-copy-number ColE1/pMB1/pBR322/pUC origin of
                     replication"
     CDS             complement(2267..3061)
                     /codon_start=1
                     /gene="aph(3')-II (or nptII)"
                     /product="aminoglycoside phosphotransferase from Tn5"
                     /label=NeoR/KanR
                     /note="confers resistance to neomycin, kanamycin, and G418
                     (Geneticin(R))"

/translation="MIEQDGLHAGSPAAWVERLFGYDWAQQTIGCSDAAVFRLSAQGRP

VLFVKTDLSGALNELQDEAARLSWLATTGVPCAAVLDVVTEAGRDWLLLGEVPGQDLLS

SHLAPAEKVSIMADAMRRLHTLDPATCPFDHQAKHRIERARTRMEAGLVDQDDLDEEHQ

GLAPAELFARLKASMPDGEDLVVTHGDACLPNIMVENGRFSGFIDCGRLGVADRYQDIA
                     LATRDIAEELGGEWADRFLVLYGIAAPDSQRIAFYRLLDEFF"
     promoter        complement(3096..3453)
                     /label=SV40 promoter
                     /note="SV40 enhancer and early promoter"
     rep_origin      3110..3245
                     /label=SV40 ori
                     /note="SV40 origin of replication"
     promoter        complement(3455..3559)
                     /gene="bla"
                     /label=AmpR promoter
ORIGIN
        1 ctaaattgta agcgttaata ttttgttaaa attcgcgtta aatttttgtt aaatcagctc
       61 atttttaac caataggccg aaatcggcaa aatcccttat aaatcaaaag aatagaccga
      121 gatagggttg agtgttgttc cagtttggaa caagagtcca ctattaaaga acgtggactc
      181 caacgtcaaa gggcgaaaaa ccgtctatca gggcgatggc ccactacgtg aaccatcacc
      241 ctaatcaagt tttttggggt cgaggtgccg taaagcacta aatcggaacc ctaaagggag
      301 cccccgattt agagcttgac ggggaaagcc ggcgaacgtg gcgagaaagg aagggaagaa
      361 agcgaaagga gcgggcgcta gggcgctggc aagtgtagcg gtcacgctgc gcgtaaccac
      421 cacacccgcc gcgcttaatg cgccgctaca gggcgcgtcc cattcgccat tcaggctgcg
      481 caactgttgg gaagggcgat cggtgcgggc ctcttcgcta ttacgccagc tggcgaaagg
      541 gggatgtgct gcaaggcgat taagttgggt aacgccaggg ttttcccagt cacgacgttg
      601 taaaacgacg gccagtgagc gcgcgtaata cgactcacta tagggcgaat tgggtacccg
```

```
     661 tctcacaggc ggatcgatcc aaggtcgggc aggaagaggg cctatttccc atgattcctt
     721 catatttgca tatacgatac aaggctgtta gagagataat tggaattaat ttgactgtaa
     781 acacaaagat attagtacaa aatacgtgac gtagaaagta ataatttctt gggtagtttg
     841 cagttttaaa attatgtttt aaaatggact atcatatgct taccgtaact tgaaagtatt
     901 tcgatttctt ggctttatat atcttgtgga aaggacgaaa caccgggtct tcgagaagac
     961 ctgttttaga gctagaaata gcaagttaaa ataaggctag tccgttatca acttgaaaaa
    1021 gtggcaccga gtcggtgcgg gagcacatga ggatcaccca tgtgccacga gcgacatgag
    1081 gatcacccat gtcgctcgtg ttccctttt ttctccgctg agcgtactga gacgccgcgg
    1141 tggagctcca gcttttgttc cctttagtga gggttaattg cgcgcttggc gtaatcatgg
    1201 tcatagctgt ttcctgtgtg aaattgttat ccgctcacaa ttccacacaa catacgagcc
    1261 ggaagcataa agtgtaaagc ctggggtgcc taatgagtga gctaactcac attaattgcg
    1321 ttgcgctcac tgcccgcttt ccagtcggga aacctgtcgt gccagctgca ttaatgaatc
    1381 ggccaacgcg cggggagagg cggtttgcgt attgggcgct cttccgcttc ctcgctcact
    1441 gactcgctgc gctcggtcgt tcggctgcgg cgagcggtat cagctcactc aaaggcggta
    1501 atacggttat ccacagaatc aggggataac gcaggaaaga acatgtgagc aaaaggccag
    1561 caaaaggcca ggaaccgtaa aaaggccgcg ttgctggcgt ttttccatag gctccgcccc
    1621 cctgacgagc atcacaaaaa tcgacgctca agtcagaggt ggcgaaaccc gacaggacta
    1681 taaagatacc aggcgtttcc ccctggaagc tccctcgtgc gctctcctgt tccgaccctg
    1741 ccgcttaccg gatacctgtc cgcctttctc ccttcgggaa gcgtggcgct ttctcatagc
    1801 tcacgctgta ggtatctcag ttcggtgtag gtcgttcgct ccaagctggg ctgtgtgcac
    1861 gaaccccccg ttcagcccga ccgctgcgcc ttatccggta actatcgtct tgagtccaac
    1921 ccggtaagac acgacttatc gccactggca gcagccactg gtaacaggat tagcagagcg
    1981 aggtatgtag gcggtgctac agagttcttg aagtggtggc ctaactacgg ctacactaga
    2041 aggacagtat ttggtatctg cgctctgctg aagccagtta ccttcggaaa aagagttggt
    2101 agctcttgat ccggcaaaca aaccaccgct ggtagcggtg gtttttttgt ttgcaagcag
    2161 cagattacgc gcagaaaaaa aggatctcaa gaagatcctt tgatcttttc tacggggtct
    2221 gacgctcagt ggaacgaaaa ctcacgttaa gggattttgg tcatgatcag aagaactcgt
    2281 caagaaggcg atagaaggcg atgcgctgcg aatcgggagc ggcgataccg taaagcacga
    2341 ggaagcggtc agcccattcg ccgccaagct cttcagcaat atcacgggta gccaacgcta
    2401 tgtcctgata gcggtccgcc acacccagcc ggccacagtc gatgaatcca gaaaagcggc
    2461 cattttccac catgatattc ggcaagcagg catcgccatg ggtcacgacg agatcctcgc
    2521 cgtcgggcat gctcgccttg agcctggcga acagttcggc tggcgcgagc ccctgatgct
    2581 cttcgtccag atcatcctga tcgacaagac cggcttccat ccgagtacgt gctcgctcga
    2641 tgcgatgttt cgcttggtgg tcgaatgggc aggtagccgg atcaagcgta tgcagccgcc
    2701 gcattgcatc agccatgatg gatactttct cggcaggagc aaggtgagat gacaggagat
    2761 cctgccccgg cacttcgccc aatagcagcc agtcccttcc cgcttcagtg acaacgtcga
    2821 gcacagctgc gcaaggaacg cccgtcgtgg ccagccacga tagccgcgct gcctcgtctt
    2881 gcagttcatt cagggcaccg gacaggtcgg tcttgacaaa aagaaccggg cgcccctgcg
    2941 ctgacagccg gaacacggcg gcatcagagc agccgattgt ctgttgtgcc cagtcatagc
    3001 cgaatagcct ctccacccaa gcggccggag aacctgcgtg caatccatct tgttcaatca
    3061 tgcgaaacga tcctcatcct gtctcttgat cgatctttgc aaaagcctag gcctccaaaa
    3121 aagcctcctc actacttctg gaatagctca gaggccgagg cggcctcggc ctctgcataa
    3181 ataaaaaaaa ttagtcagcc atggggcgga gaatgggcgg aactgggcgg agttaggggc
    3241 gggatgggcg gagttagggg cgggactatg gttgctgact aattgagatg catgctttgc
    3301 atacttctgc ctgctgggga gcctggggac tttccacacc tggttgctga ctaattgaga
    3361 tgcatgcttt gcatacttct gcctgctggg gagcctgggg actttccaca ccctaactga
    3421 cacacattcc acagctggtt ctttccgcct caggactctt cctttttcaa tattattgaa
    3481 gcatttatca gggttattgt ctcatgagcg gatacatatt tgaatgtatt tagaaaaata
    3541 aacaaatagg ggttccgcgc acatttcccc gaaaagtgcc ac
//
```

6. pCMV_^{A}CRCnu.2

```
LOCUS       Exported                9303 bp ds-DNA     circular SYN 12-AUG-2018
```

```
DEFINITION  synthetic circular DNA
ACCESSION   .
VERSION     .
KEYWORDS    .
SOURCE      synthetic DNA construct
  ORGANISM  synthetic DNA construct
REFERENCE   1  (bases 1 to 9303)
  AUTHORS   Juan-Carlos Collantes
  TITLE     Direct Submission
  JOURNAL   Exported Aug 14, 2018 from SnapGene Viewer 4.2.3
            http://www.snapgene.com
FEATURES             Location/Qualifiers
     source          1..9303
                     /organism="synthetic DNA construct"
                     /mol_type="other DNA"
     promoter        132..335
                     /label=CMV promoter
                     /note="human cytomegalovirus (CMV) immediate early
                     promoter"
     misc_feature    353..1848
                     /label=Synthetic
     promoter        377..395
                     /label=T7 promoter
                     /note="promoter for bacteriophage T7 RNA polymerase"
     regulatory      445..454
                     /regulatory_class="other"
                     /note="vertebrate consensus sequence for strong initiation
                     of translation (Kozak, 1987)"
     CDS             457..477
                     /codon_start=1
                     /product="nuclear localization signal of SV40 (simian
virus
                     40) large T antigen"
                     /label=SV40 NLS
                     /translation="PKKKRKV"
     CDS             478..1071
                     /codon_start=1
                     /gene="AICDA"
                     /product="human activation-induced cytidine deaminase
                     (Dickerson et al., 2003)"
                     /label=AID

/translation="MDSLLMNRRKFLYQFKNVRWAKGRRETYLCYVVKRRDSATSFSLD

FGYLRNKNGCHVELLFLRYISDWDLDPGRCYRVTWFTSWSPCYDCARHVADFLRGNPNL

SLRIFTARLYFCEDRKAEPEGLRRLHRAGVQIAIMTFKDYFYCWNTFVENHERTFKAWE
                     GLHENSVRLSRQLRRILLPLYEVDDLRDAFRTLGL"
     CDS             1072..1146
                     /codon_start=1
                     /label=h3
                     /translation="ELKTPLGDTTHTSPPCPAPELLGGP"
     CDS             1147..1497
                     /codon_start=1
                     /label=MCP

/translation="MASNFTQFVLVDNGGTGDVTVAPSNFANGIAEWISSNSRSQAYKV

TCSVRQSSAQNRKYTIKVEVPKGAWRSYLNMELTIPIFATNSDCELIVKAMQGLLKDGN
                     PIPSAIAANSGIY"
     CDS             1507..1560
                     /codon_start=1
                     /product="2A peptide from Thosea asigna virus capsid
```

```
                     protein"
                     /label=T2A
                     /note="Eukaryotic ribosomes fail to insert a peptide bond
                     between the Gly and Pro residues, yielding separate
                     polypeptides."
                     /translation="EGRGSLLTCGDVEENPGP"
     CDS             1567..1587
                     /codon_start=1
                     /product="nuclear localization signal of SV40 (simian
virus
                     40) large T antigen"
                     /label=SV40 NLS
                     /translation="PKKKRKV"
     CDS             1588..5688
                     /codon_start=1
                     /product="nickase mutant of the Cas9 endonuclease from the
                     Streptococcus pyogenes Type II CRISPR/Cas system"
                     /label=Cas9(D10A)
                     /note="generates RNA-guided single strand nicks in DNA due
                     to the D10A mutation in the RuvC catalytic domain"

/translation="DKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKN

LIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESF

LVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKF

RGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLE

NLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQI

GDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQ

QLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLR

KQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGN

SRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYF

TVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDS

VEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERL

KTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQ

LIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRH

KPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYL

YYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPS

EEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHV

AQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLN

AVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTE

ITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKE

SILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGIT

IMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNEL
```

```
ALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADA

NLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLD
                     ATLIHQSITGLYETRIDLSQLGGD"
     CDS             5719..5967
                     /codon_start=1
                     /product="uracil-DNA glycosylase inhibitor from a Bacillus
                     subtilis bacteriophage (Mol et al., 1995)"
                     /label=UGI

/translation="TNLSDIIEKETGKQLVIQESILMLPEEVEEVIGNKPESDILVHTA
                     YDESTDENVMLLTSDAPEYKPWALVIQDSNGENKIKML"
     CDS             5998..6246
                     /codon_start=1
                     /product="uracil-DNA glycosylase inhibitor from a Bacillus
                     subtilis bacteriophage (Mol et al., 1995)"
                     /label=UGI

/translation="TNLSDIIEKETGKQLVIQESILMLPEEVEEVIGNKPESDILVHTA
                     YDESTDENVMLLTSDAPEYKPWALVIQDSNGENKIKML"
     CDS             6289..6309
                     /codon_start=1
                     /product="nuclear localization signal of SV40 (simian
virus
                     40) large T antigen"
                     /label=SV40 NLS
                     /translation="PKKKRKV"
     CDS             6318..6335
                     /codon_start=1
                     /product="6xHis affinity tag"
                     /label=6xHis
                     /translation="HHHHHH"
     polyA_signal    6364..6588
                     /label=bGH poly(A) signal
                     /note="bovine growth hormone polyadenylation signal"
     primer_bind     complement(6659..6675)
                     /label=M13 rev
                     /note="common sequencing primer, one of multiple similar
                     variants"
     protein_bind    6683..6699
                     /label=lac operator
                     /bound_moiety="lac repressor encoded by lacI"
                     /note="The lac repressor binds to the lac operator to
                     inhibit transcription in E. coli. This inhibition can be
                     relieved by adding lactose or
                     isopropyl-beta-D-thiogalactopyranoside (IPTG)."
     promoter        complement(6707..6737)
                     /label=lac promoter
                     /note="promoter for the E. coli lac operon"
     protein_bind    6752..6773
                     /label=CAP binding site
                     /bound_moiety="E. coli catabolite activator protein"
                     /note="CAP binding activates transcription in the presence
                     of cAMP."
     rep_origin      complement(7061..7649)
                     /direction=LEFT
                     /label=ori
                     /note="high-copy-number ColE1/pMB1/pBR322/pUC origin of
                     replication"
     CDS             complement(7820..8680)
                     /codon_start=1
                     /gene="bla"
```

```
                     /product="beta-lactamase"
                     /label=AmpR
                     /note="confers resistance to ampicillin, carbenicillin,
and
                     related antibiotics"

/translation="MSIQHFRVALIPFFAAFCLPVFAHPETLVKVKDAEDQLGARVGYI

ELDLNSGKILESFRPEERFPMMSTFKVLLCGAVLSRIDAGQEQLGRRIHYSQNDLVEYS

PVTEKHLTDGMTVRELCSAAITMSDNTAANLLLTTIGGPKELTAFLHNMGDHVTRLDRW

EPELNEAIPNDERDTTMPVAMATTLRKLLTGELLTLASRQQLIDWMEADKVAGPLLRSA

LPAGWFIADKSGAGERGSRGIIAALGPDGKPSRIVVIYTTGSQATMDERNRQIAEIGAS
                     LIKHW"
     promoter        complement(8681..8785)
                     /gene="bla"
                     /label=AmpR promoter
     enhancer        join(9055..9303,1..131)
                     /label=CMV enhancer
                     /note="human cytomegalovirus immediate early enhancer"
ORIGIN
        1 atatgccaag tacgccccct attgacgtca atgacggtaa atggcccgcc tggcattatg
       61 cccagtacat gaccttatgg gactttccta cttggcagta catctacgta ttagtcatcg
      121 ctattaccat ggtgatgcgg ttttggcagt acatcaatgg gcgtggatag cggtttgact
      181 cacggggatt tccaagtctc caccccattg acgtcaatgg gagtttgttt tggcaccaaa
      241 atcaacggga ctttccaaaa tgtcgtaaca actccgcccc attgacgcaa atgggcggta
      301 ggcgtgtacg gtgggaggtc tatataagca gagctggttt agtgaaccgt cagatccgct
      361 agagatccgc ggccgctaat acgactcact atagggagag ccgccaccaa caaacggaca
      421 gccgacggaa gcgagttcga gtcagccacc atggccccca agaagaagcg gaaagtgatg
      481 gatagcctgc tgatgaaccg gagaaagttc ctgtatcagt ttaagaatgt gcgctgggca
      541 aagggcaggc gcgagaccta cctgtgctat gtggtgaagc ggagagattc cgccacatcc
      601 ttctctctgg actttggcta cctgcggaac aagaatggct gccacgtgga gctgctgttc
      661 ctgagataca tctctgactg ggatctggac ccaggcaggt gttatcgcgt gacctggttc
      721 acaagctggt ccccctgcta cgattgtgca aggcacgtgg cagactttct gaggggaaac
      781 ccaaatctgt ccctgcggat cttcaccgcc agactgtatt tttgcgagga taggaaggca
      841 gagccagagg gactgaggcg cctgcacagg gccggcgtgc agatcgccat catgaccttc
      901 aaggactact tttattgttg gaacaccttc gtggagaatc acgagcggac cttcaaggcc
      961 tgggagggac tgcacgagaa ctccgtgcgg ctgtctagac agctgcggag aatcctgctg
     1021 cctctgtacg aggtggacga tctgagggat gccttccgca ccctgggact ggagctgaag
     1081 acacccctgg gcgacaccac acacacctct ccaccttgcc cagcaccaga gctgctggga
     1141 ggccctatgg ccagcaactt cacacagttt gtgctggtgg ataatggagg aaccggcgac
     1201 gtgacagtgg caccatctaa ctttgccaat ggcatcgccg agtggatcag ctccaactct
     1261 cggagccagg cctataaggt gacctgtagc gtgcggcagt ctagcgccca gaatagaaag
     1321 tatacaatca aggtggaggt gcctaagggc gcctggagat cctacctgaa catggagctg
     1381 accatcccaa tctttgccac aaattctgat tgcgagctga tcgtgaaggc catgcagggc
     1441 ctgctgaagg acggcaaccc tatcccaagc gccatcgccg ccaatagcgg aatctacgga
     1501 tccggagagg gaagaggatc cctgctgacc tgtggcgacg tggaggagaa tccaggacct
     1561 ggaacaccaa agaagaagcg gaaagtcgac aagaagtaca gcatcggcct ggccatcggc
     1621 accaactctg tgggctgggc cgtgatcacc gacgagtaca aggtgcccag caagaaattc
     1681 aaggtgctgg gcaacaccga ccggcacagc atcaagaaga acctgatcgg agccctgctg
     1741 ttcgacagcg gcgaaacagc cgaggccacc cggctgaaga gaaccgccag aagaagatac
     1801 accagacgga agaaccggat ctgctatctg caagagatct tcagcaacga gatggccaag
     1861 gtggacgaca gcttcttcca cagactggaa gagtccttcc tggtggaaga ggataagaag
     1921 cacgagcggc accccatctt cggcaacatc gtggacgagg tggcctacca cgagaagtac
     1981 cccaccatct accacctgag aaagaaactg gtggacagca ccgacaaggc cgacctgcgg
     2041 ctgatctatc tggccctggc ccacatgatc aagttccggg gccacttcct gatcgagggc
     2101 gacctgaacc ccgacaacag cgacgtggac aagctgttca tccagctggt gcagacctac
     2161 aaccagctgt tcgaggaaaa ccccatcaac gccagcggcg tggacgccaa ggccatcctg
     2221 tctgccagac tgagcaagag cagacggctg gaaaatctga tcgcccagct gcccggcgag
     2281 aagaagaatg gcctgttcgg aaacctgatt gccctgagcc tgggcctgac ccccaacttc
     2341 aagagcaact tcgacctggc cgaggatgcc aaactgcagc tgagcaagga cacctacgac
```

2401 gacgacctgg acaacctgct ggcccagatc ggcgaccagt acgccgacct gtttctggcc
2461 gccaagaacc tgtccgacgc catcctgctg agcgacatcc tgagagtgaa caccgagatc
2521 accaaggccc ccctgagcgc ctctatgatc aagagatacg acgagcacca ccaggacctg
2581 accctgctga aagctctcgt gcggcagcag ctgcctgaga agtacaaaga gattttcttc
2641 gaccagagca agaacggcta cgccggctac attgacggcg gagccagcca ggaagagttc
2701 tacaagttca tcaagcccat cctggaaaag atggacggca ccgaggaact gctcgtgaag
2761 ctgaacagag aggacctgct gcggaagcag cggaccttcg acaacggcag catcccccac
2821 cagatccacc tgggagagct gcacgccatt ctgcggcggc aggaagattt ttacccattc
2881 ctgaaggaca accgggaaaa gatcgagaag atcctgacct tccgcatccc ctactacgtg
2941 ggccctctgg ccaggggaaa cagcagattc gcctggatga ccagaaagag cgaggaaacc
3001 atcacccccct ggaacttcga ggaagtggtg gacaagggcg cttccgccca gagcttcatc
3061 gagcggatga ccaacttcga taagaacctg cccaacgaga aggtgctgcc caagcacagc
3121 ctgctgtacg agtacttcac cgtgtataac gagctgacca aagtgaaata cgtgaccgag
3181 ggaatgagaa agcccgcctt cctgagcggc gagcagaaaa aggccatcgt ggacctgctg
3241 ttcaagacca accggaaagt gaccgtgaag cagctgaaag aggactactt caagaaaatc
3301 gagtgcttcg actccgtgga aatctccggc gtggaagatc ggttcaacgc ctccctgggc
3361 acataccacg atctgctgaa aattatcaag gacaaggact tcctggacaa tgaggaaaac
3421 gaggacattc tggaagatat cgtgctgacc ctgacactgt ttgaggacag agagatgatc
3481 gaggaacggc tgaaaaccta tgcccacctg ttcgacgaca aagtgatgaa gcagctgaag
3541 cggcggagat acaccggctg gggcaggctg agccggaagc tgatcaacgg catccgggac
3601 aagcagtccg gcaagacaat cctggatttc ctgaagtccg acggcttcgc caacagaaac
3661 ttcatgcagc tgatccacga cgacagcctg acctttaaag aggacatcca gaaagcccag
3721 gtgtccggcc agggcgatag cctgcacgag cacattgcca atctggccgg cagccccgcc
3781 attaagaagg gcatcctgca gacagtgaag gtggtggacg agctcgtgaa agtgatgggc
3841 cggcacaagc ccgagaacat cgtgatcgaa atggccagag agaaccagac cacccagaag
3901 ggacagaaga acagccgcga gagaatgaag cggatcgaag agggcatcaa agagctgggc
3961 agccagatcc tgaaagaaca ccccgtggaa aacacccagc tgcagaacga gaagctgtac
4021 ctgtactacc tgcagaatgg gcgggatatg tacgtggacc aggaactgga catcaaccgg
4081 ctgtccgact acgatgtgga ccatatcgtg cctcagagct ttctgaagga cgactccatc
4141 gacaacaagg tgctgaccag aagcgacaag aaccggggca agagcgacaa cgtgccctcc
4201 gaagaggtcg tgaagaagat gaagaactac tggcggcagc tgctgaacgc caagctgatt
4261 acccagagaa agttcgacaa tctgaccaag gccgagagag gcggcctgag cgaactggat
4321 aaggccggct tcatcaagag acagctggtg gaaacccggc agatcacaaa gcacgtggca
4381 cagatcctgg actcccggat gaacactaag tacgacgaga atgacaagct gatccgggaa
4441 gtgaaagtga tcaccctgaa gtccaagctg gtgtccgatt tccggaagga tttccagttt
4501 tacaaagtgc gcgagatcaa caactaccac cacgcccacg acgcctacct gaacgccgtc
4561 gtgggaaccg ccctgatcaa aaagtaccct aagctggaaa gcgagttcgt gtacggcgac
4621 tacaaggtgt acgacgtgcg gaagatgatc gccaagagcg agcaggaaat cggcaaggct
4681 accgccaagt acttcttcta cagcaacatc atgaactttt tcaagaccga gattaccctg
4741 gccaacggcg agatccggaa gcggcctctg atcgagacaa acggcgaaac cggggagatc
4801 gtgtgggata agggccggga ttttgccacc gtgcggaaag tgctgagcat gccccaagtg
4861 aatatcgtga aaaagaccga ggtgcagaca ggcggcttca gcaaagagtc tatcctgccc
4921 aagaggaaca gcgataagct gatcgccaga aagaaggact gggaccctaa gaagtacggc
4981 ggcttcgaca gccccaccgt ggcctattct gtgctggtgg tggccaaagt ggaaaagggc
5041 aagtccaaga aactgaagag tgtgaaagag ctgctgggga tcaccatcat ggaaagaagc
5101 agcttcgaga agaatcccat cgactttctg gaagccaagg gctacaaaga agtgaaaaag
5161 gacctgatca tcaagctgcc taagtactcc ctgttcgagc tggaaaacgg ccggaagaga
5221 atgctggcct ctgccggcga actgcagaag ggaaacgaac tggccctgcc ctccaaatat
5281 gtgaacttcc tgtacctggc cagccactat gagaagctga agggctcccc cgaggataat
5341 gagcagaaac agctgtttgt ggaacagcac aagcactacc tggacgagat catcgagcag
5401 atcagcgagt tctccaagag agtgatcctg gccgacgcta atctggacaa agtgctgtcc
5461 gcctacaaca agcaccggga taagcccatc agagagcagg ccgagaatat catccacctg
5521 tttaccctga ccaatctggg agcccctgcc gccttcaagt actttgacac caccatcgac
5581 cggaagaggt acaccagcac caaagaggtg ctggacgcca ccctgatcca ccagagcatc
5641 accggcctgt acgagacacg gatcgacctg tctcagctgg gaggtgacag cggcgggagc
5701 ggcgggagcg gggggagcac taatctgagc gacatcattg agaaggagac tgggaaacag
5761 ctggtcattc aggagtccat cctgatgctg cctgaggagg tggaggaagt gatcggcaac
5821 aagccagagt ctgacatcct ggtgcacacc gcctacgacg agtccacaga tgagaatgtg
5881 atgctgctga cctctgacgc ccccgagtat aagccttggg ccctggtcat ccaggattct
5941 aacggcgaga ataagatcaa gatgctgagc ggaggatccg gaggatctgg aggcagcacc
6001 aacctgtctg acatcatcga gaaggagaca ggcaagcagc tggtcatcca ggagagcatc
6061 ctgatgctgc ccgaagaagt cgaagaagtg atcggaaaca agcctgagag cgatatcctg
6121 gtccataccg cctacgacga gagtaccgac gaaaatgtga tgctgctgac atccgacgcc

```
     6181 ccagagtata agccctgggc tctggtcatc caggattcca acggagagaa caaaatcaaa
     6241 atgctgtctg gcggctcaaa aagaaccgcc gacggcagcg aattcgagcc caagaagaag
     6301 aggaaagtct aaccggtcat catcaccatc accattgagt ttaaacccgc tgatcagcct
     6361 cgactgtgcc ttctagttgc cagccatctg ttgtttgccc ctcccccgtg ccttccttga
     6421 ccctggaagg tgccactccc actgtccttt cctaataaaa tgaggaaatt gcatcgcatt
     6481 gtctgagtag gtgtcattct attctggggg gtggggtggg gcaggacagc aagggggagg
     6541 attgggaaga caatagcagg catgctgggg atgcggtggg ctctatggct tctgaggcgg
     6601 aaagaaccag ctggggctcg ataccgtcga cctctagcta gagcttggcg taatcatggt
     6661 catagctgtt tcctgtgtga aattgttatc cgctcacaat tccacacaac atacgagccg
     6721 gaagcataaa gtgtaaagcc tagggtgcct aatgagtgag ctaactcaca ttaattgcgt
     6781 tgcgctcact gcccgctttc cagtcgggaa acctgtcgtg ccagctgcat taatgaatcg
     6841 gccaacgcgc ggggagaggc ggtttgcgta ttgggcgctc ttccgcttcc tcgctcactg
     6901 actcgctgcg ctcggtcgtt cggctgcggc gagcggtatc agctcactca aaggcggtaa
     6961 tacggttatc cacagaatca ggggataacg caggaaagaa catgtgagca aaaggccagc
     7021 aaaaggccag gaaccgtaaa aaggccgcgt tgctggcgtt tttccatagg ctccgccccc
     7081 ctgacgagca tcacaaaaat cgacgctcaa gtcagaggtg gcgaaacccg acaggactat
     7141 aaagatacca ggcgtttccc cctggaagct ccctcgtgcg ctctcctgtt ccgaccctgc
     7201 cgcttaccgg atacctgtcc gcctttctcc cttcgggaag cgtggcgctt tctcatagct
     7261 cacgctgtag gtatctcagt tcggtgtagg tcgttcgctc caagctgggc tgtgtgcacg
     7321 aaccccccgt tcagcccgac cgctgcgcct tatccggtaa ctatcgtctt gagtccaacc
     7381 cggtaagaca cgacttatcg ccactggcag cagccactgg taacaggatt agcagagcga
     7441 ggtatgtagg cggtgctaca gagttcttga agtggtggcc taactacggc tacactagaa
     7501 gaacagtatt tggtatctgc gctctgctga agccagttac cttcggaaaa agagttggta
     7561 gctcttgatc cggcaaacaa accaccgctg gtagcggtgg tttttttgtt tgcaagcagc
     7621 agattacgcg cagaaaaaaa ggatctcaag aagatccttt gatcttttct acggggtctg
     7681 acactcagtg gaacgaaaac tcacgttaag ggattttggt catgagatta tcaaaaagga
     7741 tcttcaccta gatcctttta aattaaaaat gaagttttaa atcaatctaa agtatatatg
     7801 agtaaacttg gtctgacagt taccaatgct taatcagtga ggcacctatc tcagcgatct
     7861 gtctatttcg ttcatccata gttgcctgac tccccgtcgt gtagataact acgatacggg
     7921 agggcttacc atctggcccc agtgctgcaa tgataccgcg agacccacgc tcaccggctc
     7981 cagatttatc agcaataaac cagccagccg gaagggccga gcgcagaagt ggtcctgcaa
     8041 ctttatccgc ctccatccag tctattaatt gttgccggga agctagagta agtagttcgc
     8101 cagttaatag tttgcgcaac gttgttgcca ttgctacagg catcgtggtg tcacgctcgt
     8161 cgtttggtat ggcttcattc agctccggtt cccaacgatc aaggcgagtt acatgatccc
     8221 ccatgttgtg caaaaaagcg gttagctcct tcggtcctcc gatcgttgtc agaagtaagt
     8281 tggccgcagt gttatcactc atggttatgg cagcactgca taattctctt actgtcatgc
     8341 catccgtaag atgcttttct gtgactggtg agtactcaac caagtcattc tgagaatagt
     8401 gtatgcggcg accgagttgc tcttgcccgg cgtcaatacg ggataatacc gcgccacata
     8461 gcagaacttt aaaagtgctc atcattggaa aacgttcttc ggggcgaaaa ctctcaagga
     8521 tcttaccgct gttgagatcc agttcgatgt aacccactcg tgcacccaac tgatcttcag
     8581 catcttttac tttcaccagc gtttctgggt gagcaaaaac aggaaggcaa aatgccgcaa
     8641 aaaagggaat aagggcgaca cggaaatgtt gaatactcat actcttcctt tttcaatatt
     8701 attgaagcat ttatcagggt tattgtctca tgagcggata catatttgaa tgtatttaga
     8761 aaaataaaca aataggggtt ccgcgcacat ttccccgaaa agtgccacct gacgtcgacg
     8821 gatcgggaga tcgatctccc gatcccctag ggtcgactct cagtacaatc tgctctgatg
     8881 ccgcatagtt aagccagtat ctgctccctg cttgtgtgtt ggaggtcgct gagtagtgcg
     8941 cgagcaaaat ttaagctaca acaaggcaag gcttgaccga caattgcatg aagaatctgc
     9001 ttagggttag gcgttttgcg ctgcttcgcg atgtacgggc cagatatacg cgttgacatt
     9061 gattattgac tagttattaa tagtaatcaa ttacggggtc attagttcat agcccatata
     9121 tggagttccg cgttacataa cttacggtaa atggcccgcc tggctgaccg cccaacgacc
     9181 cccgcccatt gacgtcaata atgacgtatg ttcccatagt aacgccaata gggactttcc
     9241 attgacgtca atgggtggag tatttacggt aaactgccca cttggcagta catcaagtgt
     9301 atc
//
```

7. pCMV_^{2NA}CRCnu.2

```
LOCUS       Exported                9333 bp ds-DNA     circular SYN 12-AUG-2018
DEFINITION  synthetic circular DNA
ACCESSION   .
```

```
VERSION     .
KEYWORDS    .
SOURCE      synthetic DNA construct
  ORGANISM  synthetic DNA construct
REFERENCE   1  (bases 1 to 9333)
  AUTHORS   Juan-Carlos Collantes
  TITLE     Direct Submission
  JOURNAL   Exported Aug 14, 2018 from SnapGene Viewer 4.2.3
            http://www.snapgene.com
FEATURES             Location/Qualifiers
     source          1..9333
                     /organism="synthetic DNA construct"
                     /mol_type="other DNA"
     promoter        132..335
                     /label=CMV promoter
                     /note="human cytomegalovirus (CMV) immediate early
                     promoter"
     misc_feature    353..1878
                     /label=Synthetic
     promoter        377..395
                     /label=T7 promoter
                     /note="promoter for bacteriophage T7 RNA polymerase"
     regulatory      445..454
                     /regulatory_class="other"
                     /note="vertebrate consensus sequence for strong initiation
                     of translation (Kozak, 1987)"
     CDS             457..477
                     /codon_start=1
                     /product="nuclear localization signal of SV40 (simian
virus
                     40) large T antigen"
                     /label=SV40 NLS
                     /translation="PKKKRKV"
     CDS             478..1071
                     /codon_start=1
                     /gene="AICDA"
                     /product="human activation-induced cytidine deaminase
                     (Dickerson et al., 2003)"
                     /label=AID

/translation="MDSLLMNRRKFLYQFKNVRWAKGRRETYLCYVVKRRDSATSFSLD

FGYLRNKNGCHVELLFLRYISDWDLDPGRCYRVTWFTSWSPCYDCARHVADFLRGNPNL

SLRIFTARLYFCEDRKAEPEGLRRLHRAGVQIAIMTFKDYFYCWNTFVENHERTFKAWE
                     GLHENSVRLSRQLRRILLPLYEVDDLRDAFRTLGL"
     CDS             1072..1146
                     /codon_start=1
                     /label=h3
                     /translation="ELKTPLGDTTHTSPPCPAPELLGGP"
     CDS             1147..1497
                     /codon_start=1
                     /label=MCP

/translation="MASNFTQFVLVDNGGTGDVTVAPSNFANGIAEWISSNSRSQAYKV

TCSVRQSSAQNRKYTIKVEVPKGAWRSYLNMELTIPIFATNSDCELIVKAMQGLLKDGN
                     PIPSAIAANSGIY"
     CDS             1507..1527
                     /codon_start=1
                     /product="nuclear localization signal of SV40 (simian
virus
                     40) large T antigen"
```

```
                     /label=SV40 NLS
                     /translation="PKKKRKV"
     CDS             1537..1590
                     /codon_start=1
                     /product="2A peptide from Thosea asigna virus capsid
                     protein"
                     /label=T2A
                     /note="Eukaryotic ribosomes fail to insert a peptide bond
                     between the Gly and Pro residues, yielding separate
                     polypeptides."
                     /translation="EGRGSLLTCGDVEENPGP"
     CDS             1597..1617
                     /codon_start=1
                     /product="nuclear localization signal of SV40 (simian
virus
                     40) large T antigen"
                     /label=SV40 NLS
                     /translation="PKKKRKV"
     CDS             1618..5718
                     /codon_start=1
                     /product="nickase mutant of the Cas9 endonuclease from the
                     Streptococcus pyogenes Type II CRISPR/Cas system"
                     /label=Cas9(D10A)
                     /note="generates RNA-guided single strand nicks in DNA due
                     to the D10A mutation in the RuvC catalytic domain"

/translation="DKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKN

LIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESF

LVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKF

RGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLE

NLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQI

GDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQ

QLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLR

KQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGN

SRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYF

TVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDS

VEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERL

KTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQ

LIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRH

KPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYL

YYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPS

EEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHV

AQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLN

AVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTE

ITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKE
```

```
SILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGIT

IMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNEL

ALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADA

NLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLD
                     ATLIHQSITGLYETRIDLSQLGGD"
     CDS             5749..5997
                     /codon_start=1
                     /product="uracil-DNA glycosylase inhibitor from a Bacillus
                     subtilis bacteriophage (Mol et al., 1995)"
                     /label=UGI

/translation="TNLSDIIEKETGKQLVIQESILMLPEEVEEVIGNKPESDILVHTA
                     YDESTDENVMLLTSDAPEYKPWALVIQDSNGENKIKML"
     CDS             6028..6276
                     /codon_start=1
                     /product="uracil-DNA glycosylase inhibitor from a Bacillus
                     subtilis bacteriophage (Mol et al., 1995)"
                     /label=UGI

/translation="TNLSDIIEKETGKQLVIQESILMLPEEVEEVIGNKPESDILVHTA
                     YDESTDENVMLLTSDAPEYKPWALVIQDSNGENKIKML"
     CDS             6319..6339
                     /codon_start=1
                     /product="nuclear localization signal of SV40 (simian
virus
                     40) large T antigen"
                     /label=SV40 NLS
                     /translation="PKKKRKV"
     CDS             6348..6365
                     /codon_start=1
                     /product="6xHis affinity tag"
                     /label=6xHis
                     /translation="HHHHHH"
     polyA_signal    6394..6618
                     /label=bGH poly(A) signal
                     /note="bovine growth hormone polyadenylation signal"
     primer_bind     complement(6689..6705)
                     /label=M13 rev
                     /note="common sequencing primer, one of multiple similar
                     variants"
     protein_bind    6713..6729
                     /label=lac operator
                     /bound_moiety="lac repressor encoded by lacI"
                     /note="The lac repressor binds to the lac operator to
                     inhibit transcription in E. coli. This inhibition can be
                     relieved by adding lactose or
                     isopropyl-beta-D-thiogalactopyranoside (IPTG)."
     promoter        complement(6737..6767)
                     /label=lac promoter
                     /note="promoter for the E. coli lac operon"
     protein_bind    6782..6803
                     /label=CAP binding site
                     /bound_moiety="E. coli catabolite activator protein"
                     /note="CAP binding activates transcription in the presence
                     of cAMP."
     rep_origin      complement(7091..7679)
                     /direction=LEFT
                     /label=ori
                     /note="high-copy-number ColE1/pMB1/pBR322/pUC origin of
```

```
                     replication"
     CDS             complement(7850..8710)
                     /codon_start=1
                     /gene="bla"
                     /product="beta-lactamase"
                     /label=AmpR
                     /note="confers resistance to ampicillin, carbenicillin,
and
                     related antibiotics"

/translation="MSIQHFRVALIPFFAAFCLPVFAHPETLVKVKDAEDQLGARVGYI

ELDLNSGKILESFRPEERFPMMSTFKVLLCGAVLSRIDAGQEQLGRRIHYSQNDLVEYS

PVTEKHLTDGMTVRELCSAAITMSDNTAANLLLTTIGGPKELTAFLHNMGDHVTRLDRW

EPELNEAIPNDERDTTMPVAMATTLRKLLTGELLTLASRQQLIDWMEADKVAGPLLRSA

LPAGWFIADKSGAGERGSRGIIAALGPDGKPSRIVVIYTTGSQATMDERNRQIAEIGAS
                     LIKHW"
     promoter        complement(8711..8815)
                     /gene="bla"
                     /label=AmpR promoter
     enhancer        join(9085..9333,1..131)
                     /label=CMV enhancer
                     /note="human cytomegalovirus immediate early enhancer"
ORIGIN
        1 atatgccaag tacgccccct attgacgtca atgacggtaa atggcccgcc tggcattatg
       61 cccagtacat gaccttatgg gactttccta cttggcagta catctacgta ttagtcatcg
      121 ctattaccat ggtgatgcgg ttttggcagt acatcaatgg gcgtggatag cggtttgact
      181 cacggggatt tccaagtctc caccccattg acgtcaatgg gagtttgttt tggcaccaaa
      241 atcaacggga ctttccaaaa tgtcgtaaca actccgcccc attgacgcaa atgggcggta
      301 ggcgtgtacg gtgggaggtc tatataagca gagctggttt agtgaaccgt cagatccgct
      361 agagatccgc ggccgctaat acgactcact atagggagag ccgccaccaa caaacggaca
      421 gccgacggaa gcgagttcga gtcagccacc atggccccca agaagaagcg gaaagtgatg
      481 gatagcctgc tgatgaaccg gagaaagttc ctgtatcagt ttaagaatgt gcgctgggca
      541 aagggcaggc gcgagaccta cctgtgctat gtggtgaagc ggagagattc cgccacatcc
      601 ttctctctgg actttggcta cctgcggaac aagaatggct gccacgtgga gctgctgttc
      661 ctgagataca tctctgactg ggatctggac ccaggcaggt gttatcgcgt gacctggttc
      721 acaagctggt ccccctgcta cgattgtgca aggcacgtgg cagactttct gaggggaaac
      781 ccaaatctgt ccctgcggat cttcaccgcc agactgtatt tttgcgagga taggaaggca
      841 gagccagagg gactgaggcg cctgcacagg gccggcgtgc agatcgccat catgaccttc
      901 aaggactact tttattgttg gaacaccttc gtggagaatc acgagcggac cttcaaggcc
      961 tgggagggac tgcacgagaa ctccgtgcgg ctgtctagac agctgcggag aatcctgctg
     1021 cctctgtacg aggtggacga tctgagggat gccttccgca ccctgggact ggagctgaag
     1081 acacccctgg gcgacaccac acacacctct ccaccttgcc cagcaccaga gctgctggga
     1141 ggccctatgg ccagcaactt cacacagttt gtgctggtgg ataatggagg aaccggcgac
     1201 gtgacagtgg caccatctaa ctttgccaat ggcatcgccg agtggatcag ctccaactct
     1261 cggagccagg cctataaggt gacctgtagc gtgcggcagt ctagcgccca gaatagaaag
     1321 tatacaatca aggtggaggt gcctaagggc gcctggagat cctacctgaa catggagctg
     1381 accatcccaa tctttgccac aaattctgat tgcgagctga tcgtgaaggc catgcagggc
     1441 ctgctgaagg acggcaaccc tatcccaagc gccatcgccg ccaatagcgg aatctacggt
     1501 tctggtccga aaaagaagag gaaggtagga tccggagagg gaagaggatc cctgctgacc
     1561 tgtggcgacg tggaggagaa tccaggacct ggaacaccaa agaagaagcg gaaagtcgac
     1621 aagaagtaca gcatcggcct ggccatcggc accaactctg tgggctgggc cgtgatcacc
     1681 gacgagtaca aggtgcccag caagaaattc aaggtgctgg gcaacaccga ccggcacagc
     1741 atcaagaaga acctgatcgg agccctgctg ttcgacagcg gcgaaacagc cgaggccacc
     1801 cggctgaaga gaaccgccag aagaagatac accagacgga agaaccggat ctgctatctg
     1861 caagagatct tcagcaacga gatggccaag gtggacgaca gcttcttcca cagactggaa
     1921 gagtccttcc tggtggaaga ggataagaag cacgagcggc accccatctt cggcaacatc
     1981 gtggacgagg tggcctacca cgagaagtac cccaccatct accacctgag aaagaaactg
     2041 gtggacagca ccgacaaggc cgacctgcgg ctgatctatc tggccctggc ccacatgatc
     2101 aagttccggg gccacttcct gatcgagggc gacctgaacc ccgacaacag cgacgtggac
```

2161 aagctgttca tccagctggt gcagacctac aaccagctgt tcgaggaaaa ccccatcaac
2221 gccagcggcg tggacgccaa ggccatcctg tctgccagac tgagcaagag cagacggctg
2281 gaaaatctga tcgcccagct gcccggcgag aagaagaatg gcctgttcgg aaacctgatt
2341 gccctgagcc tgggcctgac ccccaacttc aagagcaact tcgacctggc cgaggatgcc
2401 aaactgcagc tgagcaagga cacctacgac gacgacctgg acaacctgct ggcccagatc
2461 ggcgaccagt acgccgacct gtttctggcc gccaagaacc tgtccgacgc catcctgctg
2521 agcgacatcc tgagagtgaa caccgagatc accaaggccc ccctgagcgc ctctatgatc
2581 aagagatacg acgagcacca ccaggacctg accctgctga aagctctcgt gcggcagcag
2641 ctgcctgaga agtacaaaga gattttcttc gaccagagca agaacggcta cgccggctac
2701 attgacggcg gagccagcca ggaagagttc tacaagttca tcaagcccat cctggaaaag
2761 atggacggca ccgaggaact gctcgtgaag ctgaacagag aggacctgct gcggaagcag
2821 cggaccttcg acaacggcag catcccccac cagatccacc tgggagagct gcacgccatt
2881 ctgcggcggc aggaagattt ttacccattc ctgaaggaca accgggaaaa gatcgagaag
2941 atcctgacct tccgcatccc ctactacgtg ggccctctgg ccaggggaaa cagcagattc
3001 gcctggatga ccagaaagag cgaggaaacc atcacccct ggaacttcga ggaagtggtg
3061 gacaagggcg cttccgccca gagcttcatc gagcggatga ccaacttcga taagaacctg
3121 cccaacgaga aggtgctgcc caagcacagc ctgctgtacg agtacttcac cgtgtataac
3181 gagctgacca aagtgaaata cgtgaccgag ggaatgagaa agcccgcctt cctgagcggc
3241 gagcagaaaa aggccatcgt ggacctgctg ttcaagacca accggaaagt gaccgtgaag
3301 cagctgaaag aggactactt caagaaaatc gagtgcttcg actccgtgga aatctccggc
3361 gtggaagatc ggttcaacgc ctccctgggc acataccacg atctgctgaa aattatcaag
3421 gacaaggact tcctggacaa tgaggaaaac gaggacattc tggaagatat cgtgctgacc
3481 ctgacactgt ttgaggacag agagatgatc gaggaacggc tgaaaaccta tgcccacctg
3541 ttcgacgaca aagtgatgaa gcagctgaag cggcggagat acaccggctg gggcaggctg
3601 agccggaagc tgatcaacgg catccgggac aagcagtccg gcaagacaat cctggatttc
3661 ctgaagtccg acggcttcgc caacagaaac ttcatgcagc tgatccacga cgacagcctg
3721 acctttaaag aggacatcca gaaagcccag gtgtccggcc agggcgatag cctgcacgag
3781 cacattgcca atctggccgg cagccccgcc attaagaagg gcatcctgca gacagtgaag
3841 gtggtggacg agctcgtgaa agtgatgggc cggcacaagc ccgagaacat cgtgatcgaa
3901 atggccagag agaaccagac cacccagaag ggacagaaga acagccgcga gagaatgaag
3961 cggatcgaag agggcatcaa agagctgggc agccagatcc tgaaagaaca ccccgtggaa
4021 aacacccagc tgcagaacga gaagctgtac ctgtactacc tgcagaatgg gcgggatatg
4081 tacgtggacc aggaactgga catcaaccgg ctgtccgact acgatgtgga ccatatcgtg
4141 cctcagagct ttctgaagga cgactccatc gacaacaagg tgctgaccag aagcgacaag
4201 aaccggggca agagcgacaa cgtgccctcc gaagaggtcg tgaagaagat gaagaactac
4261 tggcggcagc tgctgaacgc caagctgatt acccagagaa agttcgacaa tctgaccaag
4321 gccgagagag gcggcctgag cgaactggat aaggccggct tcatcaagag acagctggtg
4381 gaaacccggc agatcacaaa gcacgtggca cagatcctgg actcccggat gaacactaag
4441 tacgacgaga atgacaagct gatccgggaa gtgaaagtga tcaccctgaa gtccaagctg
4501 gtgtccgatt tccggaagga tttccagttt tacaaagtgc gcgagatcaa caactaccac
4561 cacgcccacg acgcctacct gaacgccgtc gtgggaaccg ccctgatcaa aaagtaccct
4621 aagctggaaa gcgagttcgt gtacggcgac tacaaggtgt acgacgtgcg gaagatgatc
4681 gccaagagcg agcaggaaat cggcaaggct accgccaagt acttcttcta cagcaacatc
4741 atgaactttt tcaagaccga gattaccctg gccaacggcg agatccggaa gcggcctctg
4801 atcgagacaa acggcgaaac cggggagatc gtgtgggata agggccggga ttttgccacc
4861 gtgcggaaag tgctgagcat gccccaagtg aatatcgtga aaaagaccga ggtgcagaca
4921 ggcggcttca gcaaagagtc tatcctgccc aagaggaaca gcgataagct gatcgccaga
4981 aagaaggact gggaccctaa gaagtacggc ggcttcgaca gccccaccgt ggcctattct
5041 gtgctggtgg tggccaaagt ggaaaagggc aagtccaaga aactgaagag tgtgaaagag
5101 ctgctgggga tcaccatcat ggaaagaagc agcttcgaga agaatcccat cgactttctg
5161 gaagccaagg gctacaaaga agtgaaaaag gacctgatca tcaagctgcc taagtactcc
5221 ctgttcgagc tggaaaacgg ccggaagaga atgctggcct ctgccggcga actgcagaag
5281 ggaaacgaac tggccctgcc ctccaaatat gtgaacttcc tgtacctggc cagccactat
5341 gagaagctga agggctcccc cgaggataat gagcagaaac agctgtttgt ggaacagcac
5401 aagcactacc tggacgagat catcgagcag atcagcgagt tctccaagag agtgatcctg
5461 gccgacgcta atctggacaa agtgctgtcc gcctacaaca agcaccggga taagcccatc
5521 agagagcagg ccgagaatat catccacctg tttaccctga ccaatctggg agcccctgcc
5581 gccttcaagt actttgacac caccatcgac cggaagaggt acaccagcac caaagaggtg
5641 ctggacgcca ccctgatcca ccagagcatc accggcctgt acgagacacg gatcgacctg
5701 tctcagctgg gaggtgacag cggcgggagc ggcgggagcg gggggagcac taatctgagc
5761 gacatcattg agaaggagac tgggaaacag ctggtcattc aggagtccat cctgatgctg
5821 cctgaggagg tggaggaagt gatcggcaac aagccagagt ctgacatcct ggtgcacacc
5881 gcctacgacg agtccacaga tgagaatgtg atgctgctga cctctgacgc ccccgagtat

```
5941 aagccttggg ccctggtcat ccaggattct aacggcgaga ataagatcaa gatgctgagc
6001 ggaggatccg gaggatctgg aggcagcacc aacctgtctg acatcatcga gaaggagaca
6061 ggcaagcagc tggtcatcca ggagagcatc ctgatgctgc ccgaagaagt cgaagaagtg
6121 atcggaaaca agcctgagag cgatatcctg gtccataccg cctacgacga gagtaccgac
6181 gaaaatgtga tgctgctgac atccgacgcc ccagagtata agccctgggc tctggtcatc
6241 caggattcca acggagagaa caaaatcaaa atgctgtctg gcggctcaaa aagaaccgcc
6301 gacggcagcg aattcgagcc caagaagaag aggaaagtct aaccggtcat catcaccatc
6361 accattgagt ttaaacccgc tgatcagcct cgactgtgcc ttctagttgc cagccatctg
6421 ttgtttgccc ctccccccgtg ccttccttga ccctggaagg tgccactccc actgtccttt
6481 cctaataaaa tgaggaaatt gcatcgcatt gtctgagtag gtgtcattct attctggggg
6541 gtggggtggg gcaggacagc aagggggagg attgggaaga caatagcagg catgctgggg
6601 atgcggtggg ctctatggct tctgaggcgg aaagaaccag ctggggctcg ataccgtcga
6661 cctctagcta gagcttggcg taatcatggt catagctgtt tcctgtgtga aattgttatc
6721 cgctcacaat tccacacaac atacgagccg gaagcataaa gtgtaaagcc tagggtgcct
6781 aatgagtgag ctaactcaca ttaattgcgt tgcgctcact gcccgctttc cagtcgggaa
6841 acctgtcgtg ccagctgcat taatgaatcg gccaacgcgc ggggagaggc ggtttgcgta
6901 ttgggcgctc ttccgcttcc tcgctcactg actcgctgcg ctcggtcgtt cggctgcggc
6961 gagcggtatc agctcactca aaggcggtaa tacggttatc cacagaatca ggggataacg
7021 caggaaagaa catgtgagca aaaggccagc aaaaggccag gaaccgtaaa aaggccgcgt
7081 tgctggcgtt tttccatagg ctccgccccc ctgacgagca tcacaaaaat cgacgctcaa
7141 gtcagaggtg gcgaaacccg acaggactat aaagatacca ggcgtttccc cctggaagct
7201 ccctcgtgcg ctctcctgtt ccgaccctgc cgcttaccgg atacctgtcc gcctttctcc
7261 cttcgggaag cgtggcgctt tctcatagct cacgctgtag gtatctcagt tcggtgtagg
7321 tcgttcgctc caagctgggc tgtgtgcacg aaccccccgt tcagcccgac cgctgcgcct
7381 tatccggtaa ctatcgtctt gagtccaacc cggtaagaca cgacttatcg ccactggcag
7441 cagccactgg taacaggatt agcagagcga ggtatgtagg cggtgctaca gagttcttga
7501 agtggtggcc taactacggc tacactagaa gaacagtatt tggtatctgc gctctgctga
7561 agccagttac cttcggaaaa agagttggta gctcttgatc cggcaaacaa accaccgctg
7621 gtagcggtgg tttttttgtt tgcaagcagc agattacgcg cagaaaaaaa ggatctcaag
7681 aagatccttt gatcttttct acggggtctg acactcagtg gaacgaaaac tcacgttaag
7741 ggatttggt catgagatta tcaaaaagga tcttcaccta gatcctttta aattaaaaat
7801 gaagttttaa atcaatctaa agtatatatg agtaaacttg gtctgacagt taccaatgct
7861 taatcagtga ggcacctatc tcagcgatct gtctatttcg ttcatccata gttgcctgac
7921 tccccgtcgt gtagataact acgatacggg agggcttacc atctggcccc agtgctgcaa
7981 tgataccgcg agacccacgc tcaccggctc cagatttatc agcaataaac cagccagccg
8041 gaagggccga gcgcagaagt ggtcctgcaa ctttatccgc ctccatccag tctattaatt
8101 gttgccggga agctagagta agtagttcgc cagttaatag tttgcgcaac gttgttgcca
8161 ttgctacagg catcgtggtg tcacgctcgt cgtttggtat ggcttcattc agctccggtt
8221 cccaacgatc aaggcgagtt acatgatccc ccatgttgtg caaaaaagcg gttagctcct
8281 tcggtcctcc gatcgttgtc agaagtaagt tggccgcagt gttatcactc atggttatgg
8341 cagcactgca taattctctt actgtcatgc catccgtaag atgcttttct gtgactggtg
8401 agtactcaac caagtcattc tgagaatagt gtatgcggcg accgagttgc tcttgcccgg
8461 cgtcaatacg ggataatacc gcgccacata gcagaacttt aaaagtgctc atcattggaa
8521 aacgttcttc ggggcgaaaa ctctcaagga tcttaccgct gttgagatcc agttcgatgt
8581 aacccactcg tgcacccaac tgatcttcag catcttttac tttcaccagc gtttctgggt
8641 gagcaaaaac aggaaggcaa aatgccgcaa aaaagggaat aagggcgaca cggaaatgtt
8701 gaatactcat actcttcctt tttcaatatt attgaagcat ttatcagggt tattgtctca
8761 tgagcggata catatttgaa tgtatttaga aaaataaaca aataggggtt ccgcgcacat
8821 ttccccgaaa agtgccacct gacgtcgacg gatcgggaga tcgatctccc gatcccctag
8881 ggtcgactct cagtacaatc tgctctgatg ccgcatagtt aagccagtat ctgctccctg
8941 cttgtgtgtt ggaggtcgct gagtagtgcg cgagcaaaat ttaagctaca acaaggcaag
9001 gcttgaccga caattgcatg aagaatctgc ttagggttag gcgttttgcg ctgcttcgcg
9061 atgtacgggc cagatatacg cgttgacatt gattattgac tagttattaa tagtaatcaa
9121 ttacggggtc attagttcat agcccatata tggagttccg cgttacataa cttacggtaa
9181 atggcccgcc tggctgaccg cccaacgacc cccgcccatt gacgtcaata atgacgtatg
9241 ttcccatagt aacgccaata gggactttcc attgacgtca atgggtggag tatttacggt
9301 aaactgccca cttggcagta catcaagtgt atc
//
```

8. pCMV_$^{N\Delta A}$CRCnu.2

```
LOCUS       Exported                9294 bp ds-DNA     circular SYN 12-AUG-2018
DEFINITION  synthetic circular DNA
ACCESSION   .
VERSION     .
KEYWORDS    .
SOURCE      synthetic DNA construct
  ORGANISM  synthetic DNA construct
REFERENCE   1  (bases 1 to 9294)
  AUTHORS   Juan-Carlos Collantes
  TITLE     Direct Submission
  JOURNAL   Exported Aug 14, 2018 from SnapGene Viewer 4.2.3
            http://www.snapgene.com
FEATURES             Location/Qualifiers
     source          1..9294
                     /organism="synthetic DNA construct"
                     /mol_type="other DNA"
     promoter        132..335
                     /label=CMV promoter
                     /note="human cytomegalovirus (CMV) immediate early
                     promoter"
     misc_feature    353..1839
                     /label=Synthetic
     promoter        377..395
                     /label=T7 promoter
                     /note="promoter for bacteriophage T7 RNA polymerase"
     regulatory      445..454
                     /regulatory_class="other"
                     /note="vertebrate consensus sequence for strong initiation
                     of translation (Kozak, 1987)"
     CDS             457..477
                     /codon_start=1
                     /product="nuclear localization signal of SV40 (simian
virus
                     40) large T antigen"
                     /label=SV40 NLS
                     /translation="PKKKRKV"
     CDS             478..1062
                     /codon_start=1
                     /label=NE-Delta_AID

/translation="MDSLLMNRRKFLYQFKNVRWAKGRRETYLCYVVKRRDSATSFSLD

FGYLRNKNGCHVELLFLRYISDWDLDPGRCYRVTWFTSWSPCYDCARHVADFLRGNPNL

SLRIFTARLYFCEDRKAEPEGLRRLHRAGVQIAIMTFKDYFYCWNTFVENHERTFKAWE
                     GLHENSVRLSRQLRRILLPLYEVDDLRDAFRT"
     CDS             1063..1137
                     /codon_start=1
                     /label=h3
                     /translation="ELKTPLGDTTHTSPPCPAPELLGGP"
     CDS             1138..1488
                     /codon_start=1
                     /label=MCP

/translation="MASNFTQFVLVDNGGTGDVTVAPSNFANGIAEWISSNSRSQAYKV

TCSVRQSSAQNRKYTIKVEVPKGAWRSYLNMELTIPIFATNSDCELIVKAMQGLLKDGN
                     PIPSAIAANSGIY"
     CDS             1498..1551
                     /codon_start=1
                     /product="2A peptide from Thosea asigna virus capsid
                     protein"
                     /label=T2A
```

```
                     /note="Eukaryotic ribosomes fail to insert a peptide bond
                     between the Gly and Pro residues, yielding separate
                     polypeptides."
                     /translation="EGRGSLLTCGDVEENPGP"
     CDS             1558..1578
                     /codon_start=1
                     /product="nuclear localization signal of SV40 (simian
virus
                     40) large T antigen"
                     /label=SV40 NLS
                     /translation="PKKKRKV"
     CDS             1579..5679
                     /codon_start=1
                     /product="nickase mutant of the Cas9 endonuclease from the
                     Streptococcus pyogenes Type II CRISPR/Cas system"
                     /label=Cas9(D10A)
                     /note="generates RNA-guided single strand nicks in DNA due
                     to the D10A mutation in the RuvC catalytic domain"

/translation="DKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKN

LIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESF

LVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKF

RGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLE

NLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQI

GDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQ

QLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLR

KQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGN

SRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYF

TVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDS

VEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERL

KTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQ

LIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRH

KPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYL

YYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPS

EEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHV

AQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLN

AVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTE

ITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKE

SILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGIT

IMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNEL

ALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADA
```

```
NLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLD
                     ATLIHQSITGLYETRIDLSQLGGD"
     CDS             5710..5958
                     /codon_start=1
                     /product="uracil-DNA glycosylase inhibitor from a Bacillus
                     subtilis bacteriophage (Mol et al., 1995)"
                     /label=UGI

/translation="TNLSDIIEKETGKQLVIQESILMLPEEVEEVIGNKPESDILVHTA
                     YDESTDENVMLLTSDAPEYKPWALVIQDSNGENKIKML"
     CDS             5989..6237
                     /codon_start=1
                     /product="uracil-DNA glycosylase inhibitor from a Bacillus
                     subtilis bacteriophage (Mol et al., 1995)"
                     /label=UGI

/translation="TNLSDIIEKETGKQLVIQESILMLPEEVEEVIGNKPESDILVHTA
                     YDESTDENVMLLTSDAPEYKPWALVIQDSNGENKIKML"
     CDS             6280..6300
                     /codon_start=1
                     /product="nuclear localization signal of SV40 (simian
virus
                     40) large T antigen"
                     /label=SV40 NLS
                     /translation="PKKKRKV"
     CDS             6309..6326
                     /codon_start=1
                     /product="6xHis affinity tag"
                     /label=6xHis
                     /translation="HHHHHH"
     polyA_signal    6355..6579
                     /label=bGH poly(A) signal
                     /note="bovine growth hormone polyadenylation signal"
     primer_bind     complement(6650..6666)
                     /label=M13 rev
                     /note="common sequencing primer, one of multiple similar
                     variants"
     protein_bind    6674..6690
                     /label=lac operator
                     /bound_moiety="lac repressor encoded by lacI"
                     /note="The lac repressor binds to the lac operator to
                     inhibit transcription in E. coli. This inhibition can be
                     relieved by adding lactose or
                     isopropyl-beta-D-thiogalactopyranoside (IPTG)."
     promoter        complement(6698..6728)
                     /label=lac promoter
                     /note="promoter for the E. coli lac operon"
     protein_bind    6743..6764
                     /label=CAP binding site
                     /bound_moiety="E. coli catabolite activator protein"
                     /note="CAP binding activates transcription in the presence
                     of cAMP."
     rep_origin      complement(7052..7640)
                     /direction=LEFT
                     /label=ori
                     /note="high-copy-number ColE1/pMB1/pBR322/pUC origin of
                     replication"
     CDS             complement(7811..8671)
                     /codon_start=1
                     /gene="bla"
                     /product="beta-lactamase"
                     /label=AmpR
```

```
                     /note="confers resistance to ampicillin, carbenicillin,
and
                     related antibiotics"

/translation="MSIQHFRVALIPFFAAFCLPVFAHPETLVKVKDAEDQLGARVGYI

ELDLNSGKILESFRPEERFPMMSTFKVLLCGAVLSRIDAGQEQLGRRIHYSQNDLVEYS

PVTEKHLTDGMTVRELCSAAITMSDNTAANLLLTTIGGPKELTAFLHNMGDHVTRLDRW

EPELNEAIPNDERDTTMPVAMATTLRKLLTGELLTLASRQQLIDWMEADKVAGPLLRSA

LPAGWFIADKSGAGERGSRGIIAALGPDGKPSRIVVIYTTGSQATMDERNRQIAEIGAS
                     LIKHW"
     promoter        complement(8672..8776)
                     /gene="bla"
                     /label=AmpR promoter
     enhancer        join(9046..9294,1..131)
                     /label=CMV enhancer
                     /note="human cytomegalovirus immediate early enhancer"
ORIGIN
        1 atatgccaag tacgccccct attgacgtca atgacggtaa atggcccgcc tggcattatg
       61 cccagtacat gaccttatgg gactttccta cttggcagta catctacgta ttagtcatcg
      121 ctattaccat ggtgatgcgg ttttggcagt acatcaatgg gcgtggatag cggtttgact
      181 cacggggatt tccaagtctc caccccattg acgtcaatgg gagtttgttt tggcaccaaa
      241 atcaacggga ctttccaaaa tgtcgtaaca actccgcccc attgacgcaa atgggcggta
      301 ggcgtgtacg gtgggaggtc tatataagca gagctggttt agtgaaccgt cagatccgct
      361 agagatccgc ggccgctaat acgactcact atagggagag ccgccaccaa caaacggaca
      421 gccgacggaa gcgagttcga gtcagccacc atggccccca agaagaagcg gaaagtgatg
      481 gatagcctgc tgatgaaccg gagaaagttc ctgtatcagt ttaagaatgt gcgctgggca
      541 aagggcaggc gcgagaccta cctgtgctat gtggtgaagc ggagagattc cgccacatcc
      601 ttctctctgg actttggcta cctgcggaac aagaatggct gccacgtgga gctgctgttc
      661 ctgagataca tctctgactg ggatctggac ccaggcaggt gttatcgcgt gacctggttc
      721 acaagctggt ccccctgcta cgattgtgca aggcacgtgg cagactttct gaggggaaac
      781 ccaaatctgt ccctgcggat cttcaccgcc agactgtatt tttgcgagga taggaaggca
      841 gagccagagg gactgaggcg cctgcacagg gccggcgtgc agatcgccat catgaccttc
      901 aaggactact tttattgttg gaacaccttc gtggagaatc acgagcggac cttcaaggcc
      961 tgggagggac tgcacgagaa ctccgtgcgg ctgtctagac agctgcggag aatcctgctg
     1021 cctctgtacg aggtggacga tctgagggat gccttccgca ccgagctgaa gacacccctg
     1081 ggcgacacca cacacacctc tccaccttgc ccagcaccag agctgctggg aggccctatg
     1141 gccagcaact tcacacagtt tgtgctggtg gataatggag gaaccggcga cgtgacagtg
     1201 gcaccatcta actttgccaa tggcatcgcc gagtggatca gctccaactc tcggagccag
     1261 gcctataagg tgacctgtag cgtgcggcag tctagcgccc agaatagaaa gtatacaatc
     1321 aaggtggagg tgcctaaggg cgcctggaga tcctacctga acatggagct gaccatccca
     1381 atctttgcca caaattctga ttgcgagctg atcgtgaagg ccatgcaggg cctgctgaag
     1441 gacggcaacc ctatcccaag cgccatcgcc gccaatagcg gaatctacgg atccggagag
     1501 ggaagaggat ccctgctgac ctgtggcgac gtggaggaga atccaggacc tggaacacca
     1561 aagaagaagc ggaaagtcga caagaagtac agcatcggcc tggccatcgg caccaactct
     1621 gtgggctggg ccgtgatcac cgacgagtac aaggtgccca gcaagaaatt caaggtgctg
     1681 ggcaacaccg accggcacag catcaagaag aacctgatcg gagccctgct gttcgacagc
     1741 ggcgaaacag ccgaggccac ccggctgaag agaaccgcca gaagaagata caccagacgg
     1801 aagaaccgga tctgctatct gcaagagatc ttcagcaacg agatggccaa ggtggacgac
     1861 agcttcttcc acagactgga agagtccttc ctggtggaag aggataagaa gcacgagcgg
     1921 caccccatct tcggcaacat cgtggacgag gtggcctacc acgagaagta ccccaccatc
     1981 taccacctga gaaagaaact ggtggacagc accgacaagg ccgacctgcg gctgatctat
     2041 ctggccctgg cccacatgat caagttccgg ggccacttcc tgatcgaggg cgacctgaac
     2101 cccgacaaca gcgacgtgga caagctgttc atccagctgg tgcagaccta caaccagctg
     2161 ttcgaggaaa accccatcaa cgccagcggc gtggacgcca aggccatcct gtctgccaga
     2221 ctgagcaaga gcagacggct ggaaaatctg atcgcccagc tgcccggcga gaagaagaat
     2281 ggcctgttcg gaaacctgat tgccctgagc ctgggcctga cccccaactt caagagcaac
     2341 ttcgacctgg ccgaggatgc caaactgcag ctgagcaagg acacctacga cgacgacctg
     2401 gacaacctgc tggcccagat cggcgaccag tacgccgacc tgtttctggc cgccaagaac
     2461 ctgtccgacg ccatcctgct gagcgacatc ctgagagtga acaccgagat caccaaggcc
```

2521 cccctgagcg cctctatgat caagagatac gacgagcacc accaggacct gaccctgctg
2581 aaagctctcg tgcggcagca gctgcctgag aagtacaaag agattttctt cgaccagagc
2641 aagaacggct acgccggcta cattgacggc ggagccagcc aggaagagtt ctacaagttc
2701 atcaagccca tcctggaaaa gatggacggc accgaggaac tgctcgtgaa gctgaacaga
2761 gaggacctgc tgcggaagca gcggaccttc gacaacggca gcatccccca ccagatccac
2821 ctgggagagc tgcacgccat tctgcggcgg caggaagatt tttacccatt cctgaaggac
2881 aaccgggaaa agatcgagaa gatcctgacc ttccgcatcc cctactacgt gggccctctg
2941 gccaggggaa acagcagatt cgcctggatg accagaaaga gcgaggaaac catcaccccc
3001 tggaacttcg aggaagtggt ggacaagggc gcttccgccc agagcttcat cgagcggatg
3061 accaacttcg ataagaacct gcccaacgag aaggtgctgc ccaagcacag cctgctgtac
3121 gagtacttca ccgtgtataa cgagctgacc aaagtgaaat acgtgaccga gggaatgaga
3181 aagcccgcct tcctgagcgg cgagcagaaa aaggccatcg tggacctgct gttcaagacc
3241 aaccggaaag tgaccgtgaa gcagctgaaa gaggactact tcaagaaaat cgagtgcttc
3301 gactccgtgg aaatctccgg cgtggaagat cggttcaacg cctccctggg cacataccac
3361 gatctgctga aaattatcaa ggacaaggac ttcctggaca atgaggaaaa cgaggacatt
3421 ctggaagata tcgtgctgac cctgacactg tttgaggaca gagagatgat cgaggaacgg
3481 ctgaaaacct atgcccacct gttcgacgac aaagtgatga agcagctgaa gcggcggaga
3541 tacaccggct ggggcaggct gagccggaag ctgatcaacg gcatccggga caagcagtcc
3601 ggcaagacaa tcctggattt cctgaagtcc gacggcttcg ccaacagaaa cttcatgcag
3661 ctgatccacg acgacagcct gacctttaaa gaggacatcc agaaagccca ggtgtccggc
3721 cagggcgata gcctgcacga gcacattgcc aatctggccg gcagccccgc cattaagaag
3781 ggcatcctgc agacagtgaa ggtggtggac gagctcgtga aagtgatggg ccggcacaag
3841 cccgagaaca tcgtgatcga aatggccaga gagaaccaga ccacccagaa gggacagaag
3901 aacagccgcg agagaatgaa gcggatcgaa gagggcatca aagagctggg cagccagatc
3961 ctgaaagaac accccgtgga aaacacccag ctgcagaacg agaagctgta cctgtactac
4021 ctgcagaatg ggcgggatat gtacgtggac caggaactgg acatcaaccg gctgtccgac
4081 tacgatgtgg accatatcgt gcctcagagc tttctgaagg acgactccat cgacaacaag
4141 gtgctgacca gaagcgacaa gaaccggggc aagagcgaca acgtgccctc cgaagaggtc
4201 gtgaagaaga tgaagaacta ctggcggcag ctgctgaacg ccaagctgat tacccagaga
4261 aagttcgaca atctgaccaa ggccgagaga ggcggcctga gcgaactgga taaggccggc
4321 ttcatcaaga gacagctggt ggaaacccgg cagatcacaa agcacgtggc acagatcctg
4381 gactcccgga tgaacactaa gtacgacgag aatgacaagc tgatccggga agtgaaagtg
4441 atcaccctga agtccaagct ggtgtccgat ttccggaagg atttccagtt ttacaaagtg
4501 cgcgagatca acaactacca ccacgcccac gacgcctacc tgaacgccgt cgtgggaacc
4561 gccctgatca aaaagtaccc taagctggaa agcgagttcg tgtacggcga ctacaaggtg
4621 tacgacgtgc ggaagatgat cgccaagagc gagcaggaaa tcggcaaggc taccgccaag
4681 tacttcttct acagcaacat catgaacttt ttcaagaccg agattaccct ggccaacggc
4741 gagatccgga agcggcctct gatcgagaca aacggcgaaa ccggggagat cgtgtgggat
4801 aagggccggg attttgccac cgtgcggaaa gtgctgagca tgccccaagt gaatatcgtg
4861 aaaaagaccg aggtgcagac aggcggcttc agcaaagagt ctatcctgcc caagaggaac
4921 agcgataagc tgatcgccag aaagaaggac tgggacccta agaagtacgg cggcttcgac
4981 agccccaccg tggcctattc tgtgctggtg gtggccaaag tggaaaaggg caagtccaag
5041 aaactgaaga gtgtgaaaga gctgctgggg atcaccatca tggaaagaag cagcttcgag
5101 aagaatccca tcgactttct ggaagccaag ggctacaaag aagtgaaaaa ggacctgatc
5161 atcaagctgc ctaagtactc cctgttcgag ctggaaaacg gccggaagag aatgctggcc
5221 tctgccggcg aactgcagaa gggaaacgaa ctggccctgc cctccaaata tgtgaacttc
5281 ctgtacctgg ccagccacta tgagaagctg aagggctccc ccgaggataa tgagcagaaa
5341 cagctgtttg tggaacagca caagcactac ctggacgaga tcatcgagca gatcagcgag
5401 ttctccaaga gagtgatcct ggccgacgct aatctggaca aagtgctgtc cgcctacaac
5461 aagcaccggg ataagcccat cagagagcag gccgagaata tcatccacct gtttaccctg
5521 accaatctgg gagcccctgc cgccttcaag tactttgaca ccaccatcga ccggaagagg
5581 tacaccagca ccaaagaggt gctggacgcc accctgatcc accagagcat caccggcctg
5641 tacgagacac ggatcgacct gtctcagctg ggaggtgaca gcggcgggag cggcgggagc
5701 ggggggagca ctaatctgag cgacatcatt gagaaggaga ctgggaaaca gctggtcatt
5761 caggagtcca tcctgatgct gcctgaggag gtggaggaag tgatcggcaa caagccagag
5821 tctgacatcc tggtgcacac cgcctacgac gagtccacag atgagaatgt gatgctgctg
5881 acctctgacg cccccgagta taagccttgg gccctggtca tccaggattc taacggcgag
5941 aataagatca agatgctgag cggaggatcc ggaggatctg gaggcagcac caacctgtct
6001 gacatcatcg agaaggagac aggcaagcag ctggtcatcc aggagagcat cctgatgctg
6061 cccgaagaag tcgaagaagt gatcggaaac aagcctgaga gcgatatcct ggtccatacc
6121 gcctacgacg agagtaccga cgaaaatgtg atgctgctga catccgacgc cccagagtat
6181 aagccctggg ctctggtcat ccaggattcc aacggagaga acaaaatcaa aatgctgtct
6241 ggcggctcaa aaagaaccgc cgacggcagc gaattcgagc ccaagaagaa gaggaaagtc

```
     6301 taaccggtca tcatcaccat caccattgag tttaaacccg ctgatcagcc tcgactgtgc
     6361 cttctagttg ccagccatct gttgtttgcc cctcccccgt gccttccttg accctggaag
     6421 gtgccactcc cactgtcctt tcctaataaa atgaggaaat tgcatcgcat tgtctgagta
     6481 ggtgtcattc tattctgggg ggtggggtgg ggcaggacag caagggggag gattgggaag
     6541 acaatagcag gcatgctggg gatgcggtgg gctctatggc ttctgaggcg gaaagaacca
     6601 gctggggctc gataccgtcg acctctagct agagcttggc gtaatcatgg tcatagctgt
     6661 ttcctgtgtg aaattgttat ccgctcacaa ttccacacaa catacgagcc ggaagcataa
     6721 agtgtaaagc ctagggtgcc taatgagtga gctaactcac attaattgcg ttgcgctcac
     6781 tgcccgcttt ccagtcggga aacctgtcgt gccagctgca ttaatgaatc ggccaacgcg
     6841 cggggagagg cggtttgcgt attgggcgct cttccgcttc ctcgctcact gactcgctgc
     6901 gctcggtcgt tcggctgcgg cgagcggtat cagctcactc aaaggcggta atacggttat
     6961 ccacagaatc aggggataac gcaggaaaga acatgtgagc aaaaggccag caaaaggcca
     7021 ggaaccgtaa aaaggccgcg ttgctggcgt ttttccatag gctccgcccc cctgacgagc
     7081 atcacaaaaa tcgacgctca agtcagaggt ggcgaaaccc gacaggacta taaagatacc
     7141 aggcgtttcc ccctggaagc tccctcgtgc gctctcctgt tccgaccctg ccgcttaccg
     7201 gatacctgtc cgcctttctc ccttcgggaa gcgtggcgct ttctcatagc tcacgctgta
     7261 ggtatctcag ttcggtgtag gtcgttcgct ccaagctggg ctgtgtgcac gaaccccccg
     7321 ttcagcccga ccgctgcgcc ttatccggta actatcgtct tgagtccaac ccggtaagac
     7381 acgacttatc gccactggca gcagccactg gtaacaggat tagcagagcg aggtatgtag
     7441 gcggtgctac agagttcttg aagtggtggc ctaactacgg ctacactaga agaacagtat
     7501 ttggtatctg cgctctgctg aagccagtta ccttcggaaa aagagttggt agctcttgat
     7561 ccggcaaaca aaccaccgct ggtagcggtg gtttttttgt ttgcaagcag cagattacgc
     7621 gcagaaaaaa aggatctcaa gaagatcctt tgatcttttc tacggggtct gacactcagt
     7681 ggaacgaaaa ctcacgttaa gggattttgg tcatgagatt atcaaaaagg atcttcacct
     7741 agatcctttt aaattaaaaa tgaagtttta aatcaatcta aagtatatat gagtaaactt
     7801 ggtctgacag ttaccaatgc ttaatcagtg aggcacctat ctcagcgatc tgtctatttc
     7861 gttcatccat agttgcctga ctccccgtcg tgtagataac tacgatacgg gagggcttac
     7921 catctggccc cagtgctgca atgataccgc gagacccacg ctcaccggct ccagatttat
     7981 cagcaataaa ccagccagcc ggaagggccg agcgcagaag tggtcctgca actttatccg
     8041 cctccatcca gtctattaat tgttgccggg aagctagagt aagtagttcg ccagttaata
     8101 gtttgcgcaa cgttgttgcc attgctacag gcatcgtggt gtcacgctcg tcgtttggta
     8161 tggcttcatt cagctccggt tcccaacgat caaggcgagt tacatgatcc cccatgttgt
     8221 gcaaaaaagc ggttagctcc ttcggtcctc cgatcgttgt cagaagtaag ttggccgcag
     8281 tgttatcact catggttatg gcagcactgc ataattctct tactgtcatg ccatccgtaa
     8341 gatgcttttc tgtgactggt gagtactcaa ccaagtcatt ctgagaatag tgtatgcggc
     8401 gaccgagttg ctcttgcccg gcgtcaatac gggataatac cgcgccacat agcagaactt
     8461 taaaagtgct catcattgga aaacgttctt cggggcgaaa actctcaagg atcttaccgc
     8521 tgttgagatc cagttcgatg taacccactc gtgcacccaa ctgatcttca gcatctttta
     8581 ctttcaccag cgtttctggg tgagcaaaaa caggaaggca aaatgccgca aaaaagggaa
     8641 taagggcgac acggaaatgt tgaatactca tactcttcct ttttcaatat tattgaagca
     8701 tttatcaggg ttattgtctc atgagcggat acatatttga atgtatttag aaaaataaac
     8761 aaataggggt tccgcgcaca tttccccgaa aagtgccacc tgacgtcgac ggatcgggag
     8821 atcgatctcc cgatccccta gggtcgactc tcagtacaat ctgctctgat gccgcatagt
     8881 taagccagta tctgctccct gcttgtgtgt tggaggtcgc tgagtagtgc gcgagcaaaa
     8941 tttaagctac aacaaggcaa ggcttgaccg acaattgcat gaagaatctg cttagggtta
     9001 ggcgttttgc gctgcttcgc gatgtacggg ccagatatac gcgttgacat tgattattga
     9061 ctagttatta atagtaatca attacggggt cattagttca tagcccatat atggagttcc
     9121 gcgttacata acttacggta aatggcccgc ctggctgacc gcccaacgac ccccgcccat
     9181 tgacgtcaat aatgacgtat gttcccatag taacgccaat agggactttc cattgacgtc
     9241 aatgggtgga gtatttacgg taaactgccc acttggcagt acatcaagtg tatc
//
```

9. pCMV_2xA_CRCnu.2

```
LOCUS       Exported                9994 bp ds-DNA     circular SYN 12-AUG-2018
DEFINITION  synthetic circular DNA
ACCESSION   .
VERSION     .
KEYWORDS    .
SOURCE      synthetic DNA construct
```

```
  ORGANISM  synthetic DNA construct
REFERENCE   1  (bases 1 to 9994)
  AUTHORS   Juan-Carlos Collantes
  TITLE     Direct Submission
  JOURNAL   Exported Aug 14, 2018 from SnapGene Viewer 4.2.3
            http://www.snapgene.com
FEATURES             Location/Qualifiers
     source          1..9994
                     /organism="synthetic DNA construct"
                     /mol_type="other DNA"
     promoter        132..335
                     /label=CMV promoter
                     /note="human cytomegalovirus (CMV) immediate early
                     promoter"
     misc_feature    353..2538
                     /label=Synthetic
     promoter        377..395
                     /label=T7 promoter
                     /note="promoter for bacteriophage T7 RNA polymerase"
     regulatory      445..454
                     /regulatory_class="other"
                     /note="vertebrate consensus sequence for strong initiation
                     of translation (Kozak, 1987)"
     CDS             457..477
                     /codon_start=1
                     /product="nuclear localization signal of SV40 (simian
virus
                     40) large T antigen"
                     /label=SV40 NLS
                     /translation="PKKKRKV"
     CDS             478..1071
                     /codon_start=1
                     /gene="AICDA"
                     /product="human activation-induced cytidine deaminase
                     (Dickerson et al., 2003)"
                     /label=AID

/translation="MDSLLMNRRKFLYQFKNVRWAKGRRETYLCYVVKRRDSATSFSLD

FGYLRNKNGCHVELLFLRYISDWDLDPGRCYRVTWFTSWSPCYDCARHVADFLRGNPNL

SLRIFTARLYFCEDRKAEPEGLRRLHRAGVQIAIMTFKDYFYCWNTFVENHERTFKAWE
                     GLHENSVRLSRQLRRILLPLYEVDDLRDAFRTLGL"
     CDS             1168..1761
                     /codon_start=1
                     /gene="AICDA"
                     /product="human activation-induced cytidine deaminase
                     (Dickerson et al., 2003)"
                     /label=AID

/translation="MDSLLMNRRKFLYQFKNVRWAKGRRETYLCYVVKRRDSATSFSLD

FGYLRNKNGCHVELLFLRYISDWDLDPGRCYRVTWFTSWSPCYDCARHVADFLRGNPNL

SLRIFTARLYFCEDRKAEPEGLRRLHRAGVQIAIMTFKDYFYCWNTFVENHERTFKAWE
                     GLHENSVRLSRQLRRILLPLYEVDDLRDAFRTLGL"
     CDS             1762..1836
                     /codon_start=1
                     /label=h3
                     /translation="ELKTPLGDTTHTSPPCPAPELLGGP"
     CDS             1837..2187
                     /codon_start=1
                     /label=MCP
```

```
/translation="MASNFTQFVLVDNGGTGDVTVAPSNFANGIAEWISSNSRSQAYKV

TCSVRQSSAQNRKYTIKVEVPKGAWRSYLNMELTIPIFATNSDCELIVKAMQGLLKDGN
                     PIPSAIAANSGIY"
     CDS             2197..2250
                     /codon_start=1
                     /product="2A peptide from Thosea asigna virus capsid
                     protein"
                     /label=T2A
                     /note="Eukaryotic ribosomes fail to insert a peptide bond
                     between the Gly and Pro residues, yielding separate
                     polypeptides."
                     /translation="EGRGSLLTCGDVEENPGP"
     CDS             2257..2277
                     /codon_start=1
                     /product="nuclear localization signal of SV40 (simian
virus
                     40) large T antigen"
                     /label=SV40 NLS
                     /translation="PKKKRKV"
     CDS             2278..6378
                     /codon_start=1
                     /product="nickase mutant of the Cas9 endonuclease from the
                     Streptococcus pyogenes Type II CRISPR/Cas system"
                     /label=Cas9(D10A)
                     /note="generates RNA-guided single strand nicks in DNA due
                     to the D10A mutation in the RuvC catalytic domain"

/translation="DKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKN

LIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESF

LVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKF

RGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLE

NLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQI

GDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQ

QLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLR

KQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGN

SRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYF

TVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDS

VEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERL

KTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQ

LIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRH

KPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYL

YYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPS

EEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHV

AQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLN
```

```
AVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTE

ITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKE

SILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGIT

IMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNEL

ALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADA

NLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLD
                     ATLIHQSITGLYETRIDLSQLGGD"
     CDS             6409..6657
                     /codon_start=1
                     /product="uracil-DNA glycosylase inhibitor from a Bacillus
                     subtilis bacteriophage (Mol et al., 1995)"
                     /label=UGI

/translation="TNLSDIIEKETGKQLVIQESILMLPEEVEEVIGNKPESDILVHTA
                     YDESTDENVMLLTSDAPEYKPWALVIQDSNGENKIKML"
     CDS             6688..6936
                     /codon_start=1
                     /product="uracil-DNA glycosylase inhibitor from a Bacillus
                     subtilis bacteriophage (Mol et al., 1995)"
                     /label=UGI

/translation="TNLSDIIEKETGKQLVIQESILMLPEEVEEVIGNKPESDILVHTA
                     YDESTDENVMLLTSDAPEYKPWALVIQDSNGENKIKML"
     CDS             6979..6999
                     /codon_start=1
                     /product="nuclear localization signal of SV40 (simian
virus
                     40) large T antigen"
                     /label=SV40 NLS
                     /translation="PKKKRKV"
     CDS             7008..7025
                     /codon_start=1
                     /product="6xHis affinity tag"
                     /label=6xHis
                     /translation="HHHHHH"
     polyA_signal    7054..7278
                     /label=bGH poly(A) signal
                     /note="bovine growth hormone polyadenylation signal"
     primer_bind     complement(7349..7365)
                     /label=M13 rev
                     /note="common sequencing primer, one of multiple similar
                     variants"
     protein_bind    7373..7389
                     /label=lac operator
                     /bound_moiety="lac repressor encoded by lacI"
                     /note="The lac repressor binds to the lac operator to
                     inhibit transcription in E. coli. This inhibition can be
                     relieved by adding lactose or
                     isopropyl-beta-D-thiogalactopyranoside (IPTG)."
     promoter        complement(7397..7427)
                     /label=lac promoter
                     /note="promoter for the E. coli lac operon"
     protein_bind    7442..7463
                     /label=CAP binding site
                     /bound_moiety="E. coli catabolite activator protein"
                     /note="CAP binding activates transcription in the presence
                     of cAMP."
```

```
     rep_origin      complement(7751..8339)
                     /direction=LEFT
                     /label=ori
                     /note="high-copy-number ColE1/pMB1/pBR322/pUC origin of
                     replication"
     CDS             complement(8510..9370)
                     /codon_start=1
                     /gene="bla"
                     /product="beta-lactamase"
                     /label=AmpR
                     /note="confers resistance to ampicillin, carbenicillin,
and
                     related antibiotics"

/translation="MSIQHFRVALIPFFAAFCLPVFAHPETLVKVKDAEDQLGARVGYI

ELDLNSGKILESFRPEERFPMMSTFKVLLCGAVLSRIDAGQEQLGRRIHYSQNDLVEYS

PVTEKHLTDGMTVRELCSAAITMSDNTAANLLLTTIGGPKELTAFLHNMGDHVTRLDRW

EPELNEAIPNDERDTTMPVAMATTLRKLLTGELLTLASRQQLIDWMEADKVAGPLLRSA

LPAGWFIADKSGAGERGSRGIIAALGPDGKPSRIVVIYTTGSQATMDERNRQIAEIGAS
                     LIKHW"
     promoter        complement(9371..9475)
                     /gene="bla"
                     /label=AmpR promoter
     enhancer        join(9745..9994,1..131)
                     /label=CMV enhancer
                     /note="human cytomegalovirus immediate early enhancer"
ORIGIN
        1 atatgccaag tacgccccct attgacgtca atgacggtaa atggcccgcc tggcattatg
       61 cccagtacat gaccttatgg gactttccta cttggcagta catctacgta ttagtcatcg
      121 ctattaccat ggtgatgcgg ttttggcagt acatcaatgg gcgtggatag cggtttgact
      181 cacggggatt tccaagtctc caccccattg acgtcaatgg gagtttgttt tggcaccaaa
      241 atcaacggga ctttccaaaa tgtcgtaaca actccgcccc attgacgcaa atgggcggta
      301 ggcgtgtacg gtgggaggtc tatataagca gagctggttt agtgaaccgt cagatccgct
      361 agagatccgc ggccgctaat acgactcact atagggagag ccgccaccaa caaacggaca
      421 gccgacggaa gcgagttcga gtcagccacc atggccccca agaagaagcg gaaagtgatg
      481 gactctttac tgatgaatcg taggaagttt ttataccagt tcaagaacgt gaggtgggcc
      541 aagggaagga gggagaccta tttatgctac gtggtgaaga ggagagacag cgctaccagc
      601 ttctctttag acttcggcta tttacgtaac aagaacggct gccacgtgga gctgctgttt
      661 ttaaggtaca tcagcgattg ggatttagac cccggtcgtt gttatcgtgt gacttggttc
      721 acctcttgga gcccttgtta cgattgcgct cgtcacgtcg ccgactttct gaggggcaac
      781 cccaacctct ctttaagaat cttcaccgct cgtctgtact tctgcgagga tcgtaaagcc
      841 gagcccgaag gcctcagaag gctgcatcgt gccggcgtgc agattgccat catgaccttt
      901 aaggattact tctactgctg gaacaccttc gtcgagaatc acgagaggac cttcaaggct
      961 tgggagggtt tacacgagaa cagcgtgagg ctgtctcgtc agctgaggag gattctgctg
     1021 cctctgtacg aggtggacga tttaagggac gcctttcgta ctttaggact cagcggagga
     1081 agctctggcg gaagcagcgg aagcgaaacc cccggtacaa gcgagagcgc tacccccgaa
     1141 tccagcggcg gtagcagcgg cggcagcatg gatagcctgc tgatgaaccg gagaaagttc
     1201 ctgtatcagt ttaagaatgt gcgctgggca aagggcaggc gcgagaccta cctgtgctat
     1261 gtggtgaagc ggagagattc cgccacatcc ttctctctgg actttggcta cctgcggaac
     1321 aagaatggct gccacgtgga gctgctgttc ctgagataca tctctgactg ggatctggac
     1381 ccaggcaggt gttatcgcgt gacctggttc acaagctggt ccccctgcta cgattgtgca
     1441 aggcacgtgg cagactttct gaggggaaac ccaaatctgt ccctgcggat cttcaccgcc
     1501 agactgtatt tttgcgagga taggaaggca gagccagagg gactgaggcg cctgcacagg
     1561 gccggcgtgc agatcgccat catgaccttc aaggactact tttattgttg gaacaccttc
     1621 gtggagaatc acgagcggac cttcaaggcc tgggagggac tgcacgagaa ctccgtgcgg
     1681 ctgtctagac agctgcggag aatcctgctg cctctgtacg aggtggacga tctgagggat
     1741 gccttccgca ccctgggact ggagctgaag acacccctgg gcgacaccac acacacctct
     1801 ccaccttgcc cagcaccaga gctgctggga ggccctatgg ccagcaactt cacacagttt
     1861 gtgctggtgg ataatggagg aaccggcgac gtgacagtgg caccatctaa ctttgccaat
```

1921 ggcatcgccg agtggatcag ctccaactct cggagccagg cctataaggt gacctgtagc
1981 gtgcggcagt ctagcgccca gaatagaaag tatacaatca aggtggaggt gcctaagggc
2041 gcctggagat cctacctgaa catggagctg accatcccaa tctttgccac aaattctgat
2101 tgcgagctga tcgtgaaggc catgcagggc ctgctgaagg acggcaaccc tatcccaagc
2161 gccatcgccg ccaatagcgg aatctacgga tccggagagg gaagaggatc cctgctgacc
2221 tgtggcgacg tggaggagaa tccaggacct ggaacaccaa agaagaagcg gaaagtcgac
2281 aagaagtaca gcatcggcct ggccatcggc accaactctg tgggctgggc cgtgatcacc
2341 gacgagtaca aggtgcccag caagaaattc aaggtgctgg gcaacaccga ccggcacagc
2401 atcaagaaga acctgatcgg agccctgctg ttcgacagcg gcgaaacagc cgaggccacc
2461 cggctgaaga gaaccgccag aagaagatac accagacgga agaaccggat ctgctatctg
2521 caagagatct tcagcaacga gatggccaag gtggacgaca gcttcttcca cagactggaa
2581 gagtccttcc tggtggaaga ggataagaag cacgagcggc accccatctt cggcaacatc
2641 gtggacgagg tggcctacca cgagaagtac cccaccatct accacctgag aaagaaactg
2701 gtggacagca ccgacaaggc cgacctgcgg ctgatctatc tggccctggc ccacatgatc
2761 aagttccggg gccacttcct gatcgagggc gacctgaacc ccgacaacag cgacgtggac
2821 aagctgttca tccagctggt gcagacctac aaccagctgt tcgaggaaaa ccccatcaac
2881 gccagcggcg tggacgccaa ggccatcctg tctgccagac tgagcaagag cagacggctg
2941 gaaaatctga tcgcccagct gcccggcgag aagaagaatg gcctgttcgg aaacctgatt
3001 gccctgagcc tgggcctgac ccccaacttc aagagcaact tcgacctggc cgaggatgcc
3061 aaactgcagc tgagcaagga cacctacgac gacgacctgg acaacctgct ggcccagatc
3121 ggcgaccagt acgccgacct gtttctggcc gccaagaacc tgtccgacgc catcctgctg
3181 agcgacatcc tgagagtgaa caccgagatc accaaggccc ccctgagcgc ctctatgatc
3241 aagagatacg acgagcacca ccaggacctg accctgctga aagctctcgt gcggcagcag
3301 ctgcctgaga agtacaaaga gattttcttc gaccagagca agaacggcta cgccggctac
3361 attgacggcg gagccagcca ggaagagttc tacaagttca tcaagcccat cctggaaaag
3421 atggacggca ccgaggaact gctcgtgaag ctgaacagag aggacctgct gcggaagcag
3481 cggaccttcg acaacggcag catcccccac cagatccacc tgggagagct gcacgccatt
3541 ctgcggcggc aggaagattt ttacccattc ctgaaggaca accgggaaaa gatcgagaag
3601 atcctgacct tccgcatccc ctactacgtg ggccctctgg ccaggggaaa cagcagattc
3661 gcctggatga ccagaaagag cgaggaaacc atcacccccct ggaacttcga ggaagtggtg
3721 gacaagggcg cttccgccca gagcttcatc gagcggatga ccaacttcga taagaacctg
3781 cccaacgaga aggtgctgcc caagcacagc ctgctgtacg agtacttcac cgtgtataac
3841 gagctgacca aagtgaaata cgtgaccgag ggaatgagaa agcccgcctt cctgagcggc
3901 gagcagaaaa aggccatcgt ggacctgctg ttcaagacca accggaaagt gaccgtgaag
3961 cagctgaaag aggactactt caagaaaatc gagtgcttcg actccgtgga aatctccggc
4021 gtggaagatc ggttcaacgc ctccctgggc acataccacg atctgctgaa aattatcaag
4081 gacaaggact tcctggacaa tgaggaaaac gaggacattc tggaagatat cgtgctgacc
4141 ctgacactgt ttgaggacag agagatgatc gaggaacggc tgaaaaccta tgcccacctg
4201 ttcgacgaca aagtgatgaa gcagctgaag cggcggagat acaccggctg gggcaggctg
4261 agccggaagc tgatcaacgg catccgggac aagcagtccg gcaagacaat cctggatttc
4321 ctgaagtccg acggcttcgc caacagaaac ttcatgcagc tgatccacga cgacagcctg
4381 acctttaaag aggacatcca gaaagcccag gtgtccggcc agggcgatag cctgcacgag
4441 cacattgcca atctggccgg cagccccgcc attaagaagg gcatcctgca gacagtgaag
4501 gtggtggacg agctcgtgaa agtgatgggc cggcacaagc ccgagaacat cgtgatcgaa
4561 atggccagag agaaccagac cacccagaag ggacagaaga acagccgcga gagaatgaag
4621 cggatcgaag agggcatcaa agagctgggc agccagatcc tgaaagaaca ccccgtggaa
4681 aacacccagc tgcagaacga gaagctgtac ctgtactacc tgcagaatgg gcgggatatg
4741 tacgtggacc aggaactgga catcaaccgg ctgtccgact acgatgtgga ccatatcgtg
4801 cctcagagct ttctgaagga cgactccatc gacaacaagg tgctgaccag aagcgacaag
4861 aaccggggca agagcgacaa cgtgccctcc gaagaggtcg tgaagaagat gaagaactac
4921 tggcggcagc tgctgaacgc caagctgatt acccagagaa agttcgacaa tctgaccaag
4981 gccgagagag gcggcctgag cgaactggat aaggccggct tcatcaagag acagctggtg
5041 gaaacccggc agatcacaaa gcacgtggca cagatcctgg actcccggat gaacactaag
5101 tacgacgaga atgacaagct gatccgggaa gtgaaagtga tcaccctgaa gtccaagctg
5161 gtgtccgatt tccggaagga tttccagttt tacaaagtgc gcgagatcaa caactaccac
5221 cacgcccacg acgcctacct gaacgccgtc gtgggaaccg ccctgatcaa aaagtaccct
5281 aagctggaaa gcgagttcgt gtacggcgac tacaaggtgt acgacgtgcg gaagatgatc
5341 gccaagagcg agcaggaaat cggcaaggct accgccaagt acttcttcta cagcaacatc
5401 atgaactttt tcaagaccga gattaccctg gccaacggcg agatccggaa gcggcctctg
5461 atcgagacaa acggcgaaac cggggagatc gtgtgggata agggccggga ttttgccacc
5521 gtgcggaaag tgctgagcat gccccaagtg aatatcgtga aaaagaccga ggtgcagaca
5581 ggcggcttca gcaaagagtc tatcctgccc aagaggaaca gcgataagct gatcgccaga
5641 aagaaggact gggaccctaa gaagtacggc ggcttcgaca gccccaccgt ggcctattct

5701 gtgctggtgg tggccaaagt ggaaaagggc aagtccaaga aactgaagag tgtgaaagag
5761 ctgctgggga tcaccatcat ggaaagaagc agcttcgaga agaatcccat cgactttctg
5821 gaagccaagg gctacaaaga agtgaaaaag gacctgatca tcaagctgcc taagtactcc
5881 ctgttcgagc tggaaaacgg ccggaagaga atgctggcct ctgccggcga actgcagaag
5941 ggaaacgaac tggccctgcc ctccaaatat gtgaacttcc tgtacctggc cagccactat
6001 gagaagctga agggctcccc cgaggataat gagcagaaac agctgtttgt ggaacagcac
6061 aagcactacc tggacgagat catcgagcag atcagcgagt tctccaagag agtgatcctg
6121 gccgacgcta atctggacaa agtgctgtcc gcctacaaca agcaccggga taagcccatc
6181 agagagcagg ccgagaatat catccacctg tttaccctga ccaatctggg agcccctgcc
6241 gccttcaagt actttgacac caccatcgac cggaagaggt acaccagcac caaagaggtg
6301 ctggacgcca ccctgatcca ccagagcatc accggcctgt acgagacacg gatcgacctg
6361 tctcagctgg gaggtgacag cggcgggagc ggcgggagcg gggggagcac taatctgagc
6421 gacatcattg agaaggagac tgggaaacag ctggtcattc aggagtccat cctgatgctg
6481 cctgaggagg tggaggaagt gatcggcaac aagccagagt ctgacatcct ggtgcacacc
6541 gcctacgacg agtccacaga tgagaatgtg atgctgctga cctctgacgc ccccgagtat
6601 aagccttggg ccctggtcat ccaggattct aacggcgaga ataagatcaa gatgctgagc
6661 ggaggatccg gaggatctgg aggcagcacc aacctgtctg acatcatcga gaaggagaca
6721 ggcaagcagc tggtcatcca ggagagcatc ctgatgctgc ccgaagaagt cgaagaagtg
6781 atcggaaaca agcctgagag cgatatcctg gtccataccg cctacgacga gagtaccgac
6841 gaaaatgtga tgctgctgac atccgacgcc ccagagtata agccctgggc tctggtcatc
6901 caggattcca acggagagaa caaaatcaaa atgctgtctg gcggctcaaa aagaaccgcc
6961 gacggcagcg aattcgagcc caagaagaag aggaaagtct aaccggtcat catcaccatc
7021 accattgagt ttaaacccgc tgatcagcct cgactgtgcc ttctagttgc cagccatctg
7081 ttgtttgccc ctccccccgtg ccttccttga ccctggaagg tgccactccc actgtccttt
7141 cctaataaaa tgaggaaatt gcatcgcatt gtctgagtag gtgtcattct attctggggg
7201 gtggggtggg gcaggacagc aagggggagg attgggaaga caatagcagg catgctgggg
7261 atgcggtggg ctctatggct tctgaggcgg aaagaaccag ctggggctcg ataccgtcga
7321 cctctagcta gagcttggcg taatcatggt catagctgtt tcctgtgtga aattgttatc
7381 cgctcacaat tccacacaac atacgagccg gaagcataaa gtgtaaagcc tagggtgcct
7441 aatgagtgag ctaactcaca ttaattgcgt tgcgctcact gcccgctttc cagtcgggaa
7501 acctgtcgtg ccagctgcat taatgaatcg gccaacgcgc ggggagaggc ggtttgcgta
7561 ttgggcgctc ttccgcttcc tcgctcactg actcgctgcg ctcggtcgtt cggctgcggc
7621 gagcggtatc agctcactca aaggcggtaa tacggttatc cacagaatca ggggataacg
7681 caggaaagaa catgtgagca aaaggccagc aaaaggccag gaaccgtaaa aaggccgcgt
7741 tgctggcgtt tttccatagg ctccgccccc ctgacgagca tcacaaaaat cgacgctcaa
7801 gtcagaggtg gcgaaacccg acaggactat aaagatacca ggcgtttccc cctggaagct
7861 ccctcgtgcg ctctcctgtt ccgaccctgc cgcttaccgg atacctgtcc gcctttctcc
7921 cttcgggaag cgtggcgctt tctcatagct cacgctgtag gtatctcagt tcggtgtagg
7981 tcgttcgctc caagctgggc tgtgtgcacg aaccccccgt tcagcccgac cgctgcgcct
8041 tatccggtaa ctatcgtctt gagtccaacc cggtaagaca cgacttatcg ccactggcag
8101 cagccactgg taacaggatt agcagagcga ggtatgtagg cggtgctaca gagttcttga
8161 agtggtggcc taactacggc tacactagaa gaacagtatt tggtatctgc gctctgctga
8221 agccagttac cttcggaaaa agagttggta gctcttgatc cggcaaacaa accaccgctg
8281 gtagcggtgg tttttttgtt tgcaagcagc agattacgcg cagaaaaaaa ggatctcaag
8341 aagatccttt gatcttttct acggggtctg acactcagtg gaacgaaaac tcacgttaag
8401 ggattttggt catgagatta tcaaaaagga tcttcaccta gatcctttta aattaaaaat
8461 gaagttttaa atcaatctaa agtatatatg agtaaacttg gtctgacagt taccaatgct
8521 taatcagtga ggcacctatc tcagcgatct gtctatttcg ttcatccata gttgcctgac
8581 tccccgtcgt gtagataact acgatacggg agggcttacc atctggcccc agtgctgcaa
8641 tgataccgcg agacccacgc tcaccggctc cagatttatc agcaataaac cagccagccg
8701 gaagggccga gcgcagaagt ggtcctgcaa ctttatccgc ctccatccag tctattaatt
8761 gttgccggga agctagagta agtagttcgc cagttaatag tttgcgcaac gttgttgcca
8821 ttgctacagg catcgtggtg tcacgctcgt cgtttggtat ggcttcattc agctccggtt
8881 cccaacgatc aaggcgagtt acatgatccc ccatgttgtg caaaaaagcg gttagctcct
8941 tcggtcctcc gatcgttgtc agaagtaagt tggccgcagt gttatcactc atggttatgg
9001 cagcactgca taattctctt actgtcatgc catccgtaag atgcttttct gtgactggtg
9061 agtactcaac caagtcattc tgagaatagt gtatgcggcg accgagttgc tcttgcccgg
9121 cgtcaatacg ggataatacc gcgccacata gcagaacttt aaaagtgctc atcattggaa
9181 aacgttcttc ggggcgaaaa ctctcaagga tcttaccgct gttgagatcc agttcgatgt
9241 aacccactcg tgcacccaac tgatcttcag catcttttac tttcaccagc gtttctgggt
9301 gagcaaaaac aggaaggcaa aatgccgcaa aaaagggaat aagggcgaca cggaaatgtt
9361 gaatactcat actcttcctt tttcaatatt attgaagcat ttatcagggt tattgtctca
9421 tgagcggata catatttgaa tgtatttaga aaaataaaca aataggggtt ccgcgcacat

```
     9481 ttccccgaaa agtgccacct gacgtcgacg gatcgggaga tcgatctccc gatcccctag
     9541 ggtcgactct cagtacaatc tgctctgatg ccgcatagtt aagccagtat ctgctccctg
     9601 cttgtgtgtt ggaggtcgct gagtagtgcg cgagcaaaat ttaagctaca acaaggcaag
     9661 gcttgaccga caattgcatg aagaatctgc ttagggttag gcgttttgcg ctgcttcgcg
     9721 atgtacgggc cagatatacg cgttgacatt gattattgac tagttattaa tagtaatcaa
     9781 ttacggggtc attagttcat agcccatata tggagttccg cgttacataa cttacggtaa
     9841 atggcccgcc tggctgaccg cccaacgacc cccgcccatt gacgtcaata atgacgtatg
     9901 ttcccatagt aacgccaata gggactttcc attgacgtca atgggtggag tatttacggt
     9961 aaactgccca cttggcagta catcaagtgt atcn
//
```

10. pCMV_^{A1}CRCnu.2

```
LOCUS       Exported                9393 bp ds-DNA     circular SYN 12-AUG-2018
DEFINITION  synthetic circular DNA
ACCESSION   .
VERSION     .
KEYWORDS    .
SOURCE      synthetic DNA construct
  ORGANISM  synthetic DNA construct
REFERENCE   1  (bases 1 to 9393)
  AUTHORS   Juan-Carlos Collantes
  TITLE     Direct Submission
  JOURNAL   Exported Aug 14, 2018 from SnapGene Viewer 4.2.3
            http://www.snapgene.com
FEATURES             Location/Qualifiers
     source          1..9393
                     /organism="synthetic DNA construct"
                     /mol_type="other DNA"
     promoter        132..335
                     /label=CMV promoter
                     /note="human cytomegalovirus (CMV) immediate early
                     promoter"
     misc_feature    353..1938
                     /label=Synthetic
     promoter        377..395
                     /label=T7 promoter
                     /note="promoter for bacteriophage T7 RNA polymerase"
     regulatory      445..454
                     /regulatory_class="other"
                     /note="vertebrate consensus sequence for strong initiation
                     of translation (Kozak, 1987)"
     CDS             457..477
                     /codon_start=1
                     /product="nuclear localization signal of SV40 (simian
virus
                     40) large T antigen"
                     /label=SV40 NLS
                     /translation="PKKKRKV"
     CDS             478..1161
                     /codon_start=1
                     /product="cytidine deaminase (C to U editing enzyme) from
                     rat"
                     /label=APOBEC-1
                     /note="can use ssDNA as a substrate (Komor et al., 2016)"

/translation="SSETGPVAVDPTLRRRIEPHEFEVFFDPRELRKETCLLYEINWGG

RHSIWRHTSQNTNKHVEVNFIEKFTTERYFCPNTRCSITWFLSWSPCGECSRAITEFLS
```

```
RYPHVTLFIYIARLYHHADPRNRQGLRDLISSGVTIQIMTEQESGYCWRNFVNYSPSNE

AHWPRYPHLWVRLYVLELYCIILGLPPCLNILRRKQPQLTFFTIALQSCHYQRLPPHIL
                     WATGLK"
     CDS             1162..1236
                     /codon_start=1
                     /label=h3
                     /translation="ELKTPLGDTTHTSPPCPAPELLGGP"
     CDS             1237..1587
                     /codon_start=1
                     /label=MCP

/translation="MASNFTQFVLVDNGGTGDVTVAPSNFANGIAEWISSNSRSQAYKV

TCSVRQSSAQNRKYTIKVEVPKGAWRSYLNMELTIPIFATNSDCELIVKAMQGLLKDGN
                     PIPSAIAANSGIY"
     CDS             1597..1650
                     /codon_start=1
                     /product="2A peptide from Thosea asigna virus capsid
                     protein"
                     /label=T2A
                     /note="Eukaryotic ribosomes fail to insert a peptide bond
                     between the Gly and Pro residues, yielding separate
                     polypeptides."
                     /translation="EGRGSLLTCGDVEENPGP"
     CDS             1657..1677
                     /codon_start=1
                     /product="nuclear localization signal of SV40 (simian
virus
                     40) large T antigen"
                     /label=SV40 NLS
                     /translation="PKKKRKV"
     CDS             1678..5778
                     /codon_start=1
                     /product="nickase mutant of the Cas9 endonuclease from the
                     Streptococcus pyogenes Type II CRISPR/Cas system"
                     /label=Cas9(D10A)
                     /note="generates RNA-guided single strand nicks in DNA due
                     to the D10A mutation in the RuvC catalytic domain"

/translation="DKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKN

LIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESF

LVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKF

RGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLE

NLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQI

GDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQ

QLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLR

KQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGN

SRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYF

TVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDS

VEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERL
```

```
KTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQ

LIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRH

KPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYL

YYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPS

EEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHV

AQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLN

AVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTE

ITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKE

SILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGIT

IMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNEL

ALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADA

NLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLD
                     ATLIHQSITGLYETRIDLSQLGGD"
     CDS             5809..6057
                     /codon_start=1
                     /product="uracil-DNA glycosylase inhibitor from a Bacillus
                     subtilis bacteriophage (Mol et al., 1995)"
                     /label=UGI

/translation="TNLSDIIEKETGKQLVIQESILMLPEEVEEVIGNKPESDILVHTA
                     YDESTDENVMLLTSDAPEYKPWALVIQDSNGENKIKML"
     CDS             6088..6336
                     /codon_start=1
                     /product="uracil-DNA glycosylase inhibitor from a Bacillus
                     subtilis bacteriophage (Mol et al., 1995)"
                     /label=UGI

/translation="TNLSDIIEKETGKQLVIQESILMLPEEVEEVIGNKPESDILVHTA
                     YDESTDENVMLLTSDAPEYKPWALVIQDSNGENKIKML"
     CDS             6379..6399
                     /codon_start=1
                     /product="nuclear localization signal of SV40 (simian
virus
                     40) large T antigen"
                     /label=SV40 NLS
                     /translation="PKKKRKV"
     CDS             6408..6425
                     /codon_start=1
                     /product="6xHis affinity tag"
                     /label=6xHis
                     /translation="HHHHHH"
     polyA_signal    6454..6678
                     /label=bGH poly(A) signal
                     /note="bovine growth hormone polyadenylation signal"
     primer_bind     complement(6749..6765)
                     /label=M13 rev
                     /note="common sequencing primer, one of multiple similar
                     variants"
     protein_bind    6773..6789
                     /label=lac operator
                     /bound_moiety="lac repressor encoded by lacI"
```

```
                     /note="The lac repressor binds to the lac operator to
                     inhibit transcription in E. coli. This inhibition can be
                     relieved by adding lactose or
                     isopropyl-beta-D-thiogalactopyranoside (IPTG)."
     promoter        complement(6797..6827)
                     /label=lac promoter
                     /note="promoter for the E. coli lac operon"
     protein_bind    6842..6863
                     /label=CAP binding site
                     /bound_moiety="E. coli catabolite activator protein"
                     /note="CAP binding activates transcription in the presence
                     of cAMP."
     rep_origin      complement(7151..7739)
                     /direction=LEFT
                     /label=ori
                     /note="high-copy-number ColE1/pMB1/pBR322/pUC origin of
                     replication"
     CDS             complement(7910..8770)
                     /codon_start=1
                     /gene="bla"
                     /product="beta-lactamase"
                     /label=AmpR
                     /note="confers resistance to ampicillin, carbenicillin,
and
                     related antibiotics"

/translation="MSIQHFRVALIPFFAAFCLPVFAHPETLVKVKDAEDQLGARVGYI

ELDLNSGKILESFRPEERFPMMSTFKVLLCGAVLSRIDAGQEQLGRRIHYSQNDLVEYS

PVTEKHLTDGMTVRELCSAAITMSDNTAANLLLTTIGGPKELTAFLHNMGDHVTRLDRW

EPELNEAIPNDERDTTMPVAMATTLRKLLTGELLTLASRQQLIDWMEADKVAGPLLRSA

LPAGWFIADKSGAGERGSRGIIAALGPDGKPSRIVVIYTTGSQATMDERNRQIAEIGAS
                     LIKHW"
     promoter        complement(8771..8875)
                     /gene="bla"
                     /label=AmpR promoter
     enhancer        join(9145..9393,1..131)
                     /label=CMV enhancer
                     /note="human cytomegalovirus immediate early enhancer"
ORIGIN
        1 atatgccaag tacgccccct attgacgtca atgacggtaa atggcccgcc tggcattatg
       61 cccagtacat gaccttatgg gactttccta cttggcagta catctacgta ttagtcatcg
      121 ctattaccat ggtgatgcgg ttttggcagt acatcaatgg gcgtggatag cggtttgact
      181 cacggggatt tccaagtctc caccccattg acgtcaatgg gagtttgttt tggcaccaaa
      241 atcaacggga ctttccaaaa tgtcgtaaca actccgcccc attgacgcaa atgggcggta
      301 ggcgtgtacg gtgggaggtc tatataagca gagctggttt agtgaaccgt cagatccgct
      361 agagatccgc ggccgctaat acgactcact atagggagag ccgccaccaa caaacggaca
      421 gccgacggaa gcgagttcga gtcagccacc atggccccca agaagaagcg gaaagtgtcc
      481 tcagagactg ggcctgtcgc cgtcgatcca accctgcgcc gccggattga acctcacgag
      541 tttgaagtgt tctttgaccc ccgggagctg agaaaggaga catgcctgct gtacgagatc
      601 aactggggag gcaggcactc catctggagg cacacctctc agaacacaaa taagcacgtg
      661 gaggtgaact tcatcgagaa gtttaccaca gagcggtact tctgccccaa taccagatgt
      721 agcatcacat ggtttctgag ctggtcccct tgcggagagt gtagcagggc catcaccgag
      781 ttcctgtcca gatatccaca cgtgacactg tttatctaca tcgccaggct gtatcaccac
      841 gcagacccaa ggaataggca gggcctgcgc gatctgatca gctccggcgt gaccatccag
      901 atcatgacag agcaggagtc cggctactgc tggcggaact tcgtgaatta ttctcctagc
      961 aacgaggccc actggcctag gtacccacac ctgtgggtgc gcctgtacgt gctggagctg
     1021 tattgcatca tcctgggcct gcccccttgt ctgaatatcc tgcggagaaa gcagccccag
     1081 ctgaccttct ttacaatcgc cctgcagtct tgtcactatc agaggctgcc accccacatc
     1141 ctgtgggcca caggcctgaa ggagctgaag acacccctgg gcgacaccac acacacctct
```

1201 ccaccttgcc cagcaccaga gctgctggga ggccctatgg ccagcaactt cacacagttt
1261 gtgctggtgg ataatggagg aaccggcgac gtgacagtgg caccatctaa ctttgccaat
1321 ggcatcgccg agtggatcag ctccaactct cggagccagg cctataaggt gacctgtagc
1381 gtgcggcagt ctagcgccca gaatagaaag tatacaatca aggtggaggt gcctaagggc
1441 gcctggagat cctacctgaa catggagctg accatcccaa tctttgccac aaattctgat
1501 tgcgagctga tcgtgaaggc catgcagggc ctgctgaagg acggcaaccc tatcccaagc
1561 gccatcgccg ccaatagcgg aatctacgga tccggagagg gaagaggatc cctgctgacc
1621 tgtggcgacg tggaggagaa tccaggacct ggaacaccaa agaagaagcg gaaagtcgac
1681 aagaagtaca gcatcggcct ggccatcggc accaactctg tgggctgggc cgtgatcacc
1741 gacgagtaca aggtgcccag caagaaattc aaggtgctgg gcaacaccga ccggcacagc
1801 atcaagaaga acctgatcgg agccctgctg ttcgacagcg gcgaaacagc cgaggccacc
1861 cggctgaaga gaaccgccag aagaagatac accagacgga agaaccggat ctgctatctg
1921 caagagatct tcagcaacga gatggccaag gtggacgaca gcttcttcca cagactggaa
1981 gagtccttcc tggtggaaga ggataagaag cacgagcggc accccatctt cggcaacatc
2041 gtggacgagg tggcctacca cgagaagtac cccaccatct accacctgag aaagaaactg
2101 gtggacagca ccgacaaggc cgacctgcgg ctgatctatc tggccctggc ccacatgatc
2161 aagttccggg gccacttcct gatcgagggc gacctgaacc ccgacaacag cgacgtggac
2221 aagctgttca tccagctggt gcagacctac aaccagctgt tcgaggaaaa ccccatcaac
2281 gccagcggcg tggacgccaa ggccatcctg tctgccagac tgagcaagag cagacggctg
2341 gaaaatctga tcgcccagct gcccggcgag aagaagaatg gcctgttcgg aaacctgatt
2401 gccctgagcc tgggcctgac ccccaacttc aagagcaact tcgacctggc cgaggatgcc
2461 aaactgcagc tgagcaagga cacctacgac gacgacctgg acaacctgct ggcccagatc
2521 ggcgaccagt acgccgacct gtttctggcc gccaagaacc tgtccgacgc catcctgctg
2581 agcgacatcc tgagagtgaa caccgagatc accaaggccc ccctgagcgc ctctatgatc
2641 aagagatacg acgagcacca ccaggacctg accctgctga aagctctcgt gcggcagcag
2701 ctgcctgaga agtacaaaga gattttcttc gaccagagca agaacggcta cgccggctac
2761 attgacggcg gagccagcca ggaagagttc tacaagttca tcaagcccat cctggaaaag
2821 atggacggca ccgaggaact gctcgtgaag ctgaacagag aggacctgct gcggaagcag
2881 cggaccttcg acaacggcag catcccccac cagatccacc tgggagagct gcacgccatt
2941 ctgcggcggc aggaagattt ttacccattc ctgaaggaca accgggaaaa gatcgagaag
3001 atcctgacct tccgcatccc ctactacgtg ggccctctgg ccaggggaaa cagcagattc
3061 gcctggatga ccagaaagag cgaggaaacc atcacccccct ggaacttcga ggaagtggtg
3121 gacaagggcg cttccgccca gagcttcatc gagcggatga ccaacttcga taagaacctg
3181 cccaacgaga aggtgctgcc caagcacagc ctgctgtacg agtacttcac cgtgtataac
3241 gagctgacca aagtgaaata cgtgaccgag ggaatgagaa agcccgcctt cctgagcggc
3301 gagcagaaaa aggccatcgt ggacctgctg ttcaagacca accggaaagt gaccgtgaag
3361 cagctgaaag aggactactt caagaaaatc gagtgcttcg actccgtgga aatctccggc
3421 gtggaagatc ggttcaacgc ctccctgggc acataccacg atctgctgaa aattatcaag
3481 gacaaggact tcctggacaa tgaggaaaac gaggacattc tggaagatat cgtgctgacc
3541 ctgacactgt ttgaggacag agagatgatc gaggaacggc tgaaaaccta tgcccacctg
3601 ttcgacgaca aagtgatgaa gcagctgaag cggcggagat acaccggctg gggcaggctg
3661 agccggaagc tgatcaacgg catccgggac aagcagtccg gcaagacaat cctggatttc
3721 ctgaagtccg acggcttcgc caacagaaac ttcatgcagc tgatccacga cgacagcctg
3781 acctttaaag aggacatcca gaaagcccag gtgtccggcc agggcgatag cctgcacgag
3841 cacattgcca atctggccgg cagccccgcc attaagaagg gcatcctgca gacagtgaag
3901 gtggtggacg agctcgtgaa agtgatgggc cggcacaagc ccgagaacat cgtgatcgaa
3961 atggccagag agaaccagac cacccagaag ggacagaaga acagccgcga gagaatgaag
4021 cggatcgaag agggcatcaa agagctgggc agccagatcc tgaaagaaca ccccgtggaa
4081 aacacccagc tgcagaacga gaagctgtac ctgtactacc tgcagaatgg gcgggatatg
4141 tacgtggacc aggaactgga catcaaccgg ctgtccgact acgatgtgga ccatatcgtg
4201 cctcagagct ttctgaagga cgactccatc gacaacaagg tgctgaccag aagcgacaag
4261 aaccggggca agagcgacaa cgtgccctcc gaagaggtcg tgaagaagat gaagaactac
4321 tggcggcagc tgctgaacgc caagctgatt acccagagaa agttcgacaa tctgaccaag
4381 gccgagagag gcggcctgag cgaactggat aaggccggct tcatcaagag acagctggtg
4441 gaaacccggc agatcacaaa gcacgtggca cagatcctgg actcccggat gaacactaag
4501 tacgacgaga atgacaagct gatccgggaa gtgaaagtga tcaccctgaa gtccaagctg
4561 gtgtccgatt tccggaagga tttccagttt tacaaagtgc gcgagatcaa caactaccac
4621 cacgcccacg acgcctacct gaacgccgtc gtgggaaccg ccctgatcaa aaagtaccct
4681 aagctggaaa gcgagttcgt gtacggcgac tacaaggtgt acgacgtgcg gaagatgatc
4741 gccaagagcg agcaggaaat cggcaaggct accgccaagt acttcttcta cagcaacatc
4801 atgaactttt tcaagaccga gattaccctg gccaacggcg agatccggaa gcggcctctg
4861 atcgagacaa acggcgaaac cggggagatc gtgtgggata agggccggga ttttgccacc
4921 gtgcggaaag tgctgagcat gccccaagtg aatatcgtga aaaagaccga ggtgcagaca

4981 ggcggcttca gcaaagagtc tatcctgccc aagaggaaca gcgataagct gatcgccaga
5041 aagaaggact gggaccctaa gaagtacggc ggcttcgaca gccccaccgt ggcctattct
5101 gtgctggtgg tggccaaagt ggaaaagggc aagtccaaga aactgaagag tgtgaaagag
5161 ctgctgggga tcaccatcat ggaaagaagc agcttcgaga agaatcccat cgactttctg
5221 gaagccaagg gctacaaaga agtgaaaaag gacctgatca tcaagctgcc taagtactcc
5281 ctgttcgagc tggaaaacgg ccggaagaga atgctggcct ctgccggcga actgcagaag
5341 ggaaacgaac tggccctgcc ctccaaatat gtgaacttcc tgtacctggc cagccactat
5401 gagaagctga agggctcccc cgaggataat gagcagaaac agctgtttgt ggaacagcac
5461 aagcactacc tggacgagat catcgagcag atcagcgagt tctccaagag agtgatcctg
5521 gccgacgcta atctggacaa agtgctgtcc gcctacaaca agcaccggga taagcccatc
5581 agagagcagg ccgagaatat catccacctg tttaccctga ccaatctggg agcccctgcc
5641 gccttcaagt actttgacac caccatcgac cggaagaggt acaccagcac caaagaggtg
5701 ctggacgcca ccctgatcca ccagagcatc accggcctgt acgagacacg gatcgacctg
5761 tctcagctgg gaggtgacag cggcgggagc ggcgggagcg gggggagcac taatctgagc
5821 gacatcattg agaaggagac tgggaaacag ctggtcattc aggagtccat cctgatgctg
5881 cctgaggagg tggaggaagt gatcggcaac aagccagagt ctgacatcct ggtgcacacc
5941 gcctacgacg agtccacaga tgagaatgtg atgctgctga cctctgacgc ccccgagtat
6001 aagccttggg ccctggtcat ccaggattct aacggcgaga ataagatcaa gatgctgagc
6061 ggaggatccg gaggatctgg aggcagcacc aacctgtctg acatcatcga gaaggagaca
6121 ggcaagcagc tggtcatcca ggagagcatc ctgatgctgc ccgaagaagt cgaagaagtg
6181 atcggaaaca agcctgagag cgatatcctg gtccataccg cctacgacga gagtaccgac
6241 gaaaatgtga tgctgctgac atccgacgcc ccagagtata agccctgggc tctggtcatc
6301 caggattcca acggagagaa caaaatcaaa atgctgtctg gcggctcaaa aagaaccgcc
6361 gacggcagcg aattcgagcc caagaagaag aggaaagtct aaccggtcat catcaccatc
6421 accattgagt ttaaacccgc tgatcagcct cgactgtgcc ttctagttgc cagccatctg
6481 ttgtttgccc ctccccccgtg ccttccttga ccctggaagg tgccactccc actgtccttt
6541 cctaataaaa tgaggaaatt gcatcgcatt gtctgagtag gtgtcattct attctggggg
6601 gtggggtggg gcaggacagc aagggggagg attgggaaga caatagcagg catgctgggg
6661 atgcggtggg ctctatggct tctgaggcgg aaagaaccag ctggggctcg ataccgtcga
6721 cctctagcta gagcttggcg taatcatggt catagctgtt tcctgtgtga aattgttatc
6781 cgctcacaat tccacacaac atacgagccg gaagcataaa gtgtaaagcc tagggtgcct
6841 aatgagtgag ctaactcaca ttaattgcgt tgcgctcact gcccgctttc cagtcgggaa
6901 acctgtcgtg ccagctgcat taatgaatcg gccaacgcgc ggggagaggc ggtttgcgta
6961 ttgggcgctc ttccgcttcc tcgctcactg actcgctgcg ctcggtcgtt cggctgcggc
7021 gagcggtatc agctcactca aaggcggtaa tacggttatc cacagaatca ggggataacg
7081 caggaaagaa catgtgagca aaaggccagc aaaaggccag gaaccgtaaa aaggccgcgt
7141 tgctggcgtt tttccatagg ctccgccccc ctgacgagca tcacaaaaat cgacgctcaa
7201 gtcagaggtg gcgaaacccg acaggactat aaagatacca ggcgtttccc cctggaagct
7261 ccctcgtgcg ctctcctgtt ccgaccctgc cgcttaccgg atacctgtcc gcctttctcc
7321 cttcgggaag cgtggcgctt tctcatagct cacgctgtag gtatctcagt tcggtgtagg
7381 tcgttcgctc caagctgggc tgtgtgcacg aaccccccgt tcagcccgac cgctgcgcct
7441 tatccggtaa ctatcgtctt gagtccaacc cggtaagaca cgacttatcg ccactggcag
7501 cagccactgg taacaggatt agcagagcga ggtatgtagg cggtgctaca gagttcttga
7561 agtggtggcc taactacggc tacactagaa gaacagtatt tggtatctgc gctctgctga
7621 agccagttac cttcggaaaa agagttggta gctcttgatc cggcaaacaa accaccgctg
7681 gtagcggtgg tttttttgtt tgcaagcagc agattacgcg cagaaaaaaa ggatctcaag
7741 aagatccttt gatcttttct acggggtctg acactcagtg gaacgaaaac tcacgttaag
7801 ggattttggt catgagatta tcaaaaagga tcttcaccta gatcctttta aattaaaaat
7861 gaagttttaa atcaatctaa agtatatatg agtaaacttg gtctgacagt taccaatgct
7921 taatcagtga ggcacctatc tcagcgatct gtctatttcg ttcatccata gttgcctgac
7981 tccccgtcgt gtagataact acgatacggg agggcttacc atctggcccc agtgctgcaa
8041 tgataccgcg agacccacgc tcaccggctc cagatttatc agcaataaac cagccagccg
8101 gaagggccga gcgcagaagt ggtcctgcaa ctttatccgc ctccatccag tctattaatt
8161 gttgccggga agctagagta agtagttcgc cagttaatag tttgcgcaac gttgttgcca
8221 ttgctacagg catcgtggtg tcacgctcgt cgtttggtat ggcttcattc agctccggtt
8281 cccaacgatc aaggcgagtt acatgatccc ccatgttgtg caaaaaagcg gttagctcct
8341 tcggtcctcc gatcgttgtc agaagtaagt tggccgcagt gttatcactc atggttatgg
8401 cagcactgca taattctctt actgtcatgc catccgtaag atgcttttct gtgactggtg
8461 agtactcaac caagtcattc tgagaatagt gtatgcggcg accgagttgc tcttgcccgg
8521 cgtcaatacg ggataatacc gcgccacata gcagaacttt aaaagtgctc atcattggaa
8581 aacgttcttc ggggcgaaaa ctctcaagga tcttaccgct gttgagatcc agttcgatgt
8641 aacccactcg tgcacccaac tgatcttcag catcttttac tttcaccagc gtttctgggt
8701 gagcaaaaac aggaaggcaa aatgccgcaa aaaagggaat aagggcgaca cggaaatgtt

```
     8761 gaatactcat actcttcctt tttcaatatt attgaagcat ttatcagggt tattgtctca
     8821 tgagcggata catatttgaa tgtatttaga aaaataaaca aataggggtt ccgcgcacat
     8881 ttccccgaaa agtgccacct gacgtcgacg gatcgggaga tcgatctccc gatcccctag
     8941 ggtcgactct cagtacaatc tgctctgatg ccgcatagtt aagccagtat ctgctccctg
     9001 cttgtgtgtt ggaggtcgct gagtagtgcg cgagcaaaat ttaagctaca acaaggcaag
     9061 gcttgaccga caattgcatg aagaatctgc ttagggttag gcgttttgcg ctgcttcgcg
     9121 atgtacgggc cagatatacg cgttgacatt gattattgac tagttattaa tagtaatcaa
     9181 ttacggggtc attagttcat agcccatata tggagttccg cgttacataa cttacggtaa
     9241 atggcccgcc tggctgaccg cccaacgacc cccgcccatt gacgtcaata atgacgtatg
     9301 ttcccatagt aacgccaata gggactttcc attgacgtca atgggtggag tatttacggt
     9361 aaactgccca cttggcagta catcaagtgt atc
//
```

11. phU6_gRNA_6xMS2

```
LOCUS       Exported                3949 bp ds-DNA     circular SYN 06-AUG-2018
DEFINITION  synthetic circular DNA
ACCESSION   .
VERSION     .
KEYWORDS    .
SOURCE      synthetic DNA construct
  ORGANISM  synthetic DNA construct
REFERENCE   1  (bases 1 to 3949)
  AUTHORS   Juan-Carlos Collantes
  TITLE     Direct Submission
  JOURNAL   Exported Aug 14, 2018 from SnapGene Viewer 4.2.3
            http://www.snapgene.com
FEATURES             Location/Qualifiers
     source          1..3949
                     /organism="synthetic DNA construct"
                     /mol_type="other DNA"
     rep_origin      complement(3..458)
                     /direction=LEFT
                     /label=f1 ori
                     /note="f1 bacteriophage origin of replication; arrow
                     indicates direction of (+) strand synthesis"
     primer_bind     600..616
                     /label=M13 fwd
                     /note="common sequencing primer, one of multiple similar
                     variants"
     promoter        626..644
                     /label=T7 promoter
                     /note="promoter for bacteriophage T7 RNA polymerase"
     promoter        696..936
                     /label=U6 promoter
                     /note="RNA polymerase III promoter for human U6 snRNA"
     misc_feature    866..1524
                     /label=Synthetic
     primer_bind     866..889
                     /label=CRC2_U6sg-F
     misc_RNA        1070..1088
                     /label=MS2 stem loop
                     /note="stem loop that binds the bacteriophage MS2 coat
                     protein"
     misc_RNA        1140..1158
                     /label=MS2 stem loop
                     /note="stem loop that binds the bacteriophage MS2 coat
                     protein"
     misc_RNA        1210..1228
                     /label=MS2 stem loop
```

```
                     /note="stem loop that binds the bacteriophage MS2 coat
                     protein"
     misc_RNA        1280..1298
                     /label=MS2 stem loop
                     /note="stem loop that binds the bacteriophage MS2 coat
                     protein"
     misc_RNA        1350..1368
                     /label=MS2 stem loop
                     /note="stem loop that binds the bacteriophage MS2 coat
                     protein"
     misc_RNA        1420..1438
                     /label=MS2 stem loop
                     /note="stem loop that binds the bacteriophage MS2 coat
                     protein"
     primer_bind     complement(1507..1524)
                     /label=CRC2_U6sg-R
     promoter        complement(1528..1546)
                     /label=T3 promoter
                     /note="promoter for bacteriophage T3 RNA polymerase"
     primer_bind     complement(1567..1583)
                     /label=M13 rev
                     /note="common sequencing primer, one of multiple similar
                     variants"
     protein_bind    1591..1607
                     /label=lac operator
                     /bound_moiety="lac repressor encoded by lacI"
                     /note="The lac repressor binds to the lac operator to
                     inhibit transcription in E. coli. This inhibition can be
                     relieved by adding lactose or
                     isopropyl-beta-D-thiogalactopyranoside (IPTG)."
     promoter        complement(1615..1645)
                     /label=lac promoter
                     /note="promoter for the E. coli lac operon"
     protein_bind    1660..1681
                     /label=CAP binding site
                     /bound_moiety="E. coli catabolite activator protein"
                     /note="CAP binding activates transcription in the presence
                     of cAMP."
     rep_origin      complement(1969..2557)
                     /direction=LEFT
                     /label=ori
                     /note="high-copy-number ColE1/pMB1/pBR322/pUC origin of
                     replication"
     CDS             complement(2634..3428)
                     /codon_start=1
                     /gene="aph(3')-II (or nptII)"
                     /product="aminoglycoside phosphotransferase from Tn5"
                     /label=NeoR/KanR
                     /note="confers resistance to neomycin, kanamycin, and G418
                     (Geneticin(R))"

/translation="MIEQDGLHAGSPAAWVERLFGYDWAQQTIGCSDAAVFRLSAQGRP

VLFVKTDLSGALNELQDEAARLSWLATTGVPCAAVLDVVTEAGRDWLLLGEVPGQDLLS

SHLAPAEKVSIMADAMRRLHTLDPATCPFDHQAKHRIERARTRMEAGLVDQDDLDEEHQ

GLAPAELFARLKASMPDGEDLVVTHGDACLPNIMVENGRFSGFIDCGRLGVADRYQDIA
                     LATRDIAEELGGEWADRFLVLYGIAAPDSQRIAFYRLLDEFF"
     promoter        complement(3463..3820)
                     /label=SV40 promoter
                     /note="SV40 enhancer and early promoter"
     rep_origin      3477..3612
```

```
                     /label=SV40 ori
                     /note="SV40 origin of replication"
     promoter        complement(3822..3926)
                     /gene="bla"
                     /label=AmpR promoter
ORIGIN
        1 ctaaattgta agcgttaata ttttgttaaa attcgcgtta aatttttgtt aaatcagctc
       61 atttttttaac caataggccg aaatcggcaa aatcccttat aaatcaaaag aatagaccga
      121 gatagggttg agtgttgttc cagtttggaa caagagtcca ctattaaaga acgtggactc
      181 caacgtcaaa gggcgaaaaa ccgtctatca gggcgatggc ccactacgtg aaccatcacc
      241 ctaatcaagt tttttggggt cgaggtgccg taaagcacta aatcggaacc ctaaagggag
      301 cccccgattt agagcttgac ggggaaagcc ggcgaacgtg gcgagaaagg aagggaagaa
      361 agcgaaagga gcgggcgcta gggcgctggc aagtgtagcg gtcacgctgc gcgtaaccac
      421 cacacccgcc gcgcttaatg cgccgctaca gggcgcgtcc cattcgccat tcaggctgcg
      481 caactgttgg gaagggcgat cggtgcgggc ctcttcgcta ttacgccagc tggcgaaagg
      541 gggatgtgct gcaaggcgat taagttgggt aacgccaggg ttttcccagt cacgacgttg
      601 taaaacgacg gccagtgagc gcgcgtaata cgactcacta tagggcgaat tgggtacccg
      661 tctcacaggc ggatcgatcc aaggtcgggc aggaagaggg cctatttccc atgattcctt
      721 catatttgca tatacgatac aaggctgtta gagagataat tggaattaat ttgactgtaa
      781 acacaaagat attagtacaa aatacgtgac gtagaaagta ataatttctt gggtagtttg
      841 cagttttaaa attatgtttt aaaatggact atcatatgct taccgtaact tgaaagtatt
      901 tcgatttctt ggctttatat atcttgtgga aaggacgaaa caccgggtct tcgagaagac
      961 ctgtttaaga gctatgctgg aaacagcata gcaagtttaa ataaggctag tccgttatca
     1021 acttgaaaaa gtggcaccga gtcggtgctc tgcaggtcga ctctagaaaa catgaggatc
     1081 acccatgtct gcagtattcc cgggttcatt agatcctaag gtacctaatt gcctagaaaa
     1141 catgaggatc acccatgtcc gacaccgaac gtgcgacaaa gctaagtgcg ccgtacggag
     1201 atccgataaa catgaggatc acccatgtcc ttgcgagatt acgcggcaca ttattggtcg
     1261 acgccgcgag ctatcgtaaa catgaggatc acccatgtct ctgtttgacg cgctggacgc
     1321 ccatgatcgt ccgatctggc ccgcgttaaa catgaggatc acccatgtcg ctccgcggct
     1381 tgaaaccata caccgaacga gacggtcaat cgcgcagaaa catgaggatc acccatgtcg
     1441 gaccgagggc tctagaaatt tttttgcggc cgcgactcta gatgctgagc gtactgagac
     1501 gccgcggtgg agctccagct tttgttccct ttagtgaggg ttaattgcgc gcttggcgta
     1561 atcatggtca tagctgtttc ctgtgtgaaa ttgttatccg ctcacaattc cacacaacat
     1621 acgagccgga agcataaagt gtaaagcctg gggtgcctaa tgagtgagct aactcacatt
     1681 aattgcgttg cgctcactgc ccgctttcca gtcgggaaac ctgtcgtgcc agctgcatta
     1741 atgaatcggc caacgcgcgg ggagaggcgg tttgcgtatt gggcgctctt ccgcttcctc
     1801 gctcactgac tcgctgcgct cggtcgttcg gctgcggcga gcggtatcag ctcactcaaa
     1861 ggcggtaata cggttatcca cagaatcagg ggataacgca ggaaagaaca tgtgagcaaa
     1921 aggccagcaa aaggccagga accgtaaaaa ggccgcgttg ctggcgtttt tccataggct
     1981 ccgcccccct gacgagcatc acaaaaatcg acgctcaagt cagaggtggc gaaacccgac
     2041 aggactataa agataccagg cgtttccccc tggaagctcc ctcgtgcgct ctcctgttcc
     2101 gaccctgccg cttaccggat acctgtccgc ctttctccct tcgggaagcg tggcgctttc
     2161 tcatagctca cgctgtaggt atctcagttc ggtgtaggtc gttcgctcca agctgggctg
     2221 tgtgcacgaa ccccccgttc agcccgaccg ctgcgcctta tccggtaact atcgtcttga
     2281 gtccaacccg gtaagacacg acttatcgcc actggcagca gccactggta acaggattag
     2341 cagagcgagg tatgtaggcg gtgctacaga gttcttgaag tggtggccta actacggcta
     2401 cactagaagg acagtatttg gtatctgcgc tctgctgaag ccagttacct tcggaaaaag
     2461 agttggtagc tcttgatccg gcaaacaaac caccgctggt agcggtggtt tttttgtttg
     2521 caagcagcag attacgcgca gaaaaaaagg atctcaagaa gatcctttga tcttttctac
     2581 ggggtctgac gctcagtgga acgaaaactc acgttaaggg attttggtca tgatcagaag
     2641 aactcgtcaa gaaggcgata gaaggcgatg cgctgcgaat cgggagcggc gataccgtaa
     2701 agcacgagga agcggtcagc ccattcgccg ccaagctctt cagcaatatc acgggtagcc
     2761 aacgctatgt cctgatagcg gtccgccaca cccagccggc cacagtcgat gaatccagaa
     2821 aagcggccat tttccaccat gatattcggc aagcaggcat cgccatgggt cacgacgaga
     2881 tcctcgccgt cgggcatgct cgccttgagc ctggcgaaca gttcggctgg cgcgagcccc
     2941 tgatgctctt cgtccagatc atcctgatcg acaagaccgg cttccatccg agtacgtgct
     3001 cgctcgatgc gatgtttcgc ttggtggtcg aatgggcagg tagccggatc aagcgtatgc
     3061 agccgccgca ttgcatcagc catgatggat actttctcgg caggagcaag gtgagatgac
     3121 aggagatcct gccccggcac ttcgcccaat agcagccagt cccttcccgc ttcagtgaca
     3181 acgtcgagca cagctgcgca aggaacgccc gtcgtggcca gccacgatag ccgcgctgcc
     3241 tcgtcttgca gttcattcag ggcaccggac aggtcggtct tgacaaaaag aaccgggcgc
     3301 ccctgcgctg acagccggaa cacggcggca tcagagcagc cgattgtctg ttgtgcccag
     3361 tcatagccga atagcctctc cacccaagcg gccggagaac ctgcgtgcaa tccatcttgt
```

```
     3421 tcaatcatgc gaaacgatcc tcatcctgtc tcttgatcga tctttgcaaa agcctaggcc
     3481 tccaaaaaag cctcctcact acttctggaa tagctcagag gccgaggcgg cctcggcctc
     3541 tgcataaata aaaaaaatta gtcagccatg gggcggagaa tgggcggaac tgggcggagt
     3601 taggggcggg atgggcggag ttaggggcgg gactatggtt gctgactaat tgagatgcat
     3661 gctttgcata cttctgcctg ctggggagcc tggggacttt ccacacctgg ttgctgacta
     3721 attgagatgc atgctttgca tacttctgcc tgctggggag cctggggact ttccacaccc
     3781 taactgacac acattccaca gctggttctt tccgcctcag gactcttcct ttttcaatat
     3841 tattgaagca tttatcaggg ttattgtctc atgagcggat acatatttga atgtatttag
     3901 aaaaataaac aaataggggt tccgcgcaca tttccccgaa aagtgccac
//
```

12. phU6_gRNA_2+4xMS2

```
LOCUS       Exported                3859 bp ds-DNA     circular SYN 06-AUG-2018
DEFINITION  synthetic circular DNA
ACCESSION   .
VERSION     .
KEYWORDS    phU6_sgRNA2.0_2xBbsI
SOURCE      synthetic DNA construct
  ORGANISM  synthetic DNA construct
REFERENCE   1  (bases 1 to 3859)
  AUTHORS   Juan-Carlos Collantes
  TITLE     Direct Submission
  JOURNAL   Exported Aug 14, 2018 from SnapGene Viewer 4.2.3
            http://www.snapgene.com
FEATURES             Location/Qualifiers
     source          1..3859
                     /organism="synthetic DNA construct"
                     /mol_type="other DNA"
     rep_origin      complement(3..458)
                     /direction=LEFT
                     /label=f1 ori
                     /note="f1 bacteriophage origin of replication; arrow
                     indicates direction of (+) strand synthesis"
     primer_bind     600..616
                     /label=M13 fwd
                     /note="common sequencing primer, one of multiple similar
                     variants"
     promoter        626..644
                     /label=T7 promoter
                     /note="promoter for bacteriophage T7 RNA polymerase"
     promoter        696..936
                     /label=U6 promoter
                     /note="RNA polymerase III promoter for human U6 snRNA"
     misc_feature    866..1434
                     /label=Synthetic
     primer_bind     866..889
                     /label=CRC2_U6sg-F
     misc_RNA        980..998
                     /label=MS2 stem loop
                     /note="stem loop that binds the bacteriophage MS2 coat
                     protein"
     misc_RNA        1050..1068
                     /label=MS2 stem loop
                     /note="stem loop that binds the bacteriophage MS2 coat
                     protein"
     misc_RNA        1120..1138
                     /label=MS2 stem loop
                     /note="stem loop that binds the bacteriophage MS2 coat
                     protein"
```

```
     misc_RNA        1190..1208
                     /label=MS2 stem loop
                     /note="stem loop that binds the bacteriophage MS2 coat
                     protein"
     misc_RNA        1260..1278
                     /label=MS2 stem loop
                     /note="stem loop that binds the bacteriophage MS2 coat
                     protein"
     misc_RNA        1330..1348
                     /label=MS2 stem loop
                     /note="stem loop that binds the bacteriophage MS2 coat
                     protein"
     primer_bind     complement(1417..1434)
                     /label=CRC2_U6sg-R
     promoter        complement(1438..1456)
                     /label=T3 promoter
                     /note="promoter for bacteriophage T3 RNA polymerase"
     primer_bind     complement(1477..1493)
                     /label=M13 rev
                     /note="common sequencing primer, one of multiple similar
                     variants"
     protein_bind    1501..1517
                     /label=lac operator
                     /bound_moiety="lac repressor encoded by lacI"
                     /note="The lac repressor binds to the lac operator to
                     inhibit transcription in E. coli. This inhibition can be
                     relieved by adding lactose or
                     isopropyl-beta-D-thiogalactopyranoside (IPTG)."
     promoter        complement(1525..1555)
                     /label=lac promoter
                     /note="promoter for the E. coli lac operon"
     protein_bind    1570..1591
                     /label=CAP binding site
                     /bound_moiety="E. coli catabolite activator protein"
                     /note="CAP binding activates transcription in the presence
                     of cAMP."
     rep_origin      complement(1879..2467)
                     /direction=LEFT
                     /label=ori
                     /note="high-copy-number ColE1/pMB1/pBR322/pUC origin of
                     replication"
     CDS             complement(2544..3338)
                     /codon_start=1
                     /gene="aph(3')-II (or nptII)"
                     /product="aminoglycoside phosphotransferase from Tn5"
                     /label=NeoR/KanR
                     /note="confers resistance to neomycin, kanamycin, and G418
                     (Geneticin(R))"

/translation="MIEQDGLHAGSPAAWVERLFGYDWAQQTIGCSDAAVFRLSAQGRP

VLFVKTDLSGALNELQDEAARLSWLATTGVPCAAVLDVVTEAGRDWLLLGEVPGQDLLS

SHLAPAEKVSIMADAMRRLHTLDPATCPFDHQAKHRIERARTRMEAGLVDQDDLDEEHQ

GLAPAELFARLKASMPDGEDLVVTHGDACLPNIMVENGRFSGFIDCGRLGVADRYQDIA
                     LATRDIAEELGGEWADRFLVLYGIAAPDSQRIAFYRLLDEFF"
     promoter        complement(3373..3730)
                     /label=SV40 promoter
                     /note="SV40 enhancer and early promoter"
     rep_origin      3387..3522
                     /label=SV40 ori
                     /note="SV40 origin of replication"
```

```
     promoter        complement(3732..3836)
                     /gene="bla"
                     /label=AmpR promoter
ORIGIN
        1 ctaaattgta agcgttaata ttttgttaaa attcgcgtta aatttttgtt aaatcagctc
       61 atttttttaac caataggccg aaatcggcaa aatcccttat aaatcaaaag aatagaccga
      121 gatagggttg agtgttgttc cagtttggaa caagagtcca ctattaaaga acgtggactc
      181 caacgtcaaa gggcgaaaaa ccgtctatca gggcgatggc ccactacgtg aaccatcacc
      241 ctaatcaagt tttttggggt cgaggtgccg taaagcacta aatcggaacc ctaaagggag
      301 cccccgattt agagcttgac ggggaaagcc ggcgaacgtg gcgagaaagg aagggaagaa
      361 agcgaaagga gcgggcgcta gggcgctggc aagtgtagcg gtcacgctgc gcgtaaccac
      421 cacacccgcc gcgcttaatg cgccgctaca gggcgcgtcc cattcgccat tcaggctgcg
      481 caactgttgg gaagggcgat cggtgcgggc ctcttcgcta ttacgccagc tggcgaaagg
      541 gggatgtgct gcaaggcgat taagttgggt aacgccaggg ttttcccagt cacgacgttg
      601 taaaacgacg gccagtgagc gcgcgtaata cgactcacta tagggcgaat tgggtacccg
      661 tctcacaggc ggatcgatcc aaggtcgggc aggaagaggg cctatttccc atgattcctt
      721 catatttgca tatacgatac aaggctgtta gagagataat tggaattaat ttgactgtaa
      781 acacaaagat attagtacaa aatacgtgac gtagaaagta ataatttctt gggtagtttg
      841 cagttttaaa attatgtttt aaaatggact atcatatgct taccgtaact tgaaagtatt
      901 tcgatttctt ggctttatat atcttgtgga aaggacgaaa caccgggtct tcgagaagac
      961 ctgtttaaga gctaggccaa catgaggatc acccatgtct gcagggccta gcaagtttaa
     1021 ataaggctag tccgttatca acttggccaa catgaggatc acccatgtct gcagggccaa
     1081 gtggcaccga gtcggtgctc tgcaggtcga ctctagaaaa catgaggatc acccatgtct
     1141 gcagtattcc cgggttcatt agatcctaag gtacctaatt gcctagaaaa catgaggatc
     1201 acccatgtcc gacaccgaac gtgcgacaaa gctaagtgcg ccgtacggag atccgataaa
     1261 catgaggatc acccatgtcc ttgcgagatt acgcggcaca ttattggtcg acgccgcgag
     1321 ctatcgtaaa catgaggatc acccatgtcg gaccgagggc tctagaaatt tttttgcggc
     1381 cgcgactcta gatgctgagc gtactgagac gccgcggtgg agctccagct tttgttccct
     1441 ttagtgaggg ttaattgcgc gcttggcgta atcatggtca tagctgtttc ctgtgtgaaa
     1501 ttgttatccg ctcacaattc cacacaacat acgagccgga agcataaagt gtaaagcctg
     1561 gggtgcctaa tgagtgagct aactcacatt aattgcgttg cgctcactgc ccgctttcca
     1621 gtcgggaaac ctgtcgtgcc agctgcatta atgaatcggc caacgcgcgg ggagaggcgg
     1681 tttgcgtatt gggcgctctt ccgcttcctc gctcactgac tcgctgcgct cggtcgttcg
     1741 gctgcggcga gcggtatcag ctcactcaaa ggcggtaata cggttatcca cagaatcagg
     1801 ggataacgca ggaaagaaca tgtgagcaaa aggccagcaa aaggccagga accgtaaaaa
     1861 ggccgcgttg ctggcgtttt tccataggct ccgcccccct gacgagcatc acaaaaatcg
     1921 acgctcaagt cagaggtggc gaaacccgac aggactataa agataccagg cgtttccccc
     1981 tggaagctcc ctcgtgcgct ctcctgttcc gaccctgccg cttaccggat acctgtccgc
     2041 ctttctccct tcgggaagcg tggcgctttc tcatagctca cgctgtaggt atctcagttc
     2101 ggtgtaggtc gttcgctcca agctgggctg tgtgcacgaa ccccccgttc agcccgaccg
     2161 ctgcgcctta tccggtaact atcgtcttga gtccaacccg gtaagacacg acttatcgcc
     2221 actggcagca gccactggta acaggattag cagagcgagg tatgtaggcg gtgctacaga
     2281 gttcttgaag tggtggccta actacggcta cactagaagg acagtatttg gtatctgcgc
     2341 tctgctgaag ccagttacct tcggaaaaag agttggtagc tcttgatccg gcaaacaaac
     2401 caccgctggt agcggtggtt tttttgtttg caagcagcag attacgcgca gaaaaaaagg
     2461 atctcaagaa gatcctttga tcttttctac ggggtctgac gctcagtgga acgaaaactc
     2521 acgttaaggg attttggtca tgatcagaag aactcgtcaa gaaggcgata gaaggcgatg
     2581 cgctgcgaat cgggagcggc gataccgtaa agcacgagga agcggtcagc ccattcgccg
     2641 ccaagctctt cagcaatatc acgggtagcc aacgctatgt cctgatagcg gtccgccaca
     2701 cccagccggc cacagtcgat gaatccagaa aagcggccat tttccaccat gatattcggc
     2761 aagcaggcat cgccatgggt cacgacgaga tcctcgccgt cgggcatgct cgccttgagc
     2821 ctggcgaaca gttcggctgg cgcgagcccc tgatgctctt cgtccagatc atcctgatcg
     2881 acaagaccgg cttccatccg agtacgtgct cgctcgatgc gatgtttcgc ttggtggtcg
     2941 aatgggcagg tagccggatc aagcgtatgc agccgccgca ttgcatcagc catgatggat
     3001 actttctcgg caggagcaag gtgagatgac aggagatcct gccccggcac ttcgcccaat
     3061 agcagccagt cccttcccgc ttcagtgaca acgtcgagca cagctgcgca aggaacgccc
     3121 gtcgtggcca gccacgatag ccgcgctgcc tcgtcttgca gttcattcag ggcaccggac
     3181 aggtcggtct tgacaaaaag aaccgggcgc ccctgcgctg acagccggaa cacggcggca
     3241 tcagagcagc cgattgtctg ttgtgcccag tcatagccga atagcctctc cacccaagcg
     3301 gccggagaac ctgcgtgcaa tccatcttgt tcaatcatgc gaaacgatcc tcatcctgtc
     3361 tcttgatcga tctttgcaaa agcctaggcc tccaaaaaag cctcctcact acttctggaa
     3421 tagctcagag gccgaggcgg cctcggcctc tgcataaata aaaaaaatta gtcagccatg
     3481 gggcggagaa tgggcggaac tgggcggagt taggggcggg atgggcggag ttaggggcgg
```

```
     3541 gactatggtt gctgactaat tgagatgcat gctttgcata cttctgcctg ctggggagcc
     3601 tggggacttt ccacacctgg ttgctgacta attgagatgc atgctttgca tacttctgcc
     3661 tgctggggag cctggggact ttccacaccc taactgacac acattccaca gctggttctt
     3721 tccgcctcag gactcttcct ttttcaatat tattgaagca tttatcaggg ttattgtctc
     3781 atgagcggat acatatttga atgtatttag aaaaataaac aaataggggt tccgcgcaca
     3841 tttccccgaa aagtgccac
//
```