MOLECULAR APPROACHES FOR CONTROLLING RNA STABILITY

Erin Katherine Borchardt

A dissertation submitted to the faculty at the University of North Carolina at Chapel Hill in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the Curriculum in Genetics and Molecular Biology in the School of Medicine.

Chapel Hill
2016

Approved by:

Aravind Asokan

William F. Marzluff

Alain Laederach

Brian Kuhlman

Dale A. Ramsden

ABSTRACT

Erin Katherine Borchardt: Molecular Approaches for Controlling RNA Stability
(Under the direction of Aravind Asokan)

Nature utilizes a number of methods for regulating gene expression via modulation of RNA stability. Factors involved in these processes include microRNAs, nucleotide modifications, structural elements, and ribonucleases, among others. In this dissertation, we aim to develop novel approaches for controlling gene expression through the manipulation of RNA stability. Prokaryotic CRISPR systems have provided a wealth of new tools for a range of molecular biology applications. A large proportion of these CRISPR-based technologies rely on CRISPR-associated nucleases to cleave either target DNA or RNA. The CRISPR endoribonuclease, Csy4 (Cas6f), is one such CRISPR protein, which specifically interacts with, and cleaves a target RNA hairpin. Here, we explore the use of Csy4 for destabilizing and stabilizing RNA transcripts in mammalian cells. We find that Csy4 is capable of knocking down gene expression when targeting the 5'UTR or coding sequence of a reporter mRNA. Additionally, Csy4 can rescue destabilized transcripts when targeted to the 3'UTR, and can promote their translation. We utilize the ability of Csy4 to stabilize target hairpin-containing cleavage products to induce RNA circularization of an engineered RNA splicing reporter. Further, we demonstrate the use of recombinant Adeno-Associated Virus (rAAV)-delivered circular RNA (circRNA) expression cassettes as platforms for gene expression *in vivo.* We find that these circRNAs can be translated to produce protein products in mice, and

we observe expression differences between heart and highly proliferative liver tissue. The new approaches presented here provide starting points for the development of a number of new technologies including genetic ‘safety switches’ and genetic cassettes capable of therapeutic gene expression at lower rAAV vector dosages.

ACKNOWLEDGEMENTS

I am profoundly grateful to have been able to work in the supportive environment of the Asokan laboratory for my graduate studies. As my advisor, Dr. Aravind Asokan allowed me the intellectual freedom to explore new research areas for the lab while guiding my development as a scientist. Aravind has been consistently approachable, encouraging, enthusiastic, and honest in his mentoring. I have learned so much from him, within science and beyond, and I feel lucky to be able to count him among my friends.

My labmates were my surrogate family here in North Carolina. I feel immensely grateful to have developed not only meaningful professional relationships with them, but also deep friendships. Garrett Berry and Giridhar Murlidharan joined the Asokan lab at the same time as I did, and we endured the struggles and joys of graduate school and science together. I will always be thankful for having the opportunity to make this journey in their company. In addition, Blake Albright and Victoria Madigan brought optimism into the laboratory while Eric Horowitz, Nagesh Pulicherla, Ruth Castellanos-Rivera, Sven Moller-Tank, and Victor Tse provided post-doctoral wisdom and set admirable examples of productive researchers. I am also grateful for the hard work and dedication of the rotation students, technicians, undergraduate students, and summer students that contributed to the work presented here. Specifically, I would like to thank

Lavanya Rao, Leonidas Vandoros, Michael Huang, Ryan Fogg, Joey Heider, Rebecca Reardon, and Rita Meganck. I owe all members of the Asokan laboratory thanks not only for their scientific support but also for making the lab an open and friendly environment. I will always treasure our times, inside the lab and out, teeming with laughter and full conversation.

I am also fortunate to have had the support and guidance of a wonderful thesis committee in Drs. William F. Marzluff, Alain Laederach, Dale Ramsden, and Brian Kuhlman. Dr. Marzluff and Dr. Laederach introduced me to the incredibly talented and diverse RNA community present in the Triangle area. This RNA community has been welcoming and I am privileged to have had the opportunity to learn from them.

I became inspired to pursue a career in molecular biology through working in the laboratory of Dr. Lisa Stubbs and through the iGEM competition during my undergraduate years at the University of Illinois at Urbana-Champaign. It was in these environments that I first learned basic laboratory techniques and scientific approaches. For this, I owe many thanks to Dr. Lisa Stubbs, Dr. Christopher Rao, Dr. Yong-Su Jin, Younguk (Calvin) Sun, Dina Leiding, and Andrea Skinner.

Lastly, I am grateful for the support of my family and friends outside of North Carolina. This includes my parents, Kathy and Mike Borchardt, my brothers, Jason, Brett, and Elliott Borchardt, my grandparents Lois Borchardt and Ted Rogachuk, and my closest, lifelong friend, Selynn Hinkle.

Aravind has always emphasized the importance of the journey over the destination during my graduate training. Thank you all for making this journey so wonderful.

TABLE OF CONTENTS

LIST OF FIGURES ix

LIST OF ABBREVIATIONS xi

CHAPTER 1: INTRODUCTION 1

1.1 CRISPR Systems 1

1.2 Pre-crRNA Processing by Csy4 5

1.3 CRISPR Components as Tools 7

1.4 Sequence Elements Involved in RNA Stability 10

1.5 Circular RNAs 11

1.6 Adeno-Associated Virus-Mediated Gene Delivery and RNA Therapeutics 15

CHAPTER 2: CONTROLLING MRNA STABILITY AND TRANSLATION WITH THE CRISPR ENDORIBONUCLEASE CYS4 20

2.1 Overview 20

2.2 Introduction 21

2.3 Materials and Methods 24

2.4 Results and Discussion 29

CHAPTER 3: INDUCING CIRCULAR RNA FORMATION USING THE CRISPR ENDORIBONUCLEASE CSY4 47

3.1 Overview 47

3.2 Introduction 48

3.3 Materials and Methods 50

3.4 Results and Discussion 54

CHAPTER 4: *IN VIVO* DELIVERY OF TRANSLATABLE CIRCULAR RNA CASSETTES USING RECOMBINANT AAV VECTORS 66

4.1 Overview 66

4.2 Introduction 67

4.3 Methods 70

4.4 Results and Discussion 74

CHAPTER 5: SYNOPSIS AND FUTURE DIRECTIONS 85

5.1 Summary 85

5.2 Further Evaluation and Regulation of Csy4 in Mammalian Cells 86

5.3 Future Applications and Engineering of Csy4 88

5.4 Improving Current CircRNA Expression Cassettes 90

5.5 CircRNAs Containing an EMCV IRES are Efficiently Translated in Mice 94

5.6 Future Evaluation of rAAV CircRNA Expression Cassettes 95

5.7 Final Remarks 95

REFERENCES 99

LIST OF FIGURES

Figure 1: General schematic of CRISPR mediated immunity 17

Figure 2: Characteristics the Csy4-hairpin interaction. 18

Figure 3: Genomic map of AAV 19

Figure 4: Csy4-mediated knockdown of 5' UTR-hairpin (HP) and ATG-HP constructs 39

Figure 5: Effect of placing the HP in the 3'UTR on Csy4-mediated knockdown 41

Figure 6: Cys4 interacts with the substrate hairpin in human cells. 43

Figure 7: Csy4 binding and cleavage are necessary for regulation of transgene expression. 44

Figure 8: Selective processing by Csy4 is essential for rescue of 3'UTR-HP-Δp(A) constructs. 45

Figure 9: Potential events outlining Csy4-mediated processing of 3'UTR-HP-Δp(A) mRNA. 46

Figure 10: Schematic of circGFP-CD splicing patterns and reporter outputs. 60

Figure 11: Csy4 induces circGFP-CD reporter expression to switch from dsRed to GFP. 61

Figure 12: Detection of circGFP-CD expression products. 63

Figure 13: Accumulation of circRNA and linear splicing products over time. 65

Figure 14: Schematic of circGFP expression. 81

Figure 15: rAAV mediated delivery of circGFP expression cassette in mice. 82

Figure 16: Molecular characterization of circGFP expression products in mice. 83

Figure 17: *In vitro* expression of circRNA reporters from viral vectors. 97

Figure 18: *in vivo* expression of circRNA reporters from viral vectors. 98

LIST OF ABBREVIATIONS

AAT	Alpha-1-antitrypsin
AAV	Adeno-associated virus
Cas	CRISPR associated
CBA	Chicken beta actin
cDNA	complementary DNA
cGFP	Coral green fluorescent protein
CIP	Calf intestinal phosphatase
circGFP-CD	circGFP-Csy4 dependent
circRNA	Circular RNA
CRISPR	Clustered regularly interspaced short palindromic repeats
crRNA	CRISPR RNA
DAB	3,3’-Diaminobenzidine
DNA	Deoxyribonucleic acid
EDTA	Ethylenediaminetretraacetic acid
EIciRNA	Exon-intron circular RNA
EMCV	encephalomyocarditis virus
ENE	Expression and nuclear retention element
ESF	Engineered splicing factor
FGF-2	Fibroblast Growth Factor-2
GFP	Green fluorescent protein
GLuc	Gaussia luciferase
gRNA	Guide RNA

HEK293	Human embryonic kidney 293
HIPK3	Homeodomain interacting protein kinase 3
HIV	Human immunodeficiency virus
HSV-TK	Herpes simplex virus thymidine kinase
IGF2BP1	Insulin-like growth factor 2 mRNA binding protein 1
IRES	Internal ribosome entry site
ITR	Inverted terminal repeat
K_d	Binding affinity
KSHV	Kaposi's sarcoma associated herpesvirus
LTR	Long terminal repeat
MALAT1	Metastasis associated lung adenocarcinoma transcript 1
MascRNA	MALAT1-associated small cytoplasmic RNA
MBL	Muscleblind
MENβ	Multiple endocrine neoplasia β
miRNA	MicroRNA
MOI	Multiplicity of infection
mRNA	Messenger RNA
ORF	Open reading frame
PABP	Poly-A binding protein
PAM	Protospacer adjacent motif
PAN RNA	Polyadenylated nuclear RNA
PBS	Phosphate buffered saline
PCR	Polymerase chain reaction

PEI	Polyethylenimine
Pre-crRNA	Precursor CRISPR RNA
PUF	Pumilio/fem-2 binding factor
qPCR	quantitative polymerase chain reaction
qRT-PCR	quantitative reverse transcription polymerase chain reaction
rAAV	Recombinant Adeno-Associated Virus
RBS	Ribosome binding site
RNA	Ribonucleic acid
RT PCR	Reverse transcription polymerase chain reaction
shRNA	Short hairpin RNA
TBST	Tris buffered saline with Tween 20
tracrRNA	trans-activating CRISPR RNA
tricRNA	tRNA intronic circular RNA
tRNA	Transfer RNA
UTR	Untranslated region
VPR	Viral protein R
ZKSCAN1	Zinc finger with KRAB and SCAN domains 1

CHAPTER 1: INTRODUCTION

1.1 CRISPR Systems

Prokaryotic organisms regularly encounter foreign genetic elements which pose a threat to their survival. As such, ~45% of bacteria and ~83% of archaea utilize the Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR) system as a mechanism of defense against these invading nucleic acids (Grissa et al., 2007).

CRISPR systems neutralize foreign nucleic acid using a combination of proteins and RNA elements encoded within CRISPR loci. These loci are typically characterized by the presence of a repeat-spacer-array, a leader sequence, and a series of CRISPR-associated (Cas) protein encoding genes. Within the repeat-spacer array, short repeating motifs flank unique sequence elements (spacers) which share sequence homology with previously encountered foreign nucleic acid (Barrangou et al., 2007a; Bolotin et al., 2005; Mojica et al., 2005; Pourcel et al., 2005). The leader sequence, which exhibits promoter activity, is involved in expressing the repeat-spacer array and in acquisition of new spacer sequences (Pul et al., 2010; Wei et al., 2015). The protein products of Cas genes are responsible for carrying out essential steps in the CRISPR pathway, ultimately leading to target neutralization. CRISPR-based immunity is carried out in three general stages: acquisition (adaptation), expression and maturation, and interference (Figure 1). However, CRISPR systems vary in their exact mechanisms to carry out each of these stages, owing to a diverse range of Cas proteins. As such,

CRISPR systems have been divided into classes, types, and subtypes, discussed below.

CRISPR Classification:

CRISPR systems are functionally categorized into two classes based on their complement of Cas proteins (Makarova et al., 2015). CRISPR classes are divided into five types denoted by roman numerals, which are further separated into 16 subtypes denoted by alphabetical letters (Makarova et al., 2015). All functional CRISPR systems require the presence of both Cas1 and Cas2, which are essential for the adaptation stage of CRISPR immunity. Accessory Cas proteins provide the basis for classification and specify the mechanistic differences of immunity between each CRISPR subtype. Class 1 CRISPR systems are characterized by a multi-subunit protein complex of Cas proteins which is utilized to carry out target neutralization. Conversely, Class 2 CRISPR systems utilize a single Cas protein to mediate CRISPR immunity.

CRISPR Adaptation:

Initial studies exploring CRISPR immunity recognized that spacer sequences were identical to sequences present in phage and plasmids (Bolotin et al., 2005; Mojica et al., 2005; Pourcel et al., 2005). In addition, studies demonstrated that phage-resistant *Streptococcus thermophilus* could be generated by challenging the bacteria with bacteriophage of interest and isolating survivors. Evaluation of the CRISPR locus in these strains revealed the addition of one or more spacer sequences corresponding to the specific bacteriophage the repeat-spacer array (Barrangou et al., 2007b). Further, if

these spacer sequences were subsequently deleted from the bacteria, their acquired resistance was lost (Barrangou et al., 2007b). Together, these data determined a relationship between the spacer sequence and CRISPR-mediated immunity. In particular, this study established the fundamental adaptability of the CRISPR system, demonstrating that new, immunizing spacers can be acquired upon exposure to a pathogen.

The acquisition (or adaptation) phase of CRISPR immunity requires Cas1 and Cas2 to incorporate a small segment of foreign nucleic acid sequence as a new spacer in the repeat-spacer array (Nunez et al., 2015; Yosef et al., 2012) (Figure 1). Integration of a new spacer into the repeat-spacer array always occurs at the leader proximal end of the array (Barrangou et al., 2007b). The original sequence of the spacer, found in the foreign nucleic acid, is known as the protospacer. A protospacer adjacent motif (PAM) immediately flanking the protospacer is utilized by the adaptation machinery for recognizing new potential spacer sequences (Heler et al., 2015).

Interestingly, CRISPR systems also appear to be primed for acquisition of new spacers against previously encountered targets. Such primed acquisition occurs when a particular target shares partial homology with a spacer that is already present in the repeat-spacer array (Datsenko et al., 2012; Swarts et al., 2012). This may represent a mechanism for combatting bacteriophage that escape from CRISPR targeting by mutating their protospacers (Datsenko et al., 2012).

CRISPR Expression and Maturation:

In the expression and maturation stage of CRISPR immunity, the repeat-spacer

array is transcribed into a single, long precursor transcript (pre-crRNA) (Figure 1). Transcription of the pre-crRNA is driven by a promoter in the leader sequence (Pul et al., 2010). The pre-crRNA is subsequently processed by Cas proteins into small CRISPR RNAs (crRNAs) such that each crRNA contains a single spacer and a partial repeat handle. The pre-crRNA processing step of the CRISPR pathway varies between CRISPR types and subtypes. For example, in the Class 1 Type-IF CRISPR system of *Pseudomonas aeruginosa*, the task of crRNA biogenesis falls to Csy4 (Cas6f) (Cady and O'Toole, 2011; Haurwitz et al., 2010). However, in Class 2 type II CRISPR systems, the combined efforts of Cas9, cellular Ribonuclease III, and an additional small RNA (trans-activating CRISPR RNA, or tracrRNA) are responsible for pre-crRNA processing (Karvelis et al., 2013).

CRISPR Interference:

During the interference phase, Cas nucleases are guided to their targets by crRNAs in a sequence specific manner (Figure 1). An important checkpoint in distinguishing target from self occurs in most CRISPR systems during the interference stage. In this step, the interference complex interrogates the sequence adjacent to the protospacer for a PAM. If the PAM is present, the Cas nuclease is licensed to cleave the target, rendering the invading nucleic acid inactive (Sternberg et al., 2014; Szczelkun et al., 2014). Verification of the PAM sequence prevents targeting of the genomic CRISPR locus, which contains crRNA binding sites but lacks a PAM.

As in the expression and maturation phase, the mechanistic details of CRISPR interference vary by CRISPR class and type. For example, PAM sequences vary

between CRISPR subtypes, both in sequence and location relative to the protospacer (5' or 3') (Shah et al., 2013). Further, in Class 1 systems, the interference complex is formed by crRNA interaction with a multi-subunit Cas protein complex. Target cleavage is carried out by Cas3 in Class 1 Type I systems while Cas10 carries out this task in Class 1 Type III systems (Brouns et al., 2008; Hatoum-Aslan et al., 2014; Samai et al., 2015; Sinkunas et al., 2011; Westra et al., 2012). In comparison, Class 2 systems utilize a single Cas protein bound to crRNA to form the effector complex (Makarova et al., 2015). In Class 2 Type II systems, this task falls to Cas9. (Barrangou et al., 2007b; Garneau et al., 2010b; Jinek et al., 2012; Sapranauskas et al., 2011).

1.2 Pre-crRNA Processing by Csy4

Csy4 (Cas6f) is a CRISPR endoribonuclease member of Class 1 Type IF CRISPR systems. Functioning in the expression stage of CRISPR immunity, Csy4 cleaves the pre-crRNA at regular intervals to generate small crRNAs (Cady and O'Toole, 2011; Haurwitz et al., 2010). In the *P. aeruginosa* Type I-F CRISPR system, each repeat in the repeat-spacer array consists of 28 nucleotides. Sixteen of these nucleotides form a short hairpin which serves as the substrate for Csy4 binding and cleavage. Csy4-mediated cleavage occurs at the 3' base of each hairpin stem, generating a terminal 3'-phosphate (Wiedenheft et al., 2011) (Figure 2C). Each resulting crRNA contains a unique 32-nucleotide spacer flanked by portions of repeat-derived sequences on either side (Wiedenheft et al., 2011). Csy4 remains tightly associated with the processed crRNA product and complexes with additional Cas proteins for the

interference stage of the CRISPR pathway (Sternberg et al., 2012; Wiedenheft et al., 2011).

Csy4 and the target hairpin form a high affinity interaction ($K_d \approx 50$ pM) which is dependent on both sequence- and structure-specific contacts (Haurwitz et al., 2010; Sternberg et al., 2012). A positively charged, arginine-rich helix docks into the major groove of the hairpin stem (Haurwitz et al., 2010) (Figure 2A, B). Within the arginine-rich helix, Arg114, Arg115, Arg118, Arg119 and His120 make contact with phosphate groups in the RNA hairpin (Haurwitz et al., 2010) (Figure 2A). When a subset of these arginine residues are mutated, the capacity of Csy4 to bind the substrate hairpin is greatly reduced (Sternberg et al., 2012). Nucleotide-specific contacts are specified by hydrogen bonding between Arg102 and G20, Gln104 and A19, and Arg115 and G11 (Haurwitz et al., 2010) (Figure 2A). Lastly, Phe155 interacts with, and positions G20 in the enzyme active site via base-stacking interactions (Haurwitz et al., 2010; Sternberg et al., 2012) (Figure 2A).

The Csy4 active site relies on two critical amino acids for catalysis, Ser148 and His29 (Figure 2A). Mutation of either of these residues to cysteine and alanine respectively renders Csy4 inactive (Haurwitz et al., 2010). Subsequent studies determined that Ser148 is required for positioning a 2' hydroxyl group within the active site while His29 acts as a general base for the cleavage reaction (Sternberg et al., 2012).

1.3 CRISPR Components as Tools

CRISPR systems have provided the basis for a breadth of new technologies ranging from bacterial genotyping to genome engineering (Selle and Barrangou, 2015; Wright et al., 2016). Much of the original focus on CRISPR was aimed at developing tools for application in the dairy industry, which relies on valuable bacterial starter cultures. Initial studies sought new techniques for vaccination of bacteria against bacteriophages and drove many seminal CRISPR findings (Barrangou et al., 2007b; Deveau et al., 2008; Horvath et al., 2008). As such, a number of CRISPR tools support the study and maintenance of bacteria commonly used in the food industry. For example, the consistent integration of new spacers at the leader proximal side of the repeat-spacer array records a sequential history of exposure to foreign nucleic acid (Barrangou et al., 2007b). This trait has been exploited for the genotyping of various bacteria (Grissa et al., 2009; Vergnaud et al., 2007).

Perhaps the most widely appreciated CRISPR-based tool is Cas9, an RNA guided nuclease present in all Class 2 type-II CRISPR systems (Makarova et al., 2015). During the interference phase, Cas9 is guided to foreign DNA targets through the specificity of a crRNA. Cas9's interaction with the crRNA is bridged by an additional small RNA called the trans-activating crRNA (tracrRNA) (Deltcheva et al., 2011). The Cas9-crRNA-tracrRNA complex is guided to foreign DNA through the sequence specified by the crRNA, where Cas9 mediates a double strand break. It was discovered that the tracrRNA and crRNA could be linked together to form a single RNA termed a guide-RNA (gRNA) (Jinek et al., 2012). This specificity of the gRNA can be easily programmed by replacing the crRNA region with 20 nucleotides targeting nearly any

region of interest. Consequently, Cas9 can be guided to cleave any sequence, provided the appropriate PAM is present.

Cas9 is an attractive tool for genome engineering, and a myriad of other applications, due to the ease of which it can be programmed. Initial applications of Cas9 focused on genome editing but have since expanded to include functions such as RNA-guided RNA imaging and transcriptional control (Gilbert et al., 2013; Mali et al., 2013a; Nelles et al., 2016). Furthermore, Cas9 seems to function in nearly every organism tested, ranging from bacteria to human cell lines (Cong et al., 2013; DiCarlo et al., 2013; Friedland et al., 2013; Gratz et al., 2013; Hwang et al., 2013; Jiang et al., 2013; Jinek et al., 2013; Li et al., 2013; Mali et al., 2013b; Nekrasov et al., 2013; Wang et al., 2013).

However, the use of Cas9 is limited in a few ways. First, target selection is restricted to sequences containing the PAM sequence of the particular Cas9 ortholog being used for cleavage. For *Streptococcus pyogenes* Cas9, the Cas9 ortholog utilized in initial Cas9 genome editing studies, the PAM requirement restricts gRNA design to sequences with a 3' NGG motif (Mojica et al., 2009). In addition, gRNA expression is restricted to RNA polymerase III promoters *in vivo*. The U6 promoter, which requires guanine as the first nucleotide in the gRNA tends to be preferred, adding an additional sequence constraint to gRNA selection. Limiting expression at the promoter level also hinders the use of tissue specific and conditional expression regulation. However, a solution to these promoter restrictions was found in another Cas protein, Csy4. Csy4 hairpins were incorporated flanking a gRNA such that co-expression with Csy4 will result in cleavage and separation from other RNA elements in a transcript (Nissim et al., 2014; Tsai et al., 2014). This allows expression from polymerase II promoters without

interference from sequences that may affect gRNA, function, stability, or localization. Further, multiple gRNA can be expressed from a single promoter if Csy4 hairpins are placed between each gRNA, allowing multiplexed Cas9 applications (Nissim et al., 2014; Tsai et al., 2014).

Csy4 has also formed the basis of a number of other tools and techniques. Inactivation of Csy4 catalytic activity through a H29A mutation has allowed for Csy4-mediated purification of hairpin-tagged RNA transcripts (Lee et al., 2013; Salvail-Lacoste et al., 2013). The purified transcripts can subsequently be analyzed for identification of proteins associated with a transcript of interest (Lee et al., 2013). Additionally, Csy4 has been adapted for gene regulation and processing RNA transcripts encoding multiple genes in *Escherichia coli, Bacillus subtilis, and Saccharomyces cerevisiae* (Qi et al., 2012). Csy4 was also utilized in regulating translation of prokaryotic gene expression (Du et al., 2016a). In this context, Csy4 hairpins separated a ribosome binding site (RBS) and open reading frame (ORF) from an upstream cis-repressive RNA. The cis-repressive RNA functions by hybridizing with, and masking the RBS, preventing ribosomal loading and translation. Upon co-expression with Csy4, the cis-repressive RNA is cleaved away, exposing the RBS, and allowing translation (Du et al., 2016a). Csy4 has also been used to regulate viral infection (Guo et al., 2015). For this application, Csy4 was fused to HIV Viral Protein R (VPR) in an effort to guide Csy4 to the HIV particle. The Csy4 hairpin was incorporated into the 3’ end of the HIV genome before the LTR, as a target for Csy4 cleavage. Expression of the Csy4-VPR fusion in HIV infected cells resulted in undetectable HIV levels, demonstrating Csy4-mediated viral inhibition (Guo et al., 2015).

Two new applications of Csy4 are delineated in this dissertation. Chapter 2 details the development of Csy4 as a tool for regulating the stability and translation of transgene-derived RNA. Chapter 3 describes the utilization of Csy4 for inducing engineered circular RNA formation.

1.4 Sequence Elements Involved in RNA Stability

RNA stability is tightly controlled in cells using a wide range of methods. Of note, are 3' end stabilizing sequences, which protect the 3' ends of RNAs from exonucleolytic degradation. The majority of mammalian cellular mRNAs are terminated at the 3' end with a poly(A) tail. The poly(A) tail ensures mRNA stability through recruitment of the poly(A) binding protein Pab1p (Coller et al., 1998). Interestingly, not all eukaryotic mRNAs possess a poly(A) tail, and these RNAs must ensure stability by distinct mechanisms. For example, replication-dependent histone mRNAs possess a 3' terminal stem loop structure in place of a poly(A) tail. The persistence of these mRNAs is influenced by a set of proteins interacting with the stem loop and stem loop binding protein (SLBP) to promote either maturation or degradation in a cell cycle dependent manner (Marzluff et al., 2008).

In addition, triple helical elements have demonstrated the capacity to mediate 3' RNA end stabilization. Triple helix elements have been identified thus far in the 3' ends of non-coding and viral RNAs. The MALAT1 (metastasis-associated lung adenocarcinoma transcript 1) RNA, also known as NEAT2, is a long noncoding RNA which accumulates to high abundance in mouse and human cells (Hutchinson et al., 2007; Zhang et al., 2012). The MALAT1 RNA gains stability via an RNase P mediated

cleavage event to remove a tRNA-like structure called the mascRNA (MALAT1-associated small cytoplasmic RNA). The resulting 3' end of the MALAT1 transcript folds into a triple helical structure which is resistant to exonucleolytic degradation (Brown et al., 2012; Wilusz et al., 2012). MALAT1 triple helix-mediated stability can be conferred to an RNA of interest by simply placing the 3' end of the MALAT1 transcript at the 3' end of the RNA. Surprisingly, when placed downstream of an ORF lacking a poly-adenylation signal, the MALAT1 3' end is also capable of supporting protein translation (Wilusz et al., 2012).

Viruses, such as Kaposi's sarcoma-associated herpesvirus (KSHV) also make use of triple helix elements. In the KSHV polyadenylated nuclear (PAN) RNA, a triple helix is formed within an expression and nuclear retention element (ENE). The ENE contains a U-rich loop which sequesters and protects the PAN RNA poly(A) tail (Conrad et al., 2006; Mitton-Fry et al., 2010). The resulting triple helical structure stabilizes the PAN RNA and protects it from exonucleolytic degradation (Conrad et al., 2006). Structure-based bioinformatics analysis has since identified similar ENEs in a range of viral genomes (Tycowski et al., 2012). The occurrence of ENEs in a diverse set of viral genomes highlights the utility and value of this stabilizing structure.

1.5 Circular RNAs

Circular RNAs (circRNAs) are highly stable, covalently closed RNAs. Typically, circRNAs are generated through backsplicing, a process in which a downstream splicing donor joins with an upstream splicing acceptor (Jeck et al., 2013). However, an alternate mechanism which proceeds through a lariat intermediate has been proposed

(Zaphiropoulos, 1996). Another class of circRNAs, tRNA intronic circular RNAs (tricRNAs), are generated using tRNA maturation machinery (Lu et al., 2015).

Components of the canonical splicing machinery and splicing signals have been identified as factors involved in circRNA formation (Starke et al., 2015). However, circRNA biogenesis can be facilitated by a number of both trans- and cis-acting factors. For instance, circularization of a subset of circRNAs is regulated by the alternative splicing factor, Quaking, during the human epithelial to mesenchymal transition (Conn et al., 2015). Similarly, the muscleblind (MBL) protein is involved in circularization of the MBL circRNA. This observation might suggest an autoregulatory role for MBL in which excess of MBL favors the circularization of its flanked exon. This would in turn reduce the amount of linear, MBL encoding transcript (Ashwal-Fluss et al., 2014). Furthermore, circMBL itself contains MBL binding sites and is capable of binding MBL protein, contributing to the potential autoregulatory feedback loop (Ashwal-Fluss et al., 2014).

Cis-acting factors are also involved in circRNA biogenesis. Of recent interest are inverted repeat sequences (e.g. Alu repeats in humans) which have been found flanking exons known to circularize. These inverted repeat sequences are essential for the circularization of the intervening exon(s) (Ashwal-Fluss et al., 2014; Ivanov et al., 2015; Jeck et al., 2013; Kramer et al., 2015; Liang and Wilusz, 2014; Zhang et al., 2014). Circularization mediated by inverted repeats is thought to occur through a mechanism in which base pairing brings splice sites in close proximity to facilitate backsplicing. Consistent with this hypothesis, ADAR enzymes, which deaminate adenosine to inosine and unwind double-stranded regions of RNAs, inhibit circRNA formation (Ivanov et al., 2015).

Regardless of their mechanism of biogenesis, circRNA's are markedly more stable than their linear counterparts. The lack of exposed 3' and 5' ends likely renders circRNAs resistant to a number of cellular exonucleases. Consequently, this inherent stability leads to sustained persistence of circRNAs (Jeck et al., 2013; Lasda and Parker, 2014; Liang and Wilusz, 2014). This is exemplified by the median half-life of circRNAs, which is at least 2.5X longer than that of their linear counterparts (Enuka et al., 2015).

Deep sequencing studies have revealed differences in the abundance of various circRNAs, which are dictated by tissue, cell type, and developmental stage (Memczak et al., 2013). CircRNAs are known to be highly abundant in the mammalian brain compared to other tissues. In particular, they appear to be enriched at the synapses and synthesized during neuronal differentiation and development (Rybak-Wolf et al., 2015; You et al., 2015). Synaptic localization of certain circRNAs also suggests that they may be specifically targeted to different sites in the cell, as their linear counterparts are localized in the cytoplasm (Rybak-Wolf et al., 2015). This suggests possible specific roles for circRNAs at the synapse. Though they are often most highly expressed in the brain, circRNAs are also expressed in non-neuronal tissues. For example, the ZKSCAN1 circRNA is also detected in the liver while the HIPK3 circRNA is detected in the kidney, heart, lung, thyroid, and uterus. (Liang and Wilusz, 2014).

Despite their abundance, the functions for most circRNAs remains to be determined (Ebbesen et al., 2015). The roles of a small number of circRNAs have been proposed, but it is likely that there are a wide variety of possible functions. One hypothesis posits that the biogenesis of circRNAs may serve as a method of alternative

splicing regulation. This regulation would be carried out by affecting exon skipping, due to the mutually exclusive relationship between circular and linear splicing of an exon (Ashwal-Fluss et al., 2014). Alternatively, some circRNAs such as circHIPK3, ciRS-7 (Cdr1as) and the Sry circRNA act as microRNA (miRNA) sponges (Hansen et al., 2013; Memczak et al., 2013; Zheng et al., 2016). However, most circRNAs are not enriched for miRNA binding sites and this is unlikely to be a general function of circRNAs (Guo et al., 2014). As described above, circMbl has been proposed as a sponge for MBL protein, and some circRNAs containing introns (EIciRNAs) have been implicated in transcription regulation (Ashwal-Fluss et al., 2014; Li et al., 2015). Additionally, circFoxo3 appears to bind the cell cycle regulators CDK2 (cyclin dependent kinase 2) and p21 to repress cell cycle progression (Du et al., 2016b). CircRNAs have also been proposed as templates for retrotranscription and reinsertion into the genome to generate pseudogenes (Dong et al., 2016).

Although most circRNAs arise from protein-coding exons, endogenously encoded circRNAs are not associated with ribosomes in the cytosol and no examples of translated endogenous circRNAs have been described (Capel et al., 1993; Guo et al., 2014; Jeck et al., 2013). However, studies suggest circRNAs can be engineered to express a protein of interest if an internal ribosome entry site (IRES) is included to drive translation (Chen and Sarnow, 1995; Wang and Wang, 2015). The inherent stability of circRNAs and variety of potential natural and engineered functions for circRNAs make these molecules attractive tools expression of therapeutic genes *in vivo*.

1.6 Adeno-Associated Virus-Mediated Gene Delivery and RNA Therapeutics

Adeno-Associated Virus (AAV) is a helper-dependent parvovirus with a 4.7 kb single stranded DNA genome. AAV utilizes three promoters to express a series of proteins though overlapping reading frames in Rep and Cap genes. The genome is flanked on either end by inverted terminal repeat (ITR) sequences which are essential for genome packaging into the viral capsid (Figure 3A). The ITRs are the only cis-elements required for genome packaging. As such, AAV can be converted into a recombinant vector for gene therapy by replacing the intervening sequence with an expression cassette of interest (Figure 3B).

As a gene therapy vector, recombinant AAV (rAAV) is particularly valuable due to long-term transgene expression, lack of pathogenicity, and inability to replicate independently (Samulski and Muzyczka, 2014). Further, the tropism of a number of serotypes is well defined and can be engineered through capsid manipulations (Asokan et al., 2012; Pulicherla et al., 2011a; Shen et al., 2013b). rAAV can be generated and purified through a triple plasmid transfection protocol in HEK293 cells, and can package any ITR-flanked cassette less than 4.7 kb in size (Xiao et al., 1998). These features make rAAV an ideal vector for delivery of therapeutic genes in mice and potentially higher order mammals. In particular, the well-studied tropism of various AAV serotypes allows for direct targeting of specific tissues and cell types. This prevents exposing the entire organism to the therapeutic in situations where it might not be necessary or may be detrimental.

rAAV-based therapeutics are poised for the treatment of a range of genetic diseases. Current studies are exploring the potential of rAAV-mediated delivery of

treatments for factor IX deficiency (hemophilia B), lysosomal storage diseases, and alpha-1-antitrypsin (AAT) deficiency, among others (Brimble et al., 2016; Hinderer et al., 2015; Samulski and Muzyczka, 2014; Wozniak et al., 2015). rAAV-mediated delivery of Cas9 and gRNA has shown promise for treatment of Duchenne's muscular dystrophy (Long et al., 2016; Nelson et al., 2016; Tabebordbar et al., 2016). Additionally, rAAV-based treatment developed for Leber's congenital amaurosis, a form of blindness, has been successful (Cideciyan et al., 2008; Maguire et al., 2008). An rAAV-delivered treatment for lipoprotein lipase deficiency (Glybera) has been approved in Europe, and represents the first approved rAAV therapeutic in the West (Bryant et al., 2013; Gaudet et al., 2010).

rAAV vectors have been successfully used for the expression of RNA therapeutics such as short hairpin RNAs (shRNA) and miRNAs in a number of animal models (Borel et al., 2014). While well-tolerated in the majority of cases, overexpression of AAV-shRNA vectors have been shown to cause hepatotoxicity or neurotoxicity in some examples (Grimm et al., 2006; van Gestel et al., 2014). This is likely due to their potential to saturate the miRNA pathway and consequently alter the normal cellular miRNA composition (Grimm et al., 2006; Valdmanis et al., 2016; van Gestel et al., 2014). Further, though many clinical trials utilizing rAAV for expression of therapeutic proteins have proved encouraging, vector dose-related toxicity remains a concern (Nathwani et al., 2014). CircRNA expression cassettes may help to alleviate dose-toxicity issues through stable expression of therapeutic genes at potentially lower clinical doses. Chapter 4 describes methods for AAV-mediated delivery of circRNA cassettes.

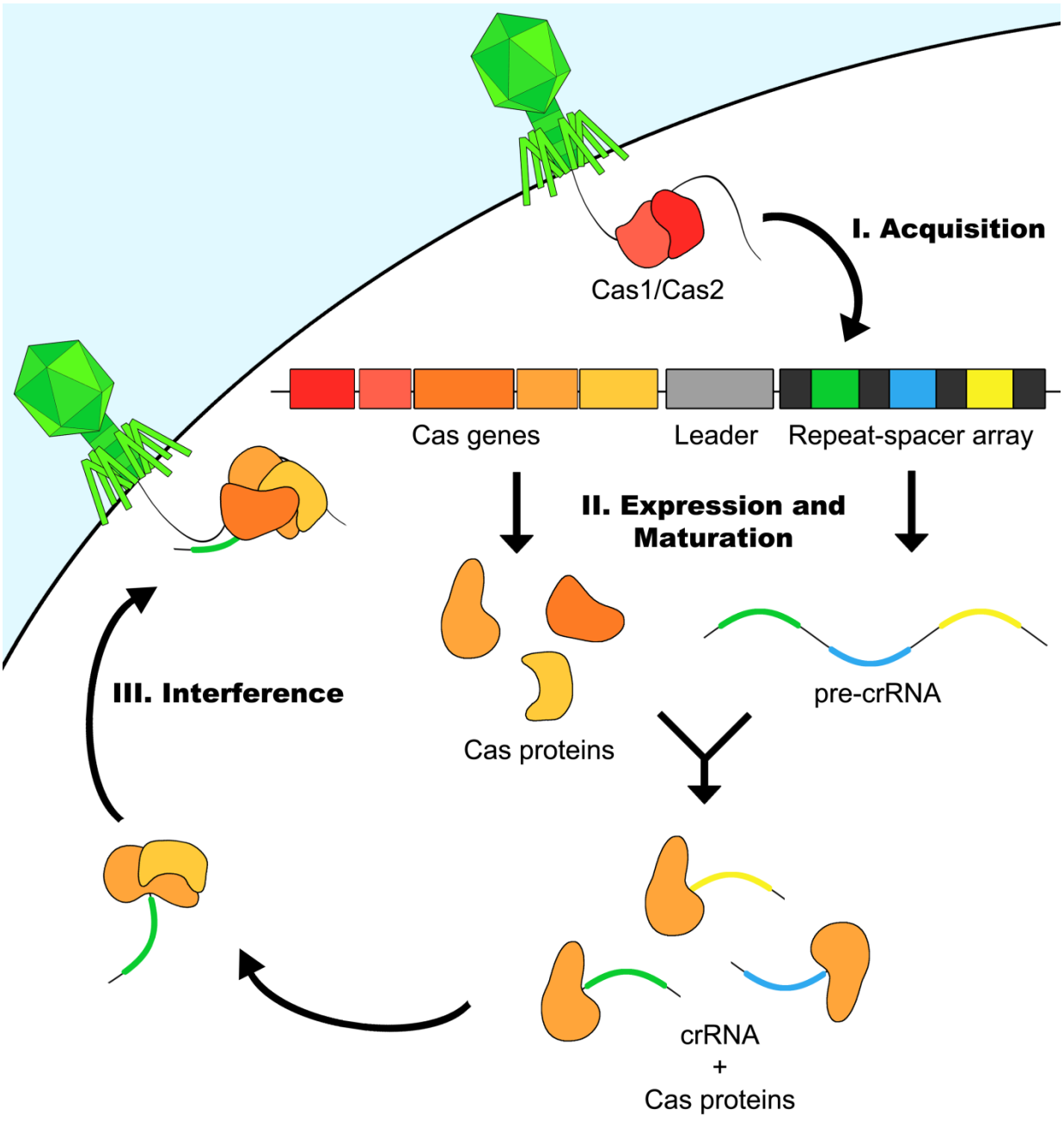

Figure 1: General schematic of CRISPR mediated immunity. Upon encountering foreign nucleic acid, such as from a phage genome, Cas1 and Cas2 mediate incorporation of a new spacer into the repeat-spacer array (acquisition). The repeat-spacer array is expressed as a pre-crRNA which is processed into mature crRNAs by Cas proteins (expression and maturation). A Cas nuclease is guided to target nucleic acid by crRNA and mediates nucleolytic cleavage (interference).

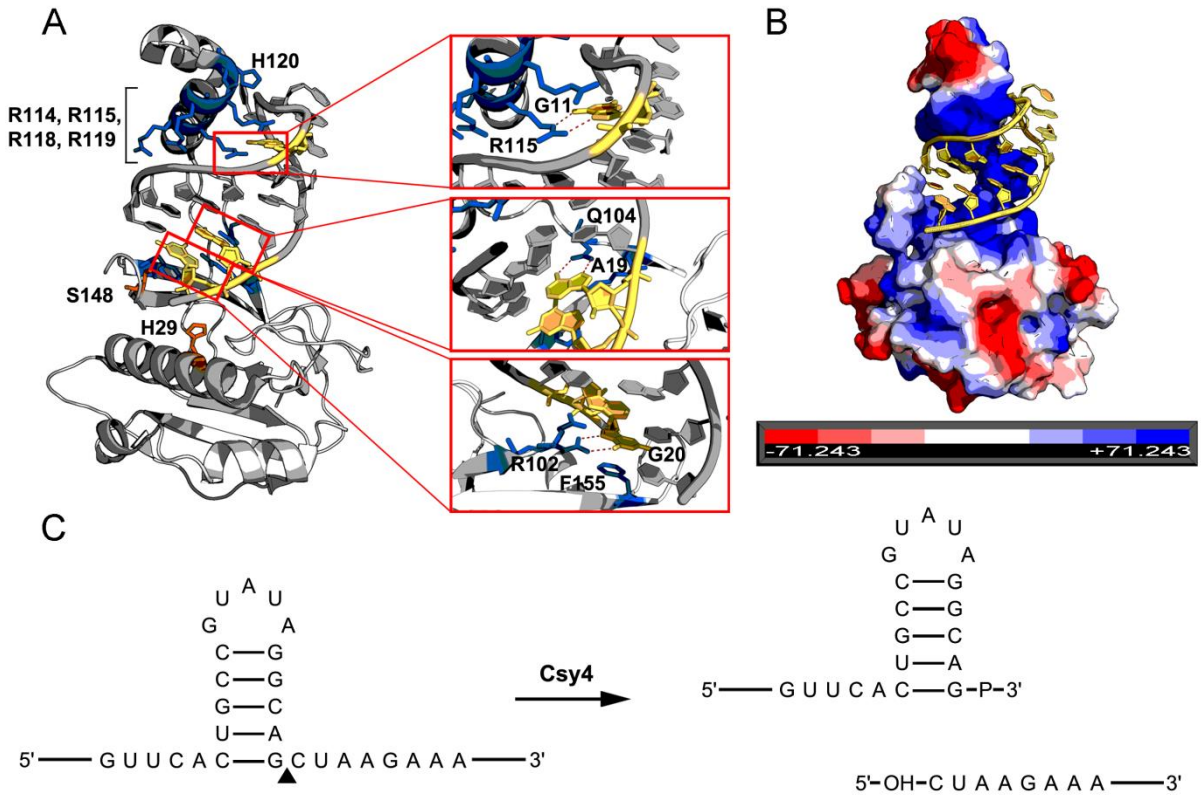

Figure 2: Characteristics the Csy4-hairpin interaction. (A) PyMOL rendering of Csy4 bound to the hairpin sequence. Catalytic residues Ser148 and His29 are displayed as sticks in orange. Residues making structure- and sequence-specific contacts are shown as sticks in blue. Hairpin ribonucleotides that are recognized in a sequence-specific manner are displayed in yellow. Sequence-specific interactions are enlarged to the right in boxes. **(B)** Charge-smoothed surface representation model of Csy4 interacting with the hairpin sequence. Renderings of Csy4 in A and B are based off of PDB coordinates 2XLK (Haurwitz et al., 2010). **(C)** Substrate (left) and products (right) of Csy4-mediated cleavage.

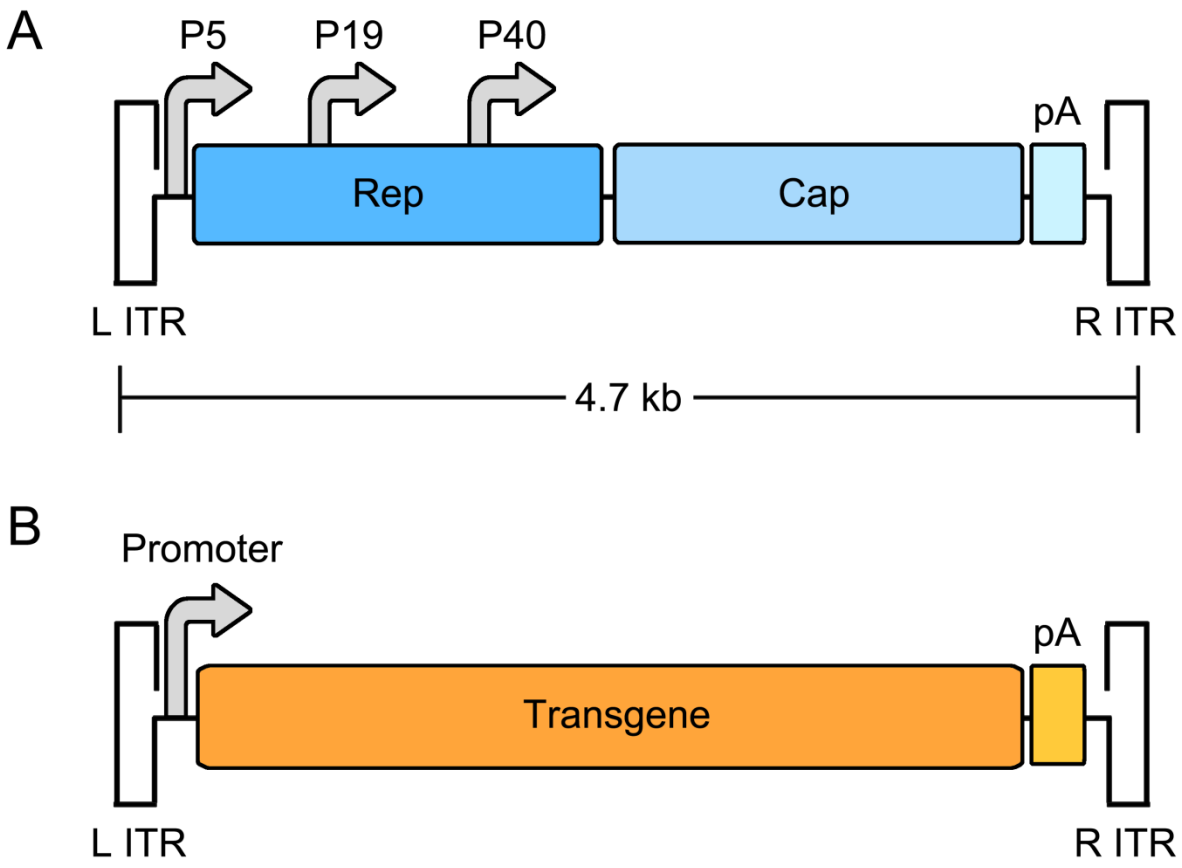

Figure 3: Genomic maps of AAV and recombinant AAV. (A) In the AAV genome, inverted terminal repeats (ITRs) flank Rep and Cap genes, which are expressed from three promoters and terminated with a poly(A) signal. **(B)** Recombinant AAV genomes can be generated by replacing the intervening sequence between the ITRs with an expression cassette of interest.

CHAPTER 2: CONTROLLING MRNA STABILITY AND TRANSLATION WITH THE CRISPR ENDORIBONUCLEASE CYS4[1]

2.1 Overview

The bacterial CRISPR endoribonuclease, Csy4 has recently been described as a potential RNA processing tool. Csy4 recognizes substrate RNA through a specific 28 nucleotide hairpin sequence and cleaves at the 3' end of the stem. To further explore applicability in mammalian cells, we introduced this hairpin at various locations in mRNAs derived from reporter transgenes and systematically evaluated the effects of Csy4-mediated processing on transgene expression. Placing the hairpin in the 5' untranslated region (UTR) or immediately after the start codon resulted in efficient degradation of target mRNA by Csy4 and knockdown of transgene expression by 20 to 40-fold. When the hairpin was incorporated in the 3' UTR prior to the poly(A) signal, the mRNA was cleaved, but only a modest decrease in transgene expression (~2.5 fold) was observed. In the absence of a poly(A) tail, Csy4 rescued the target mRNA substrate from degradation, resulting in protein expression, which suggests that the cleaved mRNA was successfully translated. In contrast, neither catalytically-inactive (H29A) nor binding-deficient (R115A/R119A) Csy4 mutants were able to exert any of the above-described effects. Generation of a similar 3' end by RNase P-mediated cleavage was unable to rescue transgene expression independent of Csy4. These

[1]This chapter previously appeared as an article in the journal RNA. The original citation is as follows: Borchardt, E.K., Vandoros, L.A., Huang, M., Lackey, P.E., Marzluff, W.F. and Asokan, A. 2015. Controlling mRNA stability and translation with the CRISPR endoribonuclease Csy4.*RNA* **21**: 1921-1930.

results support the idea that the selective generation of the Csy4/hairpin complex resulting from cleavage of target mRNA might serve as a functional poly(A)/Poly-A Binding protein (PABP) surrogate, stabilizing the mRNA and supporting translation. Although the exact mechanism(s) remain to be determined, our studies expand the potential utility of CRISPR nucleases as tools for controlling mRNA stability and translation

2.2 Introduction

Endogenous control of gene expression is achieved by regulating transcription, processing, translation and/or degradation of mRNA through a myriad of genetic elements. Artificial control of gene expression, on the other hand, requires the development of small molecule, protein or RNA-based tools and is essential for advancing synthetic biology and gene therapies. Post-transcriptional regulation of gene expression has been achieved by engineering RNA, for instance by employing riboswitches that can not only be exploited to gain insight into endogenous RNA processing mechanisms, but also as programmable elements for building gene circuits (Chang et al., 2012). Another promising platform for developing RNA regulatory tools are the Pumilio/fem-3 binding factor or PUF proteins that recognize single stranded RNA in a sequence-specific fashion (Wang et al., 2013). PUF domains can be exploited to engineer splicing factors, control translation or develop artificial site-specific RNA endonucleases (Choudhury et al., 2012). Recent studies have suggested that bacterial CRISPR systems might be useful for RNA regulatory applications in comparison to the use of hammerhead ribozymes and RNase III (Qi et al., 2012). Regardless of the

approach, controlling RNA stability and translation is a key aspect underlying these gene regulatory strategies.

The prokaryotic CRISPR (Clustered Regularly Interspaced Short Palindromic Repeats) system provides adaptive immunity against invading genetic elements in ~45% of bacteria and ~83% of archaea (Grissa et al., 2007). Briefly, CRISPR loci encode a repeat-spacer array consisting of unique sequence elements (spacers), derived from previously encountered exogenous nucleic acid (Barrangou et al., 2007a; Bolotin et al., 2005; Mojica et al., 2005; Pourcel et al., 2005). Each spacer is flanked by repeat elements and the entire array is transcribed as a single pre-crRNA (Tang et al., 2002; Tang et al., 2005). The pre-crRNA is processed into smaller crRNAs which serve as guides to target nucleic acid for neutralization by CRISPR-associated (Cas) proteins (Brouns et al., 2008; Garneau et al., 2010a). The protein content of CRISPR loci is diverse and as such, two broad classes of CRISPR systems have been described (Makarova et al., 2015). Classification is based on the utilization of either a single protein or a multi-subunit protein complex for target interference (Makarova et al., 2015). These groups are further divided into types and sub-types denoted by roman numerals and alphabetical letters respectively. The *Pseudomonas aeruginosa* Class 1 Type-IF CRISPR system employs Csy4 (Cas6f) to bind a 28-nucleotide repeat hairpin and cleave at the base of the 3' end to generate crRNAs (Cady and O'Toole, 2011; Haurwitz et al., 2010). The task of pre-crRNA processing in other systems falls to Cas5d, Cas6e (CasE/Cse3) and Cas6 in *Bacillus halodurans, Escherichia coli* and *Pyrococcus furiosus* respectively (Brouns et al., 2008; Carte et al., 2008; Nam et al., 2012).

Structural studies have provided further mechanistic insight into pre-crRNA processing enzymes (Hochstrasser and Doudna, 2014). Despite their functional similarity, these enzymes display minimal primary sequence homology. Likewise, the sequences which they process also differ both in sequence and structure, with Cas5d, Cas6e, and Csy4 associated repeat elements containing hairpin structures and Cas6 targeting a predicted unstructured sequence (Kunin et al., 2007). To carry out target neutralization, Csy4 remains bound to the processed crRNA and associates with additional Cas proteins, Csy1, Csy2, and Csy3 for target recognition (Wiedenheft et al., 2011). This complex is guided to target DNA based on sequence complementarity provided by the crRNA. In Class 1 Type-I CRISPR systems, Cas3 is then recruited to cleave and degrade the target DNA, neutralizing the invader (Sinkunas et al., 2013; Westra et al., 2012).

Recent studies have adapted the Class 1 Type-IF CRISPR endoribonuclease, Csy4 for processing of RNA encoding multigene operons and gene regulation in *E. coli, Bacillus subtilis,* and *Saccharomyces cerevisiae* (Qi et al., 2012). In particular, this study demonstrated that RNA processing by Csy4 is an effective strategy to maintain the endogenous function of different promoters, genes and regulatory elements without interference when engineered into complex genetic circuits. More recently, Csy4 has been applied for processing guide RNA (gRNA) for mammalian Cas9 technology and the programming of complex genetic circuits (Nissim et al., 2014; Tsai et al., 2014). These advancements have led to the expression of multiple gRNAs from a single promoter, removed the restriction of a 5' guanosine imposed by the U6 promoter, and permitted gRNA expression from RNA polymerase II promoters. Furthermore, Csy4 has

been utilized in the isolation of RNA-interacting proteins and has the potential to help analyze the protein-associations of diverse transcripts (Lee et al., 2013). In the current study, we systematically evaluated the ability of Csy4 to exercise post-transcriptional control of transgene expression in mammalian cells. Specifically, we investigate the positional effects of Csy4 mediated-cleavage in the 5' untranslated region (UTR), coding sequence, and 3'UTR of transcripts. Surprisingly, we find that the Csy4 processing of the 3' ends of mRNA supports translation and stabilizes the mRNA in lieu of a poly(A) signal.

2.3 Materials and Methods

Plasmids. The Csy4 gene was amplified from *P. aerugionsa* strain UCBPP-PA14 genomic DNA and cloned under the control of the chicken beta-actin (CBA) promoter using the following primers: Csy4wt Forward 5'- ATC GTC TAG AAT GGA CCA CTA CCT CGA CAT TCG CTT GC-3' and Csy4wt Reverse 5'- CGA TGC GGC CGC TCA GAA CCA GGG AAC GAA ACC TCC TTT GC-3' (IDT DNA Technologies). *P. aeruginosa* UCBPP-PA14 genomic DNA was kindly provided by Dr. Matthew Wolfgang (University of North Carolina at Chapel Hill). Csy4-H29A was amplified from pHMGWA-Pa14Csy4H29A (Addgene plasmid #41092) which was provided as a gift from Dr. Jennifer Doudna (Haurwitz et al., 2010). Csy4-H29A was cloned under the control of the CBA reporter. Csy4 R115A/R119A was synthesized as a gBlock ® (IDT DNA Technologies) and cloned under the control of the CBA promoter.

The reporter plasmid constructs containing enhanced green fluorescent protein (GFP) and Gaussia luciferase (GLuc) were modified to incorporate the Csy4 hairpin

(HP) in different locations as follows. (i) Reporter cassettes containing the HP after the start codon (ATG-HP) were generated using overlap extension PCR with primers that generated two fragments containing a single, in-frame Csy4 HP repeat sequence (5'-AGTTCACTGCCGTATAGGCAGCTAAGAAAT-3') in the 3' or 5' end. The two fragments were then gel purified and combined in equimolar quantities (40ng each) in consecutive PCR reactions without primers (35 cycles) and with flanking primers (30 additional cycles) to obtain the ATG-HP construct. The latter PCR products were then cloned under the control of the CBA promoter. (ii) The 5'UTR-HP-GFP was constructed by overlap extension PCR in a similar fashion, while the 5'UTR-HP-GLuc was synthesized as a gBlock ® (IDT DNA Technologies) and cloned under the control of the CBA promoter. (iii) 3'UTR-HP reporters and poly-A deleted 3'UTR-HP (3'UTR-HP-Δp(A)) reporters were synthesized from gBlocks® and cloned as described earlier. The cGFP-mMALAT1-3' reporter was generously provided by Dr. Jeremy Wilusz (University of Pennsylvania). A partial Csy4 hairpin (5'-GTT CAC TGC CGT ATA GGC AG-3') and mascRNA were synthesized as a gblock ® (IDT DNA Technologies) and cloned in place of the mMALAT1 3' UTR in the cGFP-mMALAT1-3' reporter to generate cGFP-HP-masc-Δp(A). Similarly, a second gblock ® (IDT DNA Technologies) was synthesized containing the partial Csy4 hairpin and mascRNA separated by 10 nucleotides (5'-CTA AAC GCG T-3') and cloned as described earlier, to generate cGFP-HP_{10}-masc-Δp(A).

Cell culture. HEK293 cells were cultured in Dulbecco's Modified Eagle's Medium (GIBCO/Life Technologies) supplemented with 10% FBS (Atlanta Biologicals), 1%

Penicillin/Streptomycin and 2.5 µg/mL amphotericin B (Sigma-Aldrich) and maintained at 37°C and 5% CO_2.

Transfection and luciferase reporter assays. Equimolar amounts (totaling 500ng) of three plasmids (different HP GLuc reporters, Csy4 and a control plasmid containing the tdTomato reporter driven by the CBA promoter) were transfected into HEK293 cells using PEI Max seeded at a density of $5x10^4$ per well in a 24 well plate. Media (50 µl) was collected from each well at different time intervals and diluted (1:100) before assessing luciferase reporter activity. For measuring GLuc activity, native coelenterazine (Nanolight) was dissolved in methanol to a concentration of 1 mg/mL and diluted (1:200) in 600 mM NaCl-Tris-EDTA buffer, following which, 50 µl of the substrate solution was added to 50 µl of collected media. Luminometric analysis was carried out using a Perkin Elmer Victor 3 ® plate reader.

Fluorescence microscopy. HEK293 cells were transfected with different HP GFP reporter cassettes as described earlier and the cells imaged at different time intervals post-transfection using an EVOS ® FL epifluorescence cell imaging system (AMC/Life Technologies) using the GFP light cube (excitation 470nm, emission 510nm).

mRNA analysis. Processing of mRNA was analyzed by monitoring the levels of DNA obtained through reverse transcriptase (RT) PCR. Briefly, HEK293 cells seeded overnight in 6 well plates at a density of $3x10^5$ per well were transfected with a total of 2.5µg of different plasmids (HP GFP reporters, Csy4wt or tdTomato control) as outlined

earlier. Total RNA from each well was isolated at 48 hours post-transfection using the Total RNA Purification Kit (#17200, Norgen Biotek). The purified RNA was then treated with DNase using the Turbo DNA-free kit (Ambion/Life Technologies). Equal ng amounts of the purified total RNA product were utilized as template for reverse transcriptase PCR using the High Capacity RNA-to-cDNA kit (Applied Biosystems/Life Technologies). Products of this reaction were used as template for further PCR amplification with gene specific primers and visualized on an agarose gel. Forward primer for amplifying ATG-HP-GFP cDNA: 5'-GCC ACC ATG AGT TCA CTG CCG-3'; Forward and reverse primers for amplifying all other GFP reporter cDNAs 5'-GAA ATG TGA GCA AGG GCG AGG AGC-3'; 5'-GCG GAC TTG AAG AAG TCG TGC TGC-3'; Forward and reverse primers for glyceraldehyde 3-phosphate dehydrogenase (GAPDH) cDNA: 5'-CCA CTC CTC CAC CTT TGA C-3'; 5'-ACC CTG TTG CTG TAG C-3'.

Detection of poly-adenylated mRNA. HEK293 cells seeded overnight in 6 well plates at a density of $3x10^5$ cells per plate were transfected with a total of 3 µg DNA as indicated. Prior to transfection, DNA was prepared by digestion with ClaI, BglII, and CIP (New England Biolabs) to remove the GLuc expression cassette from its backbone and to prevent interference from potential downstream cryptic poly(A) signals. At 24 hours post-transfection, RNA was harvested using Total RNA Purification Kit (#17200, Norgen Biotek) and DNase treated using the Turbo DNA-free kit (Ambion/Life Technologies). A total of 1.3 µg DNase-treated RNA was converted to cDNA using SMARTScribe reverse transcriptase (Clontech) according to the manufacturer's protocol. An oligo dT_{20} primer was used to prime the reverse transcription reactions. Products of this reaction were

utilized as template for further PCR amplification with gene specific primers and visualized on an agarose gel. Forward and reverse primers for reporter cDNAs (GLuc) 5'-CAA CTT CGC GAC CAC GGA TCT CG-3'; 5'-CGG CAG CCA CTT CTT GAG CAG G-3'. Forward and reverse primers for GAPDH are listed above.

RNA immunoprecipitation. HEK293 cells seeded overnight in 15 cm plates at a density of $5x10^6$ cells per plate were transfected with 10 ug of His6-Csy4-H29A and 10 ug of ATG-HP-GFP plasmids. Cells were lysed 48 hours post-transfection and RNA immunoprecipitation (RNA-IP) was carried out using the Magna RIP RNA Binding Protein Immunoprecipitation Kit (Millipore) according to the manufacturer's protocol (α-His6 antibody, ab18184, Abcam). Purified RNA was DNase treated using the Turbo DNA free kit (Ambion/Life Technologies) and cDNA was generated from treated RNA via the Peregrine method (Langevin et al., 2013). cDNA was used as template for PCR with GFP specific primers 5'-GCC ACC ATG AGT TCA CTG CCG-3'; 5'-GCG GAC TTG AAG AAG TCG TGC TGC-3' and PCR products were visualized on an agarose gel.

Northern Blot. HEK293 cells seeded overnight in 10 cm plates at a density of $2.2x10^6$ cells per plate were transfected with a total of 6 µg DNA as indicated. RNA was purified using Trizol reagent (Invitrogen/Life Technologies) and treated with DNase using the Turbo DNA-free kit (Ambion/Life Technologies). 1 microgram of RNA was separated on a 4% polyacrylamide/8M urea gel and transferred to Hybond N+ membrane (GE Healthcare). To generate radiolabeled probe, cGFP was digested out of the cGFP-MALAT1-3' backbone and used as template for the incorporation of radiolabeled CTP

via a Random Primed DNA Labeling Kit (Roche Diagnostics). Hybridization of radiolabeled probe to the membrane was carried out using Rapid-Hyb buffer (GE Healthcare).

Western Blot. HEK293 cells seeded overnight in 6 well plates at a density of $3x10^5$ cells per plate were transfected with a total of 2.5 µg DNA as indicated. Lysates were recovered 48 hours post-transfection using Passive Lysis Buffer (Promega). Lysates were diluted 1:50 and separated on a 10% Bis-Tris gel. Blots were probed with mouse monoclonal anti-GFP antibody (1:000 dilution, Santa Cruz) or mouse monoclonal anti-Actin (1:2000 dilution, Abcam) as primary antibody. Stabilized peroxidase-conjugated goat anti-mouse antibody (1:5000 dilution, ThermoScientific/Life Technologies) was used as secondary antibody. Blots were developed using SuperSignal West Femto substrate (ThermoScientific/Life Technologies).

2.4 Results and Discussion

Csy4-mediated knockdown of 5' UTR-hairpin (HP) and ATG-HP constructs.

We first evaluated the effect of Csy4-mediated cleavage of the substrate HP incorporated within the 5' untranslated region (UTR) or HP inserted in-frame immediately following the start codon of reporter transcripts. To this end, we cloned the Csy4 target hairpin (Figure 4A) derived from the *P. aeruginosa* strain UCBPP-PA14 CRISPR locus, upstream (5'UTR-HP) or immediately downstream (ATG-HP) of the initiator AUG in two separate reporter genes, GFP and Gaussia Luciferase (GLuc) and compared the activity of these constructs to the control cassettes lacking the Csy4

hairpin (Figure. 4B). The reporters were transfected into HEK293 cells with or without a plasmid expressing Csy4. While no change in mRNA levels was observed as indicated by random primed RT-PCR in the case of the control plasmid lacking the Csy4 HP substrate, we observed a marked decrease in the case of the 5'UTR-HP and the ATG-HP templates (Figure 4C). This observation was supported by a decrease in GFP signal from the 5'UTR-HP and ATG-HP templates, but not the control substrate as assessed by fluorescence microscopy (Figure 4D). Moreover, there was a ~24-fold decrease in GLuc reporter transgene expression mediated by Csy4 for the 5'UTR-HP construct and a ~37-fold decrease in signal for the ATG-HP construct (Figure 4E). In contrast, the control cassette displayed a minimal change in gene expression in the presence of Csy4. It is important to consider that the ATG-HP approach results in the incorporation of a 10 amino acid residue peptide tag with the sequence SSLPYRQLRN into the GFP or luciferase proteins. Further, it should be noted that we also observed a significant decrease in GLuc reporter activity of the 5'UTR-HP reporter construct in comparison with the unmodified or ATG-HP GLuc construct, even in the absence of Csy4. This is likely due to decreased efficiency of translational initiation by incorporation of a RNA hairpin structure in the 5'UTR region (Babendure et al., 2006; Kozak, 1986; Kozak, 1989). These results are consistent in part with studies in *S. cerevisiae*, *B. subtilis,* and *E. coli* (Qi et al., 2012). In summary, our results confirm that Csy4 is functional in mammalian cells and can process RNA in a highly selective fashion mediated by the cognate hairpin substrate.

Effect of placing the HP in the 3'UTR on Csy4-mediated knockdown.

In contrast to insertions near the 5'UTR or the start codon, placement of the Csy4 target HP following the stop codon and prior to the poly(A) signal (3'UTR-HP, Figure 5A, left column) resulted in only a modest decrease in mRNA levels (as indicated by random primed RT-PCR) and GFP expression, when co-expressed with Csy4 (Figures 5B, 5C, left columns). Further, quantitation of GLuc reporter activity indicated only a ~2.5 fold reduction in transgene expression upon treatment with Csy4 (Figure 5D, left column). However, there was a substantial reduction in poly-adenylated reporter RNA levels as measured by oligo-dT_{20} primed RT-PCR, consistent with mRNA cleavage by Csy4 (Figure 5E, left column). These results are particularly intriguing, since removal of the poly(A) signal is expected to destabilize mRNA.

The latter observations might arise from the fact that following cleavage Csy4 remains bound to the cognate HP substrate (Haurwitz et al., 2010), which in turn could protect the 3' end of the transcript from degradation, despite removal of the poly(A) tail. Since translation is only slightly reduced, the Csy4/RNA complex must also be compatible with translation. To test this hypothesis, we generated a second set of reporters identical to the 3'UTR-HP constructs, but lacking a poly(A) signal (3'UTR-HP-Δp(A)) (Figure 5A, right column). As seen in Figure 5B (right column), due to the lack of an efficient 3' processing signal, the 3'UTR-HP-Δp(A) constructs accumulate only a small amount of RNA. However, when co-expressed with Csy4, we observed an increase in mRNA levels by random primed RT-PCR, as well as an increase in GFP transgene expression (Figures 5B, 5C, right columns). As described earlier, we then assessed the poly-adenylation status of this reporter mRNA by generating reverse

transcribed cDNA using an oligo-dT_{20} primer. Regardless of the presence or absence of Csy4, 3'UTR-HP-Δp(A) reporter mRNA levels primed by oligo-dT_{20} appeared low relative to 3'UTR-HP (Figure 5E, right column). This is consistent with the 3'UTR-HP-Δp(A) RNA lacking a poly(A) signal. Importantly, Csy4 co-expression does not alter the amount of poly-adenylated RNA detected. Furthermore, quantitation of GLuc activity revealed a 9-fold increase in signal with Csy4 co-expression relative to a control lacking Csy4 (Figure 5D). This corroborates the notion that Csy4 can potentially stabilize mRNA containing the HP substrate in the 3'UTR and support translation.

Csy4 binding and cleavage are necessary for regulation of transgene expression.

We next sought to determine whether Csy4 binding to the substrate HP alone was sufficient or if binding followed by enzymatic cleavage was essential for regulation of transgene expression. Specifically, we used two mutants of Csy4; Csy4-R115A/R119A, a binding deficient mutant that still retains catalytic activity, and Csy4-H29A, a catalytically inactive, but binding competent mutant (Haurwitz et al., 2010; Sternberg et al., 2012).

We first demonstrated Csy4 association with the target hairpin via RNA immunoprecipitation (RNA-IP). Hemagglutainin (HA)-tagged Csy4-H29A was expressed with the ATG-HP-GFP reporter and Csy4-RNA complexes were isolated from cell lysates using an α-His6 antibody (Figure 6A). We were able to detect GFP RNA from the immunoprecipitates, indicating that Csy4 is interacting with the hairpin-tagged reporter RNA in mammalian cells (Figure 6B).

To evaluate the role of Csy4 binding and cleavage in regulating transgene expression, each reporter construct including the unmodified control, 5'UTR-HP, ATG-HP, 3'UTR-HP and the 3'UTR-HP-Δp(A) reporter (Figure 7A) was expressed in the presence or absence of native Csy4, Csy4-H29A or Csy4-R115A/R119A. Expression from unmodified GFP and GLuc constructs was unaffected by either mutant. The 5'UTR-HP GFP construct was also unaffected by co-expressing either the catalytically inactive or the binding deficient mutant (Figure 7B, 7D, 7E). Similar observations were made in case of the ATG-HP GFP reporter expressed with the catalytically inactive mutant, suggesting that steric hindrance from binding alone was insufficient to block translation initiation (Figure 7B, 7D). Further, the Csy4-R115A/R119A mutant also did not seem to substantially affect ATG-HP reporter expression (Figure 7B, 7E). Treatment with Csy4-H29A and Csy4-R115A/R119A did not dramatically affect expression of the corresponding GLuc reporters, supporting the previous data (Figure 7F). Taken together, these data suggest that both binding and cleavage are essential for Csy4-mediated knockdown of transgene expression. In addition, neither mutant was able to rescue reporter signal in case of the 3'UTR-HP and the 3'UTR-HP-Δp(A) reporters (Figures 7B-F). These results support the notion that rescue of transgene expression from reporter constructs lacking a poly(A) tail requires both Csy4 binding to and cleavage of target mRNA. Thus, Csy4 bound to the cleaved hairpin at the 3' end of an mRNA appears to have the remarkable property of stabilizing the mRNA, translocating to the cytoplasm, and supporting translation.

Selective processing by Csy4 is essential for rescue of 3'UTR-HP-Δp(A) constructs.

To further confirm the role of Csy4-mediated recognition and cleavage in rescuing 3'UTR-HP-Δp(A) reporters, we attempted to generate a similar cleaved 3' end product using a different endoribonuclease. Specifically, we engineered a novel reporter cassette (cGFP-HP-masc-Δp(A)), wherein a tRNA-like structure (mascRNA) was placed adjacent to the Csy4 HP (Figure 8A). This mascRNA motif is derived from the 3' end of a previously described MALAT1 RNA and is selectively cleaved by RNase P (Wilusz et al., 2008), releasing the capped 5' region of the transcript. In the endogenous MALAT1 RNA, the 3' end generated by RNase P processing folds into a triple helix structure that is capable of stabilizing the RNA and supporting translation in the absence of a poly(A) signal (Brown et al., 2012; Wilusz et al., 2012). The cGFP-HP-masc-Δp(A) construct was engineered in such a manner that RNase P would cleave at the same nucleotide as Csy4, at the base of the Csy4 HP stem, releasing the mascRNA motif, but generating the same RNA product as Csy4 would generate. In addition, we also constructed a cassette, wherein the Csy4 HP and the mascRNA motif were separated by 10 nucleotides to allow RNA processing at two different sites (Figure 8A, cGFP-cHP$_{10}$-masc-Δp(A)). In this construct, the initial RNaseP product can be subsequently cleaved by Csy4. A construct containing the full MALAT1 3’ end, including the 3' end stabilizing triple helix element in place of the Csy4 hairpin was used as a control (Figure 8A, cGFP-mMALAT1-3').

When transfected into HEK293 cells, cGFP-mMALAT1-3' reporter expression was not affected by co-expression with Csy4 wild type (Figures 8B, 8C). This is

expected due to the absence of a Csy4 target hairpin in this construct. In contrast, neither of the hairpin-containing reporters, cGFP-HP-masc-Δp(A) or cGFP-HP_{10}-masc-Δp(A), alone was able to support transgene expression due to the lack of a poly(A) signal or the MALAT1 triple helical motif (Figure 8B). However, co-expression of native Csy4 with these reporters rescued GFP expression, consistent with earlier results demonstrating the ability of Csy4 to stabilize the mRNA 3' end (Figure 8C). Northern blot using a cGFP probe confirmed these observations (Figure 8E). When expressed independently of Csy4, no cGFP RNA is detected for cGFP-HP-masc-Δp(A) or cGFP-HP_{10}-masc-Δp(A) (Figure 8E, lanes 2 and 4). However, when co-expressed with Csy4 (lanes 3 and 5), a band is detected for each reporter that runs slightly below that of cGFP-mMALAT1-3' (Figure 8E, lane 1). This is expected as the cleavage products of cGFP-HP-masc-Δp(A) and cGFP-HP_{10}-masc-Δp(A) are expected at 754 nt while the processed cGFP-mMALAT1-3' transcript runs at 844 nt. Our observations were further confirmed by western blot for cGFP on lysates expressing cGFP-mMALAT1-3' (lane 1), cGFP-HP-masc-Δp(A) (lane 2), or cGFP-HP-masc-Δp(A) with Csy4 (lane 3) (Figure 8F). Further, we expressed the cGFP-mMALAT1-3' construct, cGFP-HP-masc-Δp(A) or cGFP-HP_{10}-masc-Δp(A) construct with the catalytically inactive Csy4-H29A mutant. In this setting, the reporter mRNA would only be processed by RNase P to generate a 3' end, however, Csy4-H29A is still expected to bind the HP substrate. As seen in Figure 8D, while cGFP-mMALAT1-3' expression is unaffected, rescue of GFP transgene expression from both Csy4 HP-containing constructs using the catalytically inactive Csy4 mutant was markedly reduced. Interestingly, a few cells expressing GFP were still detectable, when Csy4-H29A was expressed with these reporter constructs. This

observation suggests that generation of a similar 3' end using RNase P in conjunction with binding by the Csy4-H29A mutant can at least partially rescue expression as seen with wild type Csy4. It should be noted that Csy4-mediated cleavage of the hairpin results in a 3'-phosphate at the base of the HP stem, while RNase P processing generates a 3'-hydroxyl group at the same position (Wiedenheft et al., 2011; Wilusz et al., 2008). A specific role for the 3’-phosphate group, if any, on Csy4-mediated rescue of non-polyadenylated RNAs is the subject of further exploration. Nevertheless, when taken together, these results affirm that both Csy4 binding and 3' end processing at the base of the HP stem are essential for stabilizing mRNA.

Implications for studying and manipulating mRNA.

Our studies suggest that Csy4 is a robust tool for knockdown of transgene-derived mRNA. Potential applications for this versatile system include (i) post-transcriptional control of transgenes, e.g., the development of ‘safety’ switches that can turn off gene expression (Di Stasi et al., 2011;Ketzer et al., 2014); (ii) spatio-temporal knockdown of over-expressed transgenes in animal models, where RNAi-mediated approaches to knockdown endogenous transgenes might not be effective; and (iii) multiplex processing of noncoding RNAs such as microRNAs without risking saturation of endogenous nucleus-to-cytosol transport machinery (Grimm et al., 2006). Another intriguing possibility is the potential to utilize different combinations of Class 1 Type I CRISPR endoribonucleases such as Cas6, Cas5d, Cas6e, and their corresponding substrates in multiplexed RNA processing applications (Brouns et al., 2008; Carte et al., 2008; Nam et al., 2012). In addition to providing tools for controlling transgene

expression, the current studies might also help expand the application of CRISPR-based tools for understanding RNA processing.

Potential cellular processing events for 3'UTR constructs containing the Csy4 HP substrate and lacking a poly(A) signal (3'UTR-HP-Δp(A)) are shown (Figure 9). First, within the nucleus, poly(A)-deficient constructs are likely degraded by cellular exonucleases in the absence of Csy4 (steps (i) and (ii)). However, Csy4 binding (iii) followed by cleavage of the substrate (iv) appears to stabilize the mRNA and enable cytosolic transport as well as translation (v) by unknown mechanisms. It is well known that the poly(A) tail plays a multifunctional role in protecting mRNA from degradation, transcriptional termination, nuclear-to-cytosolic transport and translation. Thus, the absence of a poly(A) tail is generally associated with rapid mRNA degradation and a lack (or exceedingly low levels) of transgene product. In the current study, we clearly demonstrate the ability of Csy4 to stabilize such poly(A) deleted mRNA constructs subsequent to binding and processing of the substrate hairpin at the 3' end. Secondly, cytoplasmic export of the Csy4-bound transcript from the nucleus and translation are likely essential for rescuing transgene expression. The mechanisms underlying these critical events are currently under investigation. Similar paradigms have been proposed in relation to 3'UTR processing for two nuclear noncoding RNAs – MALAT1 and MEN beta, which form a triple helical structure capable of stabilizing transgene-derived mRNA, supporting nuclear export and translation (Wilusz et al., 2012). Other examples of such tertiary RNA structures have also been reported in case of histone and viral-derived RNAs (Marzluff, 2012). For instance, Kaposi’s sarcoma associated herpesvirus (KSHV) is one such virus that employs an expression and nuclear retention element

(ENE) which also forms a triple helical structure (Mitton-Fry et al., 2010). Together, our studies and other examples in the literature strongly support the notion that the Csy4-hairpin complex can potentially function as a surrogate poly(A) tail. From a broader perspective, our studies expand the potential utility of CRISPR nucleases as tools for controlling mRNA stability and translation.

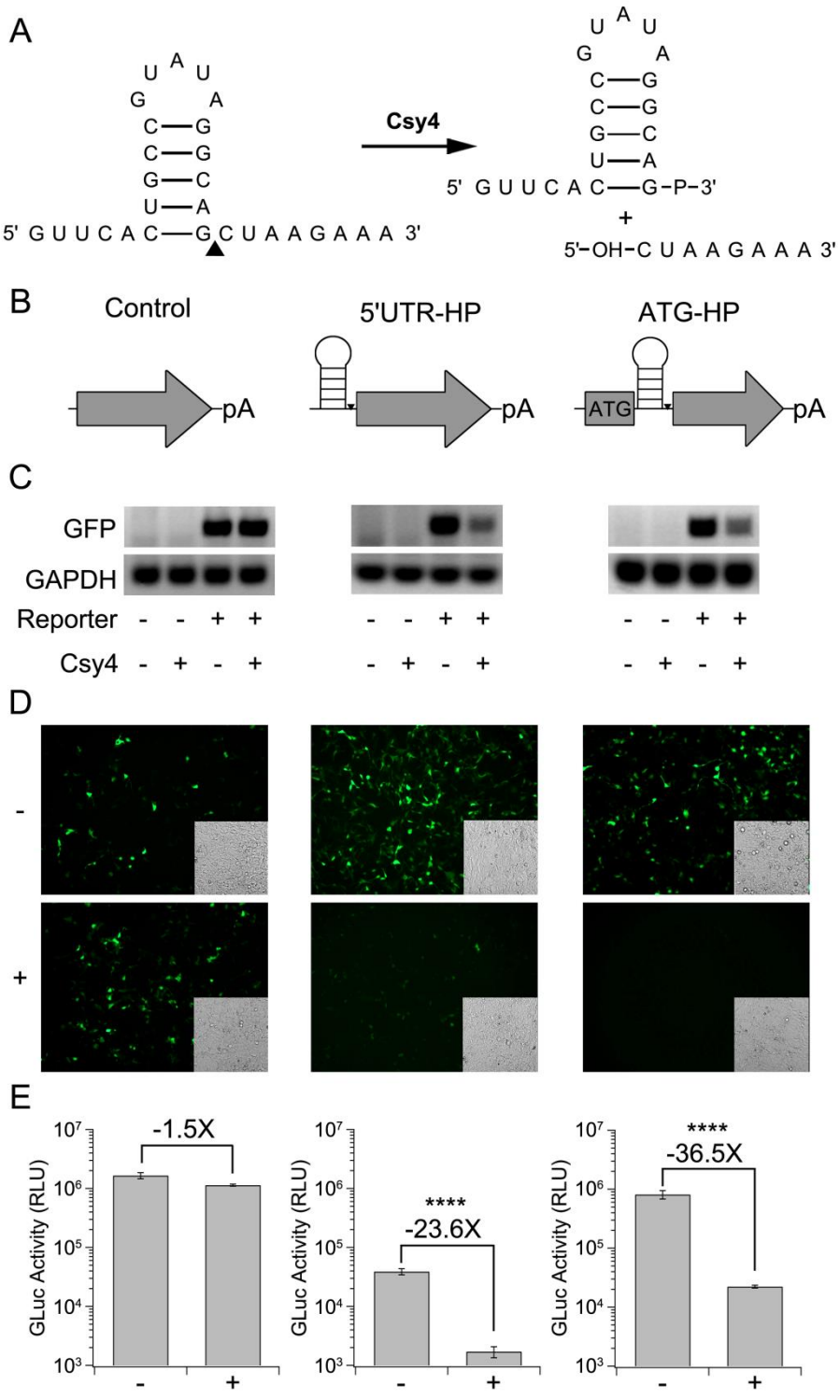

Figure 4: Csy4-mediated knockdown of 5' UTR-hairpin (HP) and ATG-HP constructs. (A) Sequence of the Csy4 hairpin substrate and fragments after enzymatic cleavage. The cleavage site is indicated by a black triangle on the unprocessed hairpin.

(B) Schematics of the unmodified control (left), 5'UTR-HP (middle), and ATG-HP (right) reporters. **(C)** PCR products of randomly primed RNA isolated from HEK293 cells transfected as indicated. (Left) unmodified control, (middle) 5'UTR-HP-GFP and (right) ATG-HP-GFP. **(D)** Fluorescence images of HEK293 cells expressing unmodified control (left), 5'UTR-HP (middle), and ATG-HP GFP (right) reporters in the absence (-) or presence (+) of Csy4. Corresponding transmitted light images are shown as insets in each fluorescence image. **(E)** Quantitation of GLuc activity by luminometric analysis at 24 hours post-transfection for unmodified control (left), 5'UTR-HP (middle), and ATG-HP (right), GLuc reporters. Error bars indicate standard deviation of four replicates. Statistical significance was calculated using a two-tailed Student's t-test (****$p \leq 0.0001$).

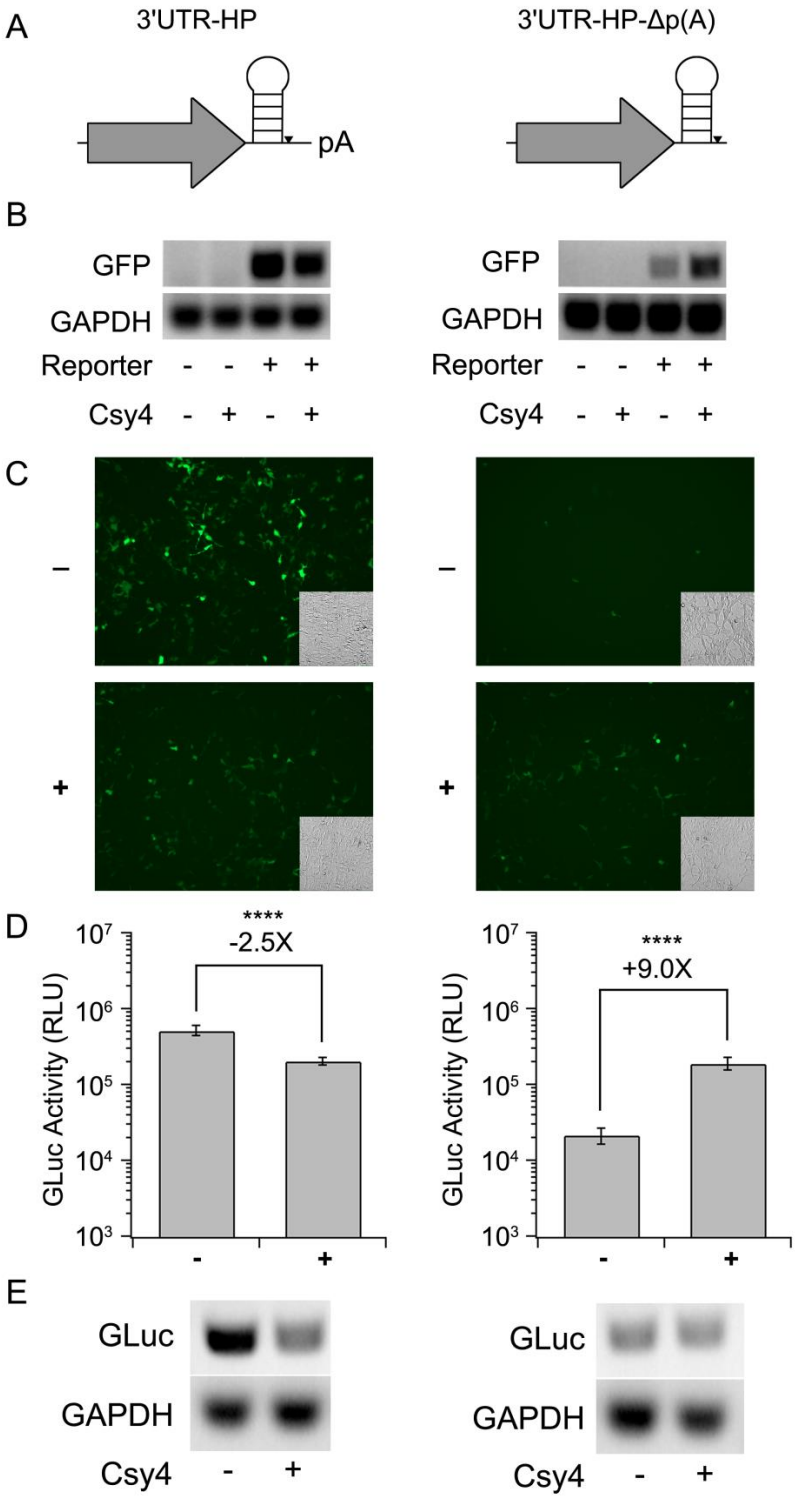

Figure 5: Effect of placing the HP in the 3'UTR on Csy4-mediated knockdown. (A) Schematics of the 3'UTR-HP reporter (left) and 3'UTR-HP-Δp(A) reporter (right). **(B)** PCR products of randomly primed RNA isolated from HEK293 cells transfected with the

GFP-3'UTR-HP reporter (left) or GFP-3'UTR-HP-Δp(A) reporter (right) and Csy4 as indicated. **(C)** Fluorescence images of HEK293 cells expressing GFP-3'UTR-HP (left) or GFP-3'UTR-HP-Δp(A) (right) in the absence (-) or presence (+) of Csy4. Corresponding transmitted light images are shown as insets in each fluorescence image. **(D)** Quantitation of Gluc activity by luminometric analysis at 24 hours post-transfection for GLuc-3'UTR-HP (left) or GLuc-3'UTR-HP-Δp(A) (right) reporters. Error bars indicate standard deviation of four replicates. Statistical significance was calculated using a two-tailed Student's t test (****$p \leq 0.0001$). **(E)** PCR products of oligo-$d(T)_{20}$ primed RNA isolated from HEK293 cells expressing GLuc-3'UTR-HP (left) or Gluc-3'UTR-HP-Δp(A) (right) with or without Csy4**.**

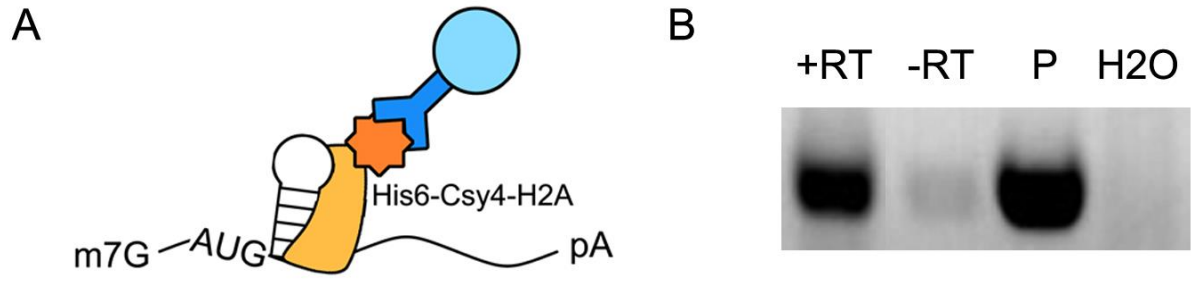

Figure 6: Cys4 interacts with the substrate hairpin in human cells. (A) Schematic representation of RNA immunoprecipitation (RNA-IP) detecting RNA associated with Csy4. An α-His6 antibody is used to pull down His6-tagged Csy4-H29A and associated transcripts. **(B)** GFP-primed PCR products of cDNA generated from RNA purified from α-His6 (Csy4-H29A) immunoprecipitates. Lysate from cells expressing ATG-HP-GFP and His6-Csy4-H29A served as RNA-IP input. Lanes are labelled as follows: reverse transcriptase (+RT), no reverse transcriptase (-RT), GFP plasmid positive PCR control (P), negative PCR control (H2O).

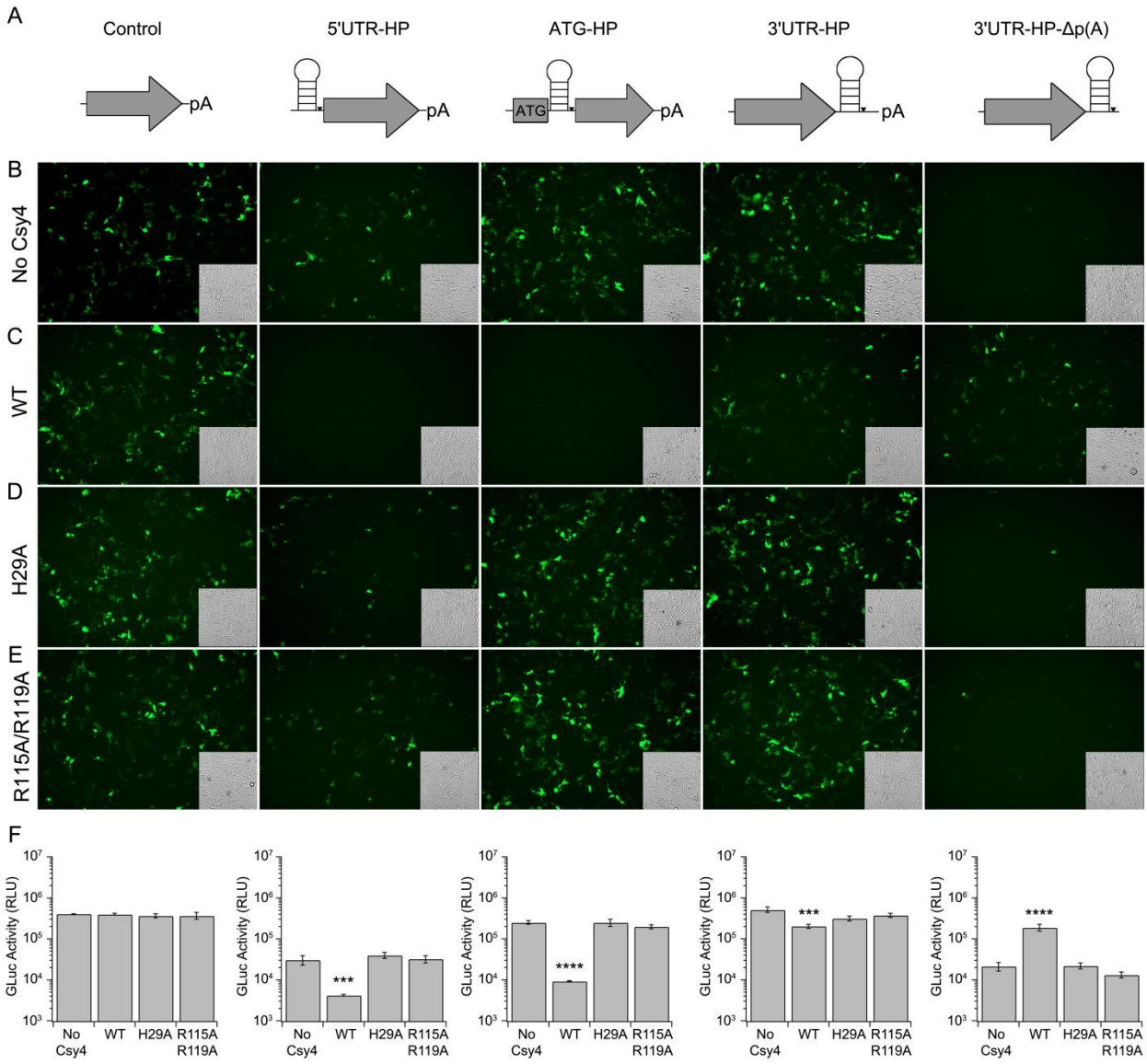

Figure 7: Csy4 binding and cleavage are necessary for regulation of transgene expression. (A) Schematics of reporters (left to right) - unmodified control, 5'UTR-HP, ATG-HP, 3'UTR-HP, and 3'UTR-HP-Δp(A). Fluorescence images of HEK293 cells expressing each reporter in the absence of Csy4 **(B)**, or presence of either native Csy4 **(C)**, Csy4 H29A **(D)** or Csy4 R115A/R119A **(E)**. Corresponding transmitted light images are shown as insets in each fluorescence image. **(F)** Quantitation of GLuc activity at 24 hours post-transfection by luminometric analysis for each reporter. Error bars indicate standard deviation of four replicates. Statistical significance was expressed relative to the "No Csy4" control and calculated using a two-tailed Student's t-test (****p≤0.0001, *** p≤0.001).

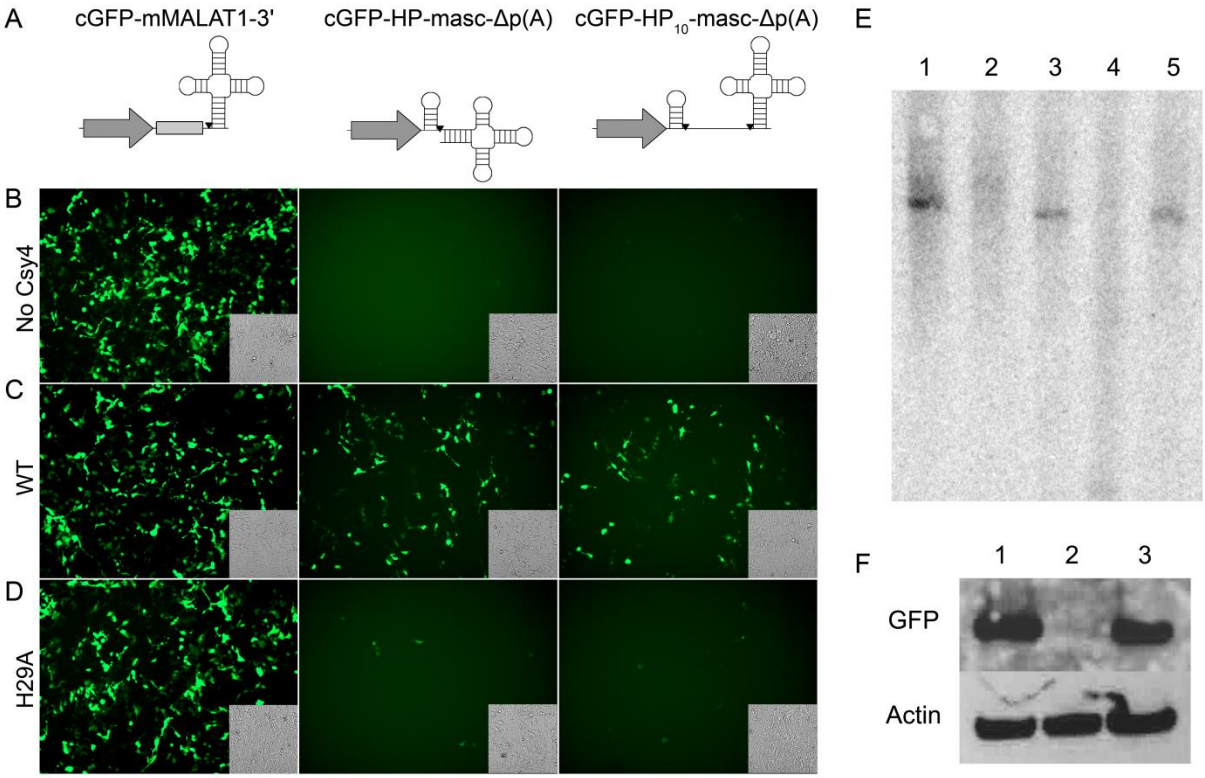

Figure 8: Selective processing by Csy4 is essential for rescue of 3'UTR-HP-Δp(A) constructs. **(A)** Schematics of the unmodified cGFP-mMALAT1-3' reporter (left), cGFP-HP-masc-Δp(A) (middle) and cGFP-HP_{10}-masc-Δp(A) (right) reporters. Cleavage sites are indicated by black inverted triangles. Fluorescence images of HEK293 cells expressing either cGFP-mMALAT3' (left), cGFP-HP-masc-Δp(A) (middle), or cGFP-HP_{10}-masc-Δp(A) (right) in the absence of Csy4 **(B)**, with Csy4 wild type **(C)**, or Csy4-H29A **(D)**. Corresponding transmitted light images are shown as insets in each fluorescence image. **(E)** Northern blot of cGFP based reporters in the presence or absence of Csy4. cGFP-mMALAT1-3' (lane 1), cGFP-HP-mascΔp(A) (lane 2), cGFP-HP-mascΔp(A) with Csy4 (lane 3), cGFP-HP_{10}-masc-Δp(A) (lane 4), and cGFP-HP_{10}-masc-Δp(A) with Csy4 (lane 5). **(F)** A Western blot probing for cGFP using cell lysates of HEK293s expressing either cGFP-mMALAT1-3' (lane 1), cGFP-HP-mascΔp(A) (lane 2), or cGFP-HP-mascΔp(A) with Csy4 (lane 3). A western blot for Actin is provided as a loading control.

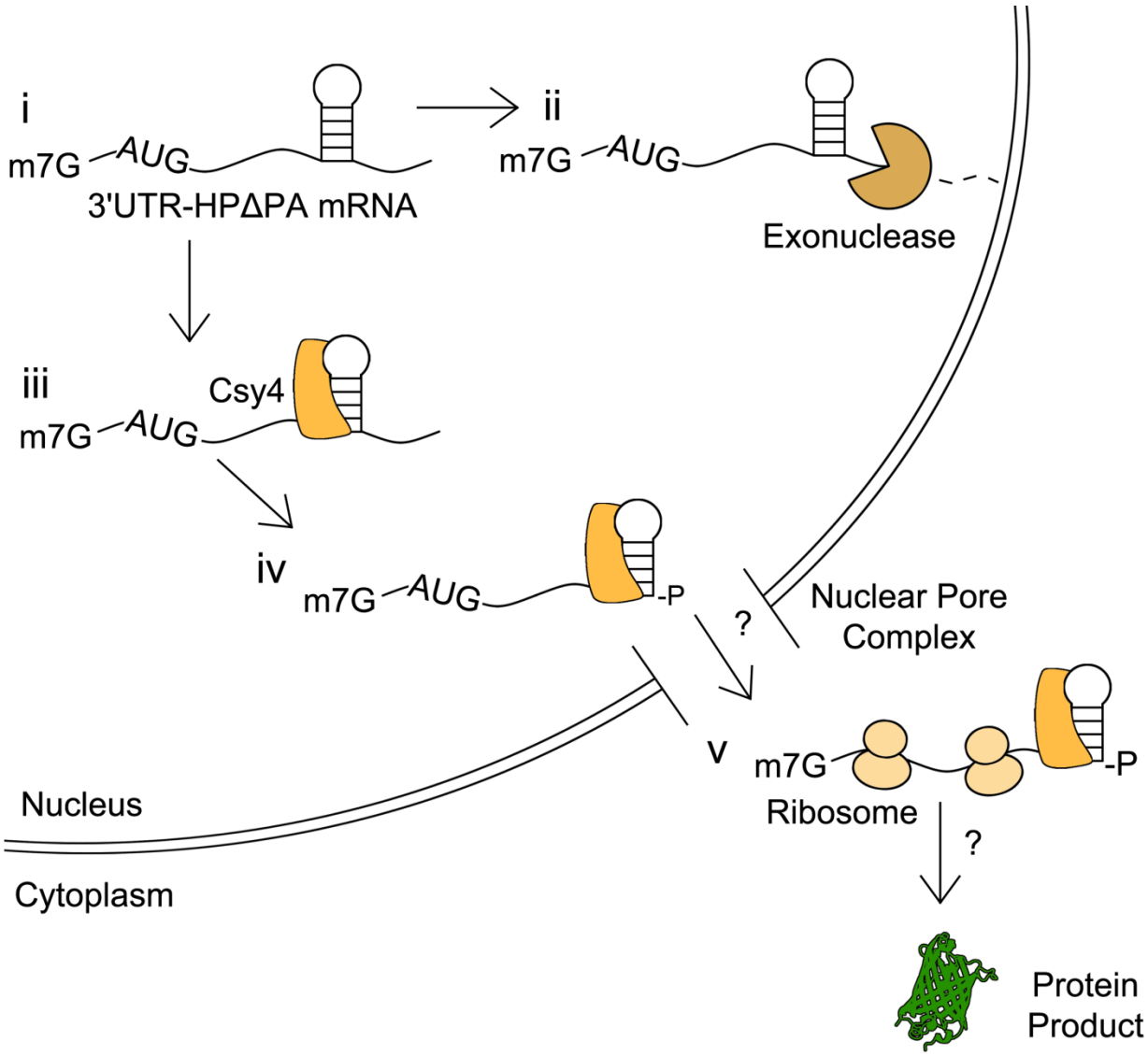

Figure 9: Potential events outlining Csy4-mediated processing of 3'UTR-HP-Δp(A) mRNA. First, within the nucleus, poly(A)-deficient constructs are likely degraded by cellular exonucleases in the absence of Csy4 (steps (i) and (ii)). However, Csy4 binding (iii) followed by cleavage of the substrate (iv) appears to stabilize the mRNA and enable cytosolic transport as well as translation (v) by unknown mechanisms.

CHAPTER 3: INDUCING CIRCULAR RNA FORMATION USING THE CRISPR ENDORIBONUCLEASE CSY4

3.1 Overview

Circular RNAs (circRNAs) are highly stable, covalently closed RNAs that are regulated in a spatiotemporal manner. With wide-ranging functions, these molecules have the potential to be incorporated into engineered systems with broad technological implications. Here we describe a switch for activating backsplicing of an engineered circRNA that relies on the CRISPR endoribonuclease, Csy4, as an activator of circularization. The endoribonuclease activity and 3' end-stabilizing property of Csy4 are particularly suited for this task. Co-expression of Csy4 and the circRNA switch allows for the removal of downstream competitive splice sites and stabilization of the 5' cleavage product. This subsequently results in backsplicing of the 5' cleavage product into a circRNA that can efficiently translate a reporter protein from an internal ribosome entry site (IRES) element. Our switch reports RNA isoform status with red fluorescence indicating linear RNA, and green fluorescence signaling circRNA formation. Circular status is confirmed by RNase R digestion. We observe steady accumulation of circRNA products over the course of 96 hours in response to Csy4 co-expression. Ultimately, our platform outlines a straightforward approach towards regulating splicing and could find potential applications in synthetic biology as well as in studying the properties of different circRNAs.

3.2 Introduction

Circular RNAs (circRNAs) comprise an emerging large class of noncoding, covalently closed RNAs present in a wide variety of organisms ranging from archaea to humans (Danan et al., 2012; Wang et al., 2014). They are typically generated through the covalent joining of a downstream splicing donor site with an upstream splicing acceptor site through a process called direct backsplicing (Jeck et al., 2013; Salzman et al., 2012). Due to their circular nature, circRNAs exhibit marked stability relative to their corresponding linear isoforms (Enuka et al., 2015; Jeck et al., 2013; Lasda and Parker, 2014; Liang and Wilusz, 2014). The median half-life of circRNAs is at least 2.5X longer than their linear counterparts (Enuka et al., 2015). Furthermore, for some genes, the abundance of the circRNA exceeds that of the associated linear mRNA by a factor of 10 (Jeck et al., 2013; Salzman et al., 2012).

Despite their prevalence, the exact function of a large number of known circRNAs remains to be determined (Ebbesen et al., 2015). Functions appear to vary from circRNA to circRNA, highlighting the potential range of applications for engineered circRNAs. Some circRNAs such as ciRS-7 (Cdr1as) and the *Sry* circRNA appear to act as microRNA sponges, while circMbl has been proposed as a sponge for MBL protein (Ashwal-Fluss et al., 2014; Hansen et al., 2013; Memczak et al., 2013). Though most circRNAs arise from protein-coding exons, endogenously encoded circRNAs do not appear to be associated with ribosomes or translated (Capel et al., 1993; Guo et al., 2014; Jeck et al., 2013). Engineered circRNAs can be translated, however, if an internal ribosome entry site (IRES) is included in the construct to drive translation (Chen and Sarnow, 1995; Wang and Wang, 2015). This property allows for expression of proteins

from stable, persistent RNA molecules. Accordingly, the development of molecular tools to exogenously control the circularization of a particular RNA of interest could prove useful.

In this regard, an interesting RNA targeting enzyme that has garnered recent interest from the synthetic biology community is Csy4 (Cas6f). Csy4 is an endoribonuclease belonging to the Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR) family. CRISPR systems utilize small RNAs (crRNAs) to guide CRISPR nucleases to nucleic acid, mediating sequence specific cleavage (Marraffini, 2015). Csy4 recognizes a 16 ribonucleotide hairpin which repeats throughout a precursor CRISPR RNA, cleaving at the 3' base of the hairpin stem to generate crRNAs (Cady and O'Toole, 2011; Haurwitz et al., 2010). Importantly, Csy4 remains associated with the 5' cleavage product (Sternberg et al., 2012). These properties of Csy4 have been adapted recently to process operon-encoding RNAs for gene expression without interference from cis sequences in *E. coli*, *B. subtilis*, and *S. cerevisiae* (Qi et al., 2012). Further, Csy4 was employed for processing RNA encoding multiple guide-RNA (gRNA) to single gRNAs for Cas9-based applications (Nissim et al., 2014; Tsai et al., 2014). Csy4 was also used to isolate and identify proteins associated with a specific RNA of interest (Lee et al., 2013). We recently demonstrated that Csy4 can be exploited to modulate the stability and subsequent protein expression of specific RNAs of interest (Borchardt et al., 2015). In particular, we demonstrated that Csy4 can stabilize the cognate 5’ cleavage product to allow translation of a reporter protein.

Here, we take advantage of the latter property of Csy4 to induce circularization of RNA. CircRNA biogenesis is measured using a Csy4-dependent fluorescent reporter

system (circGFP-CD). This reporter may serve as a switch-based platform for the regulation of a variety of natural and engineered circRNA-mediated processes while also providing a new tool for shifting expression from one gene to another. Additionally, the Csy4-inducible circularization system has the potential for application in the study of factors involved in circRNA biogenesis.

3.3 Materials and Methods

Plasmids: The Csy4 gene was amplified from the pseudomonas aeruginosa UCBP-PA14 genome using primers 5'-ATC GTC TAG AAT GGA CCA CTA CCT CGA CAT TCG CTT GC-3' and 5'-CGA TGC GGC CGC TCA GAA CCA GGG AAC GAA ACC TC CTT TGC-3'. The PCR product was cloned under the control of the CBA promoter in a plasmid backbone. Csy4-H29A was amplified from pHMGWA-Pa14Csy4H29A (Addgene plasmid #41092) which was a gift from Dr. Jennifer Doudna (Haurwitz et al., 2010). Csy4-H29A was cloned under the control of the CBA reporter. CircGFP was a gift from Dr. Zefeng Wang (Wang and Wang, 2015). To construct circGFP-CD, a gBlock ® (IDT DNA Technologies) was synthesized consisting of the Csy4 targeted hairpin followed by the canonical intron 12 of IGF2BP1 and the P2A sequence in-frame with a dsRed coding sequence. This gblock was then cloned downstream of the split-GFP cassette such that the entire transcript is driven by the CMV promoter and terminated with a SV40 poly-A signal.

Cell culture: HEK293 cells were cultured in Dulbecco's Modified Eagle's Medium (GIBCO/Life Technologies) supplemented with 10% FBS (Atlanta Biologicals), 1%

penicillin/streptomycin, and 2.5 µg/mL amphotericin B (Sigma-Aldrich) and maintained at 37ºC and 5% CO_2.

Fluorescence microscopy and quantification of fluorescence: 5 X 10^4 HEK293 cells were seeded overnight into 24 well plates and transfected with the indicated plasmids at equimolar quantities (500 ng). Cells were imaged at 72 hours post-transfection using an EVOS FL epifluorescence cell imaging system (AMC/Life Technologies) with the GFP light cube (excitation 470 nm, emission 510 nm), or RFP light cube (excitation 530, emission 590). Three images were analyzed per replicate using the FIJI image processing package to measure integrated density (Schindelin et al., 2012). Values for all nine images were averaged. Error bars indicate standard deviation of integrated density calculated from nine images from three biological replicates (three images per replicate). Statistical significance was calculated using a two-tailed Student's t-test ([****] P≤.0001, [*] P≤.05).

RT-PCR: 3 X 10^5 HEK293 cells were seeded overnight into wells of a 6-well plate and transfected with the indicated plasmids at equimolar quantities (totaling 2500 ng). RNA was harvested from these cells 48 hours post-transfection using the Total RNA Purification Kit (Norgen Biotek) and DNase treated using the Turbo DNA-free kit (Ambion/Life Technologies). Equal nanogram amounts of DNase treated RNA was converted to cDNA using the High Capacity RNA-to-cDNA kit (Applied Biosystems/Life Technologies). Products of this reverse transcription reaction were utilized as template for PCR amplification using gene-specific primers for GFP (5'-AGT GCT TCA GCC GCT

ACC C-3', 5'-GTT GTA CTC CAG CTT GTG CC-3') and glyceraldehyde 3-phosphate dehydrogenase (GAPDH) (5'-CCA CTC CTC CAC CTT TGA C-3', 5'-ACC CTG TTG CTG TAG CC-3'). PCR products were visualized on an agarose gel.

Western blot: HEK293 cells seeded overnight in 6 well plates at a density of 3 x 10^5 cells per well were transfected with a total of 2 µg DNA as indicated. Lysates were recovered 48 hours post-transfection using 1X Passive Lysis Buffer (Promega) and diluted 1:10. Samples were heated to 100C before separation on a 10% Bis-Tris gel. Membranes were blocked overnight in 2% milk in TBST. After overnight incubation, membranes were blotted with primary antibody against either GFP (1:1000 Santa Cruz, SC9996) or Actin (1:2000, Abcam, Ab3280). Stabilized peroxidase-conjugated goat anti-mouse antibody was used as secondary antibody (1:20,000, Jackson Immunologicals, 31430). Blots were developed using SuperSignal West Femto substrate (Thermo Scientific/Life Technologies).

RNaseR digestion: 2.2 X 10^6 HEK293 cells were seeded overnight into 10 cm plates and transfected with the indicated plasmids at equimolar quantities (totaling 6 micrograms). RNA was harvested at 72 hours post-transfection via TRIzol Reagent (Invitrogen/ThermoFisher Scientific) according to the manufacturer's protocol. 10 micrograms of RNA was treated with 10 units of RNaseR (Epicentre) at 37C for 10 minutes. Enzyme was inactivated at 95C for 5 minutes. The resulting RNA was subject to DNase treatment and RT-PCR as outlined above. Products were PCR amplified using primers detecting circular products (5'-CTG CTT GTC GGC CAT GAT ATA GAC

GTT GTG GC-3', 5'-CAA GCT GAC CCT GAA GTT CAT CTG CAC CAC C-3') or linear products (5'-CTT GGT CAC CTT CAG CTT GGC GGT CTG -3', 5'-GCT ACG TCC AGG GAT CCG GCG-3').

qRT-PCR: 3 X 10^5 HEK293 cells were seeded overnight into wells of a six-well plate. Before transfection, control RNA was harvested from cells using the Total RNA Purification Kit (#17200, Norgen Biotek). Remaining cells in separate wells were transfected with equimolar amounts (totaling 6 µg) of the indicated plasmids using PEI Max. RNA from each condition was harvested at time intervals of 8, 24, 48, 72 and 96 hours post-transfection using the Total RNA Purification Kit (Norgen Biotek) and treated with DNase using the Turbo DNA-free kit (Ambion/Life Technologies). 250 ng of DNase treated RNA was converted to cDNA using the High Capacity RNA-to-cDNA kit (Applied Biosystems/Life Technologies). Products of this reaction were diluted 1:100 and used as template for qPCR. qPCR was carried out using the Faststart Essential DNA Green Master Mix (Roche) and primers for detecting GFP (5'-AGT GCT TCA GCC GCT ACC C-3', 5'-GTT GTA CTC CAG CTT GTG CC-3'), dsRed (5'-GCT ACG TCC AGG GAT CCG GCG-3', 5'-CTT GGT CAC CTT CAG CTT GGC GGT CTG-3') or GAPDH (5'-CCA CTC CTC CAC CTT TGA C-3', 5'-ACC CTG TTG CTG TAG CC-3'). Data was analyzed using the ΔΔCt method and GFP and dsRed values are expressed relative to GAPDH values and normalized to t=0 control RNA. Error bars represent standard deviation of three biological replicates. Statistical significance was calculated using a two-tailed Student's t-test ([****] P≤.0001, [***] P≤.001, [**] P≤.01, [*] P≤.05).

3.4 Results and Discussion

To visualize the switch from linear to circular RNA products, we generated a dual-fluorescence reporter, circGFP-CD (Figure 10). The circGFP-CD reporter expresses dsRed upon forward splicing (linear RNA) and GFP upon backsplicing (circRNA). The reporter design was based on an circGFP construct reported earlier (Wang and Wang, 2015). Briefly, the original circGFP cassette consists of a split GFP in which the N terminal fragment of GFP lies downstream of the C terminal fragment. The C terminal fragment is preceded by a canonical splice acceptor sequence while the N terminal fragment is immediately followed by a canonical splicing donor sequence. The two fragments are separated by an IRES sequence such that, upon backsplicing, full length GFP is reconstituted and can be expressed as protein. Further, the split-GFP cassette is flanked by intron sequences which have been engineered to base pair in order to facilitate circularization (Wang and Wang, 2015). In the current study, we added an intron and competing splice acceptor site downstream from the split-GFP cassette. The splice acceptor is followed by a P2A-dsRed cassette, where the P2A peptide signal disrupts translation, while simultaneously allowing dsRed expression without being directly fused to GFP (Szymczak and Vignali, 2005). To toggle from linear RNA to circRNA, the split GFP and P2A-dsRed cassettes were separated by a Csy4 substrate hairpin (circGFP-CD; Figure 10A). Expression of circGFP-CD in the absence of Csy4 should favor forward splicing and result in dsRed expression (Figure 10A). Conversely, when circGFP-CD is co-expressed with Csy4, we expect the competing splice acceptor and P2A-dsRed encoding sequence to be cleaved away, leaving behind the 5' splice acceptor site (Figure 10B). We then hypothesized that the stabilization of the 5'

cleavage product by Csy4 could permit the cleaved split-GFP RNA to persist long enough for back-splicing to occur and support GFP expression (Figure 10B).

We first tested the circGFP-CD construct following plasmid transfection in mammalian cells and observed reporter expression via fluorescence microscopy. CircGFP, which is constitutively backspliced for GFP expression, is included as a control (Figure 11). When circGFP-CD is expressed alone, forward splicing appears to be favored as the cells strongly express dsRed but lack GFP fluorescence (Figure 11A). This expression pattern is consistent with predominant forward splicing and minimal backsplicing. It is important to note that the forward splice acceptor is identical to the backsplice acceptor in sequence. Therefore, the abundance of dsRed signal indicates that forward splicing is favored over backsplicing in this intron/exon configuration and host cell type. In contrast, when Csy4 is co-expressed with circGFP-CD, dsRed signal decreases while pronounced GFP signal appears, consistent with activation of circRNA biogenesis (Figure 11A). Fluorescence was quantified using FIJI image processing software (Schindelin et al., 2012) and the decrease in dsRed signal upon co-expression with Csy4 was measured at -6 fold. The corresponding increase in GFP signal was measured at +116 fold relative to Csy4 absent conditions (Figure 11B, 11C). The appearance of GFP signal concurrent with diminishing dsRed signal is consistent with Csy4 cleaving away the P2A-dsRed element and stabilizing the 5’ cleavage product. This allows back-splicing and subsequent GFP expression to occur.

To confirm that RNA cleavage and removal of the forward splicing acceptor is essential for favoring back-splicing, we co-expressed circGFP-CD with a catalytically dead Csy4 mutant, Csy4-H29A (H29A). This mutant is capable of binding to the target

hairpin, but is unable to cleave the RNA (Haurwitz et al., 2010). In the presence of H29A, GFP signal did not increase to the levels observed with native Csy4 (Figure 11A). This is consistent with Csy4 cleavage being required for predominant backsplicing to occur. However, when quantified, GFP signal did increase slightly by 1.8 fold (Figure 11C). Interestingly, H29A co-expression also induced a 2.4 fold increase in dsRed signal relative to Csy4 absent conditions (Figures 11A, 11B).

We then characterized the different RNA species generated from circGFP-CD using PCR and RNase R treatment. Two PCR primer sets were designed for the detection of either circular or linear products. The primers for circRNA detection will only result in product when the GFP fragments have been spliced into the appropriate orientation, for example, via RNA circularization. In the absence of Csy4, a faint band was detected using circular detection primers, indicating that back-splicing from circGFP-CD RNA occurs at a low basal rate (Figure 12A). However, when circGFP-CD RNA is co-expressed with Csy4, a robust signal is detected, demonstrating an increase in circRNA levels (Figure 12A). A modest increase in circRNA levels was also observed upon co-expression with H29A (Figure 12A). This increased detection of circRNA is consistent with fluorescence data shown earlier (Figure 11C). Together, these results demonstrate a slight improvement in overall splicing in the presence of Csy4-H29A. This effect may result from H29A-mediated RNA stabilization, although further studies will need to be conducted to fully elucidate the potential mechanism. Still, western blot analysis demonstrated that detectable amounts of GFP protein were only observed when circGFP-CD RNA was co-expressed with native Csy4 (Figure 12B). These results

corroborate the requirement of Csy4 for circularization and consequently GFP expression in our reporter system.

We further confirmed circRNA production by RNase R digestion. RNase R is a 3'- to 5'- exonuclease to which circRNAs are inherently resistant due to their lack of a 3' end (Suzuki et al., 2006). RNA derived from circGFP, which is constitutively back-spliced into circRNA, is included as a control and yields an RNase R resistant product (Figure 12C). When circGFP-CD RNA was expressed in the absence of Csy4, RNA was primarily detected using primers against linear template and was susceptible to RNase R (Figure 12C). This is consistent with predominant forward splicing in the absence of Csy4. Detection of RNase R susceptible linear product decreased when RNA from cells co-expressing Csy4 with circGFP-CD was analyzed (Figure 12C). Under these conditions, an RNase R resistant product was observed using primers designed for circRNA detection (Figure 12C). These results are consistent with Csy4-activated circRNA production. Taken together, these data indicate that the predominant splicing product of circGFP-CD switches from linear to circular when co-expressed with Csy4.

To observe the kinetics of circRNA formation from the circGFP-CD construct, we measured RNA levels over time by qRT-PCR. CircRNA and forward splicing products were differentiated by primers designed against circular and linear splicing products as depicted in Figure 13. Consistent with earlier studies, we observed circGFP control circular RNA levels rising over the course of 72 hours (Wang and Wang, 2015) (Figure 13B). Accumulation of circRNA from cells co-transfected with circGFP-CD and Csy4 also increased over time in a manner similar to circGFP (Figure 13B). This observation suggests that modulating circRNA formation with Csy4 might not affect efficiency of

circRNA synthesis. Importantly, in the absence of Csy4, circRNA levels remained low when expressed from the circGFP-CD cassette (Figure 13B). We next performed qRT-PCR with primers detecting forward splicing products (Figure 13A). Forward splicing products were expressed at high levels in case of the circGFP-CD without Csy4 (Figure 13A). Consistent with imaging and RNaseR studies, circGFP-CD co-transfected with Csy4 also expressed low levels of forward splice products. However, expression of these products remained consistently low relative to circGFP-CD transfected alone (Figure 13A). These results may indicate that a small proportion of RNA expressed from circGFP-CD is forward spliced before Csy4 cleaves away the forward splice acceptor.

In conclusion, we have designed a CRISPR-based, inducible system for circularization of RNA in mammalian cells. Csy4 acts as a reliable regulatory part for controlling the circGFP-CD switch, activating circRNA biogenesis, and protein expression. Likewise, CircGFP-CD is an adaptable device for the expression of a variety of proteins or functional non-coding RNA sequences for use in genetic circuits. A wide variety of RNA-based tools have been engineered for manipulating gene expression (Liang et al., 2011). For example, riboswitches and ribozymes are popular ligand responsive tools for controlling gene expression at the RNA level (Groher and Suess, 2014). Other tools for controlling RNA splicing include Engineered Splicing Factors (ESFs) (Wang et al., 2009). ESFs may serve as alternative options for regulating backsplicing efficiency. Here, we expand the RNA toolkit by developing a Csy4-dependent circRNA platform that provides a unique approach to regulate splicing, RNA stability, and protein translation. The functional utility of this inducible circRNA platform could potentially be expanded by replacing the fluorescent reporter ORFs with

other functional components. Such components include miRNA or protein sponge sequences as well as ORFs of proteins of interest. While the current system enables toggling from linear to circular RNA using Csy4, it might be possible to further optimize this system with the addition of cis-acting sequences and target sequences for trans-acting factors that have been shown to favor circRNA formation (Ashwal-Fluss et al., 2014; Conn et al., 2015; Ivanov et al., 2015; Jeck et al., 2013; Kramer et al., 2015; Liang and Wilusz, 2014; Zhang et al., 2014). As a circRNA expression tool, the addition of inverted repeat sequences known to facilitate circularization could improve the back-splicing efficiency of the reporter circRNA. In addition to temporal regulation, it might be possible to regulate circRNA expression in a cell/tissue-specific manner by changing promoter elements. Ultimately, CRISPR-induced circRNA formation could serve as a versatile platform for a range of potential applications in synthetic biology.

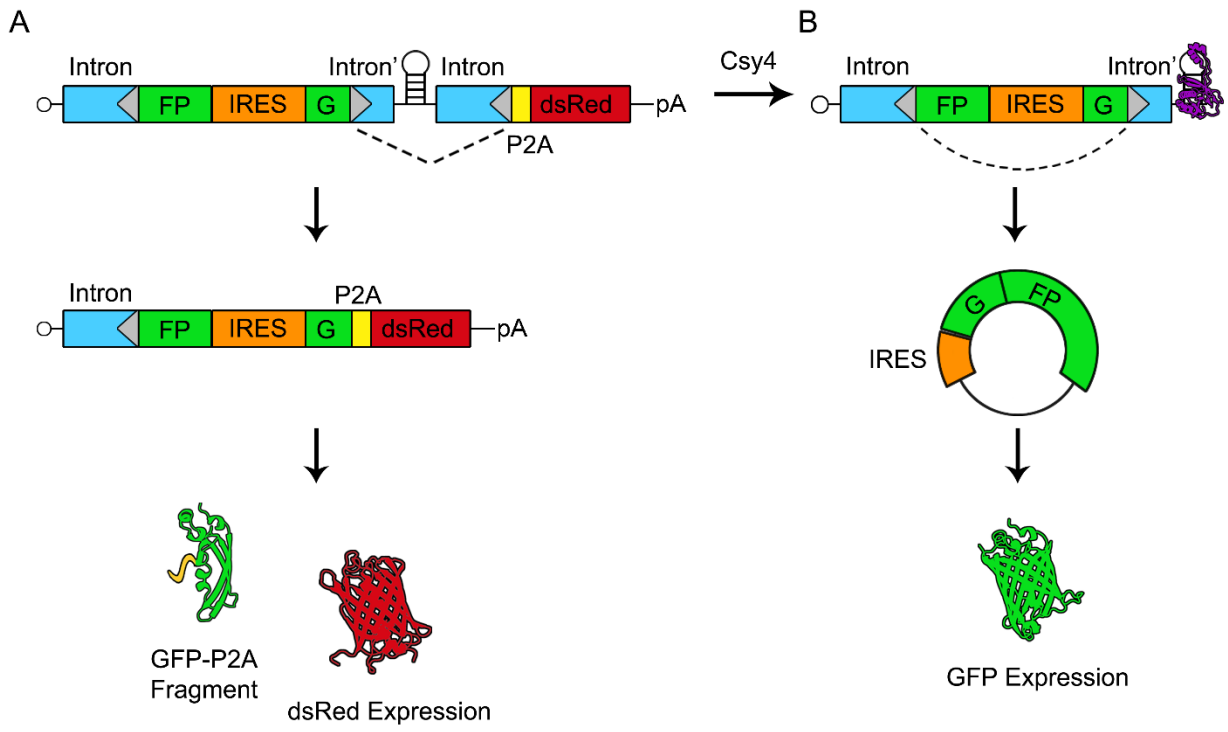

Figure 10: Schematic of circGFP-CD splicing patterns and reporter outputs. The mRNA expressed from circGFP-CD contains a split-GFP cassette flanked by engineered introns (intron and intron'). This cassette is followed by a Csy4 targeted hairpin, another intron and a P2A-dsRed cassette. Splice sites are represented by grey triangles and the dotted line indicates the expected predominant splicing pattern. In the absence of Csy4 **(A)**, forward splicing is favored and an IRES drives expression of dsRed. A non-fluorescing GFP fragment by-product is released via the P2A sequence. In the presence of Csy4 **(B)**, Csy4 cleaves the RNA at the base of the hairpin stem, releasing the forward splice acceptor and P2A-dsRed cassette. Csy4 remains associated with the cleaved hairpin, stabilizing the RNA and allowing back-splicing to occur. The IRES subsequently drives full length GFP expression from the circular RNA.

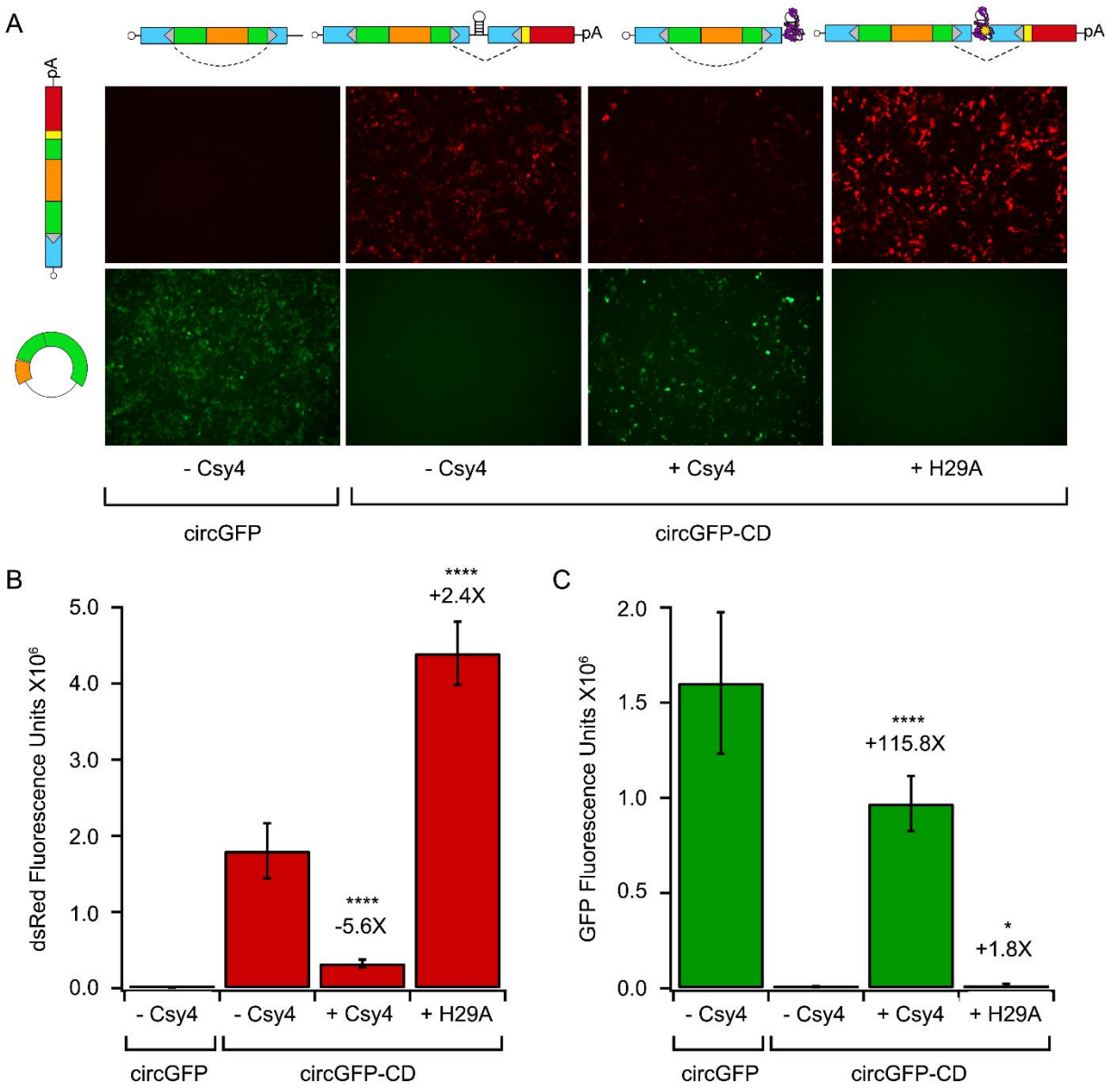

Figure 11: Csy4 induces circGFP-CD reporter expression to switch from dsRed to GFP. (A) Representative images of dsRed (top) and GFP (bottom) fluorescence from HEK293 cells expressing circGFP control (far left), or circGFP-CD in the absence of Csy4 (middle left), presence of native Csy4 (middle right) or presence of Csy4-H29A (far right) 72 hours post-transfection. Schematics of expected RNA species pre-splicing are depicted above each image column for each condition while schematics of expected RNA species post-splicing are depicted on the left of each row. Quantification of dsRed

(B) and GFP **(C)** fluorescence from images taken 72 hours post-transfection of HEK293 cells using the FIJI image processing package (Schindelin et al., 2012). Error bars indicate standard deviation of integrated density calculated from nine images from three biological replicates (three images per replicate). Statistical significance was expressed relative to the circGFP-CD “-Csy4” control and calculated using a two-tailed Student’s t-test ([****] P≤.0001, [*] P≤.05).

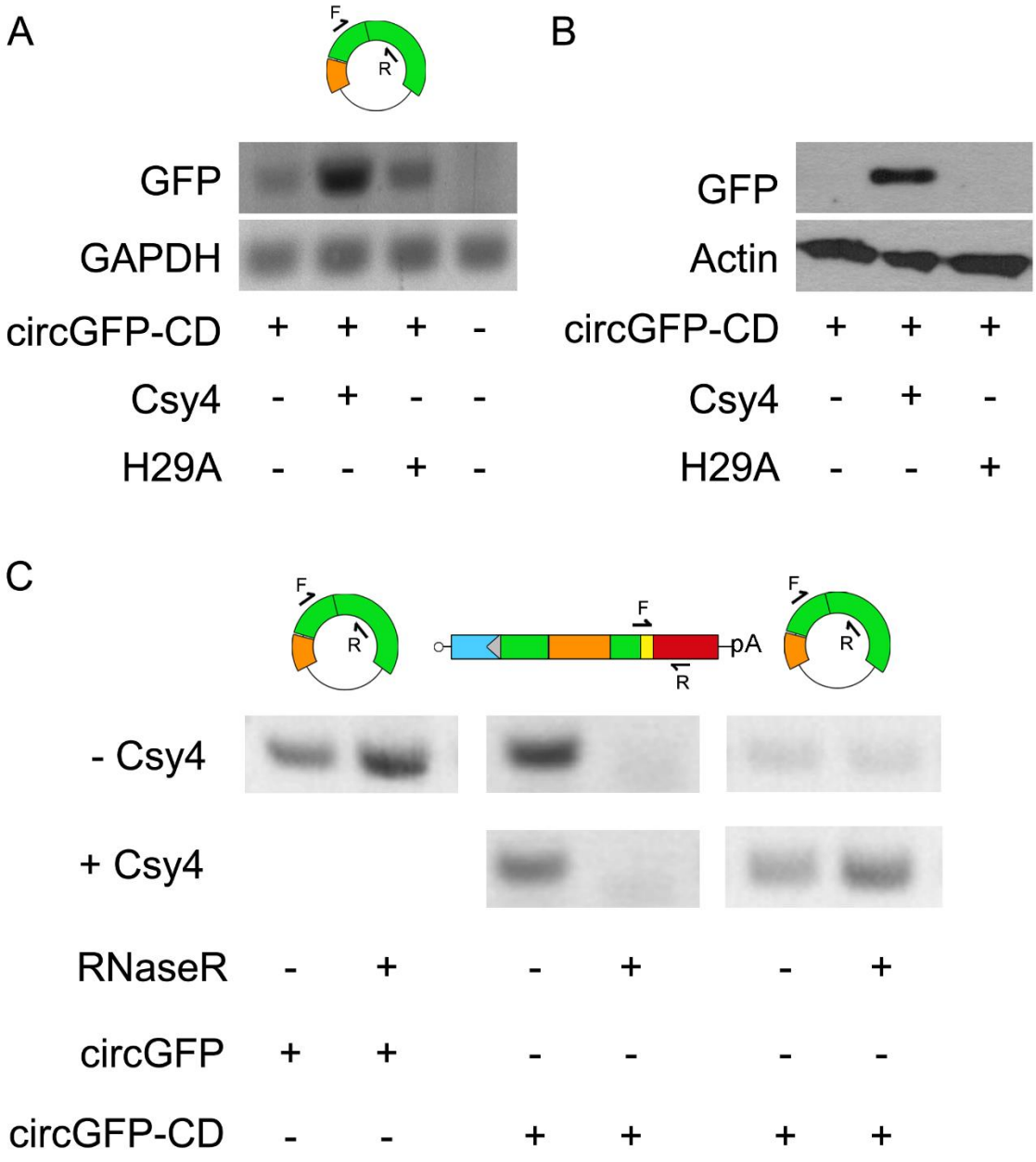

Figure 12: Detection of circGFP-CD expression products. Primers designed against either circular RNA or linear RNA products are depicted in each panel. **(A)** PCR products obtained using cDNA template generated from randomly primed RNA that was isolated from HEK293 cells transfected as indicated. GAPDH PCR products were amplified as loading controls from the same templates. **(B)** Western blot detecting GFP or loading control Actin for lysates of HEK293 cells transfected as indicated. **(C)** PCR

products amplified using primers specific for linear or circular splicing patterns. Template cDNA was generated using RNA isolated from HEK293 cells transfected with either circGFP, circGFP-CD, or circGFP-CD and Csy4 and treated with RNaseR as indicated.

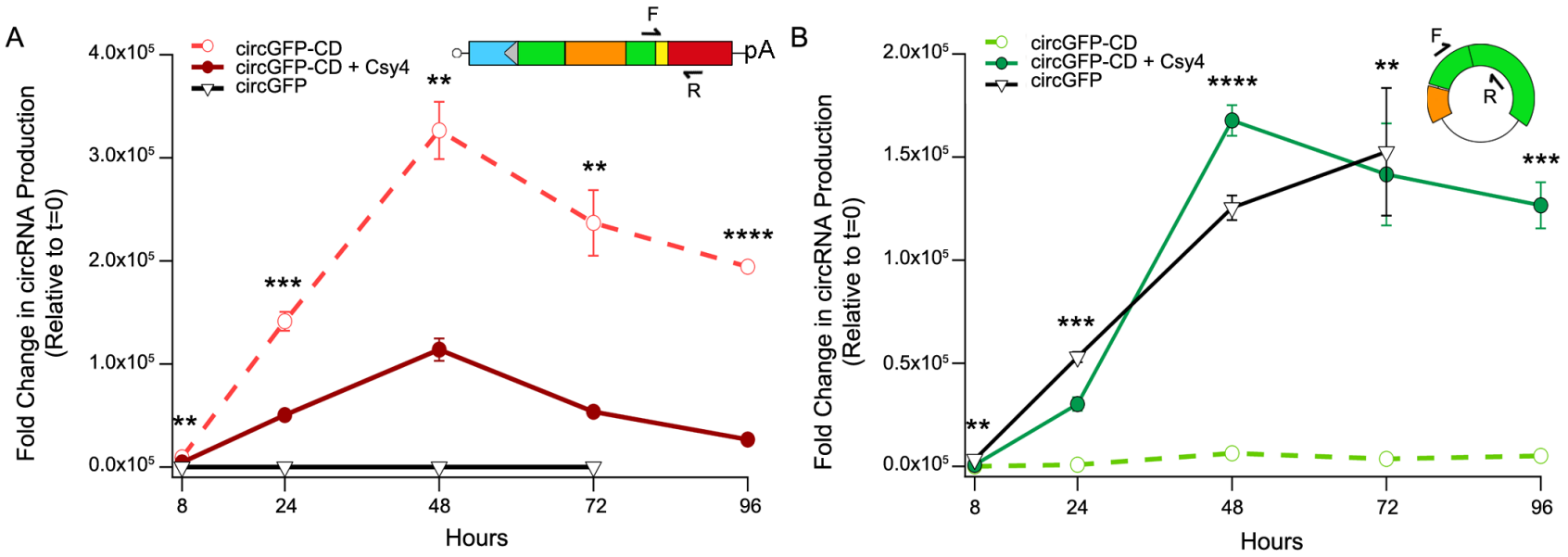

Figure 13: Accumulation of circRNA and linear splicing products over time. qRT-PCR of cDNA generated from RNA isolated at 8, 24, 48, 72, and 96 hours normalized to glyceraldehyde 3-phosphate dehydrogenase (GAPDH) and 0 hour transfection controls. Cells were transfected with either circGFP-CD (open circles, dotted line), circGFP-CD and Csy4 (closed circles, solid line), or with a circGFP control plasmid (open triangles, solid line). qRT-PCR was carried out using primers specific for either linear **(A)** or circular **(B)** splicing products as depicted above each graph. Error bars indicate standard deviation of three biological replicates. Statistical significance for the “circGFP-CD + Csy4” condition was expressed relative to “circGFP-CD” and was calculated using a two-tailed Student’s t-test ([****] P≤.0001, [***] P≤.001, [**] P≤.01, [*] P≤.05).

CHAPTER 4: *IN VIVO* DELIVERY OF TRANSLATABLE CIRCULAR RNA CASSETTES USING RECOMBINANT AAV VECTORS

4.1 Overview

Circular RNAs (circRNAs) are highly stable, persistent RNAs which serve a wide variety of cellular roles. With broad utility and potential for long-lasting effects due to their stability, circRNAs possess great promise as therapeutic agents. As currently understood, the natural functions of circRNAs are non-coding. However, circRNAs can be engineered for protein expression by the inclusion of an Internal Ribosome Entry Site (IRES) to drive translation. Recombinant Adeno-Associated Virus (rAAV) is commonly utilized in the field of gene therapy for the delivery of therapeutic cassettes. rAAV is an ideal candidate for transgene delivery due to well-defined tropism and minimal genome packaging requirements. Still, dose-related toxicity requires careful consideration when utilizing rAAV for therapeutic applications. Means to lower vector dosage while expressing gene products at therapeutic levels would alleviate some concern. With pronounced stability and long-term persistence, circRNAs are poised for maintaining long-term gene expression. This quality could prove essential for lowering effective vector dosage. Here we demonstrate expression of a circRNA reporter in mice through rAAV-mediated gene delivery. Unlike naturally occurring circRNAs, our reporter circRNA expresses a protein (GFP) through the inclusion of an IRES sequence. The GFP coding sequence is split such that full length GFP is only expressed upon RNA circularization. When administered via rAAV vectors, we observed GFP expression in the mouse heart

and liver. CircRNA biogenesis was confirmed via RNase R digestion. These results provide the foundation for AAV-mediated delivery of circRNA cassettes for a variety of therapeutic applications.

4.2 Introduction

Circular RNAs (circRNAs) comprise a class of covalently closed, exceedingly stable non-coding RNAs (ncRNAs). Initially overlooked as splicing byproducts, circRNAs are now recognized as highly abundant molecules with unique and complex expression patterns (Memczak et al., 2013; Salzman et al., 2012; Salzman et al., 2013) CircRNAs are found throughout the tree of life, with representation in organisms ranging from archaea, yeast, and amoeba, to more complex systems such as *Caenorhabditis elegans, Drosophila melanogaster,* and humans (Danan et al., 2012; Memczak et al., 2013; Salzman et al., 2013; Wang et al., 2014). The persistence of circRNAs throughout this wide range of organisms highlights the importance of these RNAs. It additionally supports further investigation into their biogenesis, function, and potential applications.

CircRNA biogenesis is thought to utilize a mechanism called backsplicing in order to join a downstream splicing donor to an upstream splicing acceptor. (Jeck et al., 2013). However, an alternate mechanism which employs a lariat intermediate has also been proposed (Zaphiropoulos, 1996).

CircRNAs are inherently stable and persistent, likely as a consequence of reduced exonucleolytic degradation resulting from their lack of exposed ends. The stability of circRNAs is demonstrated by a median half-life at least 2.5X longer than that of their linear counterparts (Enuka et al., 2015). In some cases, circRNAs outnumber

their associated linear RNAs by a factor of 10 (Jeck et al., 2013; Lasda and Parker, 2014).

Though the abundance of circRNAs has been well established, the roles of most circRNAs remain undetermined. However, from the few circRNAs with well-characterized functions, it is clear that circRNAs possess wide regulatory capacity. The circRNA Cdr1as (ciRS-7) acts as a miRNA sponge by interacting with miR-7 through 63 conserved miR-7 binding sites (Hansen et al., 2013; Memczak et al., 2013). Similarly, circSry also acts as a microRNA (miRNA) sponge by binding to miR-138 (Hansen et al., 2013). CircHIPK3 is an example of a circRNA capable of binding multiple different miRNAs and interacts with nine miRNAs over 18 binding sites (Zheng et al., 2016). CircMBL, on the other hand, is an example of a circRNA that binds protein. In an autoregulatory fashion, circMBL interacts with the MBL (Muscleblind) protein (a product of the MBL linear RNA), through MBL binding sites. This in turn regulates circMBL formation, which relies on MBL protein (Ashwal-Fluss et al., 2014).

The miRNA and protein sponging activity of circRNAs could affect a number of downstream processes. In one example, circFoxo3 binds to the cell cycle regulators p21 (cyclin dependent kinase inhibitor 1) and CDK2 (cyclin dependent kinase 2) (Du et al., 2016b). This interaction sequesters CDK2 from interacting with additional cell cycle proteins cyclin A and cyclin E, effectively preventing cell cycle entry (Du et al., 2016b). Similarly, circHIPK3 interacts with growth suppressive miRNAs. By binding these miRNAs, circHIPK3 is able to promote cell proliferation (Zheng et al., 2016). Additional proposed functions for circRNAs include transcription regulation and pseudogene formation (Dong et al., 2016; Li et al., 2015).

It is clear from their extensive breadth of innate functions, that circRNAs may provide a basis for number of therapeutic applications. Beyond their natural roles, circRNAs can be engineered for expressing proteins, provided that an Internal Ribosome Entry Site (IRES) is included for initiating translation (Chen and Sarnow, 1995; Wang and Wang, 2015). Further, in the previous chapter, we demonstrated that circRNA formation can be regulated by the CRISPR protein Csy4. The stability and utility of circRNAs thus provides and attractive platform for therapeutic gene expression of both protein and non-coding effectors.

Adeno-Associated Virus (AAV) is a helper-dependent parvovirus which is commonly adapted for applications in gene therapy and gene delivery (Samulski and Muzyczka, 2014). A number of natural and engineered AAV serotypes have been thoroughly characterized with well-defined tropisms (Asokan et al., 2012; Asokan and Samulski, 2013; Murlidharan et al., 2014). Further, The AAV genome can easily be tailored for delivery of nearly any cassette of interest provided it meets two requirements. First, the cassette must be flanked by AAV inverted terminal repeat sequences (ITRs) which are necessary for AAV genome packaging. Second, the vector genome must fit within the AAV packaging capacity of 4.7 kb. In addition, AAV's lack of pathogenicity and inability to replicate independently make recombinant AAV (rAAV) vectors excellent tools for gene delivery.

In this study, we demonstrate rAAV-mediated delivery of circRNA cassettes in mice. Further, we show that circRNAs can be translated *in vivo* to produce protein. Our findings can ultimately be applied towards the therapeutic design and usage of circRNA cassettes as both non-coding RNAs and for protein expression.

4.3 Methods

Plasmids. A split-GFP plasmid was provided as a generous gift from Dr. Zefeng Wang. To generate circGFP, the entire CMV-splitGFP-poly(A) sequence was PCR amplified out of the split-GFP cassette. The PCR product was subsequently cloned into an AAV vector ITR backbone for viral packaging. IRES-GFP was generated by overlap-extension PCR to join the IRES sequence with full length GFP. Fragments with overlapping GFP sequence were PCR amplified using the circGFP plasmid as template. The fragments were gel purified and combined in equimolar quantities (40 ng each) in consecutive PCR reactions without primers (35 cycles) and with flanking primers (29 additional cycles). The PCR product was cloned into the circGFP AAV vector backbone directly following the CMV promoter and before the poly(A) signal.

Cell culture. HEK293 cells were cultured in Dulbecco's Modified Eagle's Medium (GIBCO/Life Technologies) supplemented with 10% FBS (Atlanta Biologicals), 1% penicillin/streptomycin, and 2.5 µg/mL amphotericin B (Sigma-Aldrich) and maintained at 37ºC and 5% CO_2.

Detection of the circGFP splicing junction. HEK293 cells were seeded at a density of $3x10^5$ per well in a 6 well plate 24 hour prior to transfection. One microgram of circGFP plasmid was transfected into cells using PEI Max. Purified RNA was isolated using the Total RNA Purification Kit (#17200, Norgen) 72 hours post-transfection. RNA was treated with DNase using the Turbo DNA-free kit (Ambion/Life Technologies). DNase treated RNA was converted to cDNA using the High Capacity RNA-to-cDNA kit (Applied

Biosystems/Life Technologies). cDNA product was used as template for PCR amplification of full length GFP using primers (5'-GTA CAG CTC GTC CAT GCC GAG AGT GAT CC-3', and 5'-AGC AAG GGC GAG GAG CTG TTC ACC-3'). The PCR product was visualized on an agarose gel, gel purified, and Sanger sequenced (Eton Biosciences).

Recombinant AAV vector production. Recombinant AAV vectors were generated by a triple-plasmid transfection protocol (Shen et al., 2013a). Briefly, plasmid containing ITR flanked expression cassettes, a pXR9 helper plasmid, and a pXX6-80 adenoviral helper plasmid were transfected into HEK293 cells. After 72 hours, media was harvested off of these cells and subjected to vector purification by iodixanol gradient centrifugation. Fractioned virus was concentrated using a Pierce Concentrator (MWCO 150K, Thermo Scientific). Virus was then titered by qPCR using a Roche Lightcycler 96 (Roche Applied Sciences) with primers specific for the AAV2-ITR (5′-AAC ATG CTA CGC AGA GAG GGA GTG G-3′; 5′-CAT GAG ACA AGG AAC CCC TAG TGA TGG AG-3′) (IDT Technologies).

Intraperitoneal injections. BALB/c were mice bred and maintained in accordance with NIH guidelines and as approved by the University of North Carolina (UNC) Institutional Animal Care and Use Committee (IACUC). Neonatal postnatal day 2 (P2) pups were anesthetized by hypothermia by placement on ice for two minutes prior to injection. Mice were injected intraperitoneally with a dose of 5.75 X 10^{10} vector-genome containing particles (volume of 25 uL) of rAAV9 packaging the indicated virus. After

injection, mice were revived under a heat lamp, rubbed in bedding, and placed back with their mother. 5.5, 10, and 16 weeks post-injection, heart and liver were harvested from each mouse, post-fixed in paraformaldehyde, sectioned, and immunostained as described below.

Tissue processing and 3,3'-Diaminobenzidine (DAB) immunohistological analysis. Mice were sacrificed at the indicated timepoints with an overdose of tribromoethanol (Avertin) (0.2 mL of 1.25% solution) and subsequent transcardial perfusion of 4% paraformaldehyde in PBS. Heart and liver were removed and post-fixed overnight in 4% paraformaldehyde and then sectioned into 50-micron sections using a Leica VT 1200S vibrating blade microtome (Leica Biosystems). To stain, sections were blocked in 1% Triton X 100 and 10% goat serum in PBS, followed by overnight incubation with primary monoclonal antibody against GFP at a dilution of 1:750 (Life Technologies, G10362). Following overnight incubation, sections were washed with PBS and staining continued using the Vectastain ABC kit (Rabbit IgG, Vector Laboratories) according to the manufacturer's protocol. Stained sections were mounted onto slides and then visualized via light microscopy using an Evos FL epifluorescence cell imaging system. Slide scans were obtained by the UNC Translational Pathology Laboratory. Slide scan images were acquired using a bright field Aperio ScanScope XT and analyzed using Aperio ImageScope software.

Western blot. 10 tissue sections for each condition (CMV-GFP, IRES-GFP, or circGFP) at the 10-week timepoint were disrupted in 1X Passive Lysis Buffer (Promega) using a

TissueLyser (Qiagen). Disrupted tissue was incubated at room temperature for 5 minutes, with occasional vortexing, to facilitate lysis. Debris was pelleted by centrifugation and supernatant transferred to a fresh tube. Protein content of lysates was measured using a Bradford assay and equal amounts of protein from three biological replicates were combined to a total of 0.5 micrograms. Samples were heated to 100ºC before separation on a 10% Bis-Tris gel. Membranes were blocked overnight in 2% milk in TBST. After overnight incubation, membranes were blotted with primary antibody against either GFP (1:1000 Santa Cruz, SC9996) or Actin (1:2000, Abcam, Ab3280). Stabilized peroxidase-conjugated goat anti-mouse antibody was used as secondary antibody (1:20,000, Jackson Immunologicals, 31430). Blots were developed using SuperSignal West Femto substrate (Thermo Scientific/Life Technologies).

RNA isolation from mouse tissue. Mice were sacrificed at 10 weeks post-injection with an overdose of tribromoethanol (Avertin) (0.2 mL of 1.25% solution) and subsequent transcardial perfusion of PBS. Heart was removed and stored in RNAlater stabilization solution (Ambion/Life Technologies) until RNA harvest or DNA harvest (see below). To isolate RNA, tissue was disrupted in TRIzol (Invitrogen) using a TissueLyser (Qiagen). Once homogenized, samples were processed using the manufacturer's protocol for TRIzol extraction of RNA from tissue.

RNase R digestion. Five micrograms of RNA were DNase treated using the Turbo DNA-free kit (Ambion/Life Technologies). One microgram of DNase treated RNA was next preheated to 55ºC for 5 minutes and treated with 1 unit of RNaseR (Epicentre) at

37ºC for 10 minutes. Enzyme was inactivated at 95ºC for 5 minutes. The resulting RNA was subject to RT-PCR as outlined below.

RT-PCR. Equal volume amounts of RNase R treated RNA was converted to cDNA using the High Capacity RNA-to-cDNA kit (Applied Biosystems/Life Technologies). Products of this reverse transcription reaction were utilized as template for PCR amplification using gene-specific primers for GFP (5'-CTG CTT GTC GGC CAT GAT ATA GAC GTT GTG GC-3', 5'- CAA GCT GAC CCT GAA GTT CAT CTG CAC CAC C-3') and glyceraldehyde 3-phosphate dehydrogenase (GAPDH) (5'-CCA CTC CTC CAC CTT TGA C-3', 5'-ACC CTG TTG CTG TAG CC-3'). PCR products were visualized on an agarose gel.

Vector genome quantification. DNA was extracted from 10-week timepoint tissue lysates using the DNeasy kit (Qiagen). Vector genome (vg) copies were quantified via qPCR using primers against the AAV ITR (5'-AAC ATG CTA CGC AGA GAG GGA GTG G-3', 5'-CAT GAG ACA AGG AAC CCC TAG TGA TGG AG-3'). Vg values were normalized to copies of the mouse Lamin B2 locus (primers 5'-GGA CCC AAG GAC TAC CTC AAG GG-3', 5'-AGG GCA CCT CCA TCT CGG AAA C-3') for vg/cell values.

4.4 Results and Discussion

Design of an rAAV circRNA reporter.

To visualize circRNA expression in mice, we cloned a split-GFP reporter into an rAAV vector backbone (circGFP) (Wang and Wang, 2015). In the circGFP reporter,

GFP is divided into two parts with the N terminal fragment placed downstream of the C terminal fragment (Figure 14A). The upstream fragment of GFP is preceded by a splicing acceptor sequence at its 5' end while the downstream fragment is followed by a splicing donor at its 3' end. The two GFP fragments are separated by an encephalomyocarditis virus (EMCV) IRES sequence. This ensures that, upon back-splicing, full length GFP can be translated (Figure 14). Base pairing between flanking introns has been shown to facilitate circRNA formation (Ashwal-Fluss et al., 2014; Ivanov et al., 2015; Jeck et al., 2013; Kramer et al., 2015; Liang and Wilusz, 2014; Zhang et al., 2014). To take advantage of this property, split-GFP is flanked by IGF2BP1 introns which have been engineered to base pair (Wang and Wang, 2015) (Figure 14C). The entire cassette is driven by a CMV promoter and terminated with an SV40 p(A) signal (Figure 14A). Reconstituted, full length GFP RNA can be isolated from cells expressing the circGFP plasmid and, when sequenced, shows an intact splicing junction (Figure 14D). This indicates reliable and consistent backsplicing to form an in-frame GFP open reading frame (ORF).

We also generated two control rAAV vectors, the first of which expresses a full length GFP ORF (CMV-GFP) (Figure 15A). The second control vector, meant to account for IRES driven translation, consists of an EMCV IRES driving full length GFP (IRES-GFP) (Figure 15A). All constructs are expressed using the CMV promoter and SV40 poly-A signal.

rAAV-mediated delivery of circGFP in mice.

rAAV packaging circGFP, CMV-GFP or IRES-GFP was administered via

intraperitoneal injection in postnatal day 2 (P2) BALB/c mouse pups. We chose to utilize mouse pups in order to observe the effects of rapid cell proliferation on expression of circRNA products. In more proliferative tissues, we expected dilution of both AAV episomes and circRNA over time. At 5.5, 10, and 16 weeks post-injection, GFP expression was detected by 3,3'-Diaminobenzidine (DAB) staining. Consistent with the robust transduction mediated by AAV9 throughout mouse tissue, GFP was discerned throughout the heart and liver from CMV-GFP at all timepoints (Pulicherla et al., 2011b) (Figures 15B, 15C). IRES-GFP treatment, however, resulted in minimal GFP detection (Figures 15B, 15C). Importantly, GFP was observed in the mouse heart out to 16 weeks post-injection and liver up to 10 weeks in circGFP treated tissue (Figures 15B, 15C). Full length GFP can only be expressed from circGFP upon RNA circularization. Therefore, GFP detection suggests that circRNAs were generated in these tissues.

Detection of GFP from both CMV-GFP and circGFP was more robust in cardiac tissue relative to the liver, and in the case of circGFP, persisted to later timepoints (Figures 15B, 15C). This may be attributed to the highly proliferative nature of the liver. rAAV genomes, and presumably circRNA products, are subject to dilution as they are distributed throughout each cell division. The observed expression differences between heart and liver could also result from differential expression from the EMCV IRES in these tissues (Creancier et al., 2000).

Molecular validation of circRNA production in mouse heart.

We chose to focus our molecular analysis on 10-week post-injection cardiac tissue. Slides of DAB-stained heart sections from each timepoint were scanned to

visualize global expression (Figure 16A). Expression of CMV-GFP was distributed evenly throughout the heart, while circGFP was detected at lower frequency (Figure 16A). While equal doses of rAAV vector were administered in each mouse, it is possible that injection precision was not consistent. To detect potential variation in vector delivery, vector genomes (vg) were quantified via qPCR. The largest difference in genome copy in the heart was detected between circGFP and CMV-GFP treated tissue (Figure 16B). However, the two-fold difference between these two conditions is relatively low and statistically insignificant ($p=0.294$, two-tailed Student's t-test). Vg copies detected from IRES-GFP treated cardiac tissue did not differ significantly from CMV-GFP or circGFP conditions (Figure 16B). Therefore, minimal detection of GFP from IRES-GFP vectors is likely not due to variation in vector delivery in the heart.

We performed RNase R digestion to confirm that circRNAs were, indeed, being produced in the heart of treated mice. While linear RNAs are susceptible to this 3' to 5' exonuclease, circRNAs are inherently resistant due to their lack of exposed ends (Suzuki et al., 2006). We isolated RNA from heart tissue of mice injected with rAAV CMV-GFP, IRES-GFP or circGFP and treated with RNase R. Products were converted to cDNA and amplified with either GAPDH or GFP specific primers. The RNase R negative conditions reflect relative expression of the RNA from each vector. Consistent with GFP staining, CMV-GFP produced the strongest GFP band while IRES-GFP and circGFP products were lower in intensity (Figure 16C). Importantly, the CMV-GFP and IRES-GFP products were both susceptible to RNase R while circGFP RNA remained RNase R resistant. This is consistent with circGFP forming a circular RNA (Figure 16C). As expected, linear GAPDH controls were susceptible to RNase R (Figure 16C).

Western blotting demonstrated that GFP is only detectable from CMV-GFP and circGFP tissues (Figure 16D). Further, though the CMV-GFP RNA is certainly more abundant relative to the circGFP RNA, GFP protein for both conditions is detected by western blot at similar levels (Figures 16C, 16D). This may suggest that the stability of circRNAs allows protein expression at comparable levels to that of linear mRNA, but with lower transcript abundance. Despite detection of the linear IRES-GFP RNA, GFP protein is not detectable from these tissues via western blot (Figures 16C, 16D). This highlights the importance of circularization for stable expression. Together, these observations support the hypothesis that the stability of circRNA constructs contributes to long term protein expression as compared to linear RNAs.

Future prospects and implications for gene therapy.

Here we demonstrated the delivery and expression of translatable circRNA cassettes in mice using rAAV vectors. In the heart, circGFP expressed GFP at levels nearing that of a control full length GFP reporter (CMV-GFP) out to 16 weeks post-injection (Figures 15C, 16D). Further, the GFP expression from circGFP was robust relative to that of a non-circularized control RNA (IRES-GFP) (Figures 15C, 16D). The difference between IRES-GFP and circGFP protein expression is likely a consequence of the stability of the circRNA over its linear IRES-GFP counterpart. These findings demonstrate the promise of circRNA-mediated protein expression in animal models.

While we demonstrate circRNA formation *in vivo*, the efficiency of circRNA biogenesis can likely be improved. One approach towards enhanced circularization could involve employing cis-acting sequences known to facilitate circRNA formation. For

example, Alu repeats have been identified in the introns flanking a number of exons known to circularize (Ashwal-Fluss et al., 2014; Ivanov et al., 2015; Jeck et al., 2013; Kramer et al., 2015; Liang and Wilusz, 2014; Zhang et al., 2014). The introns are thought to contribute to circRNA biogenesis by hybridizing through nucleotide-complementarity, bringing the two splice sites into closer proximity. While our circGFP vector makes use of a similar strategy through engineered introns, the use of naturally occurring intron sequences may improve circularization efficiency.

Many rAAV-based therapeutics utilize short hairpin RNAs (shRNA) or microRNAs (miRNAs) to regulate gene expression, and have largely proved effective (Borel et al., 2014). However, in some cases, overexpression of shRNAs yields hepatotoxicity or neurotoxicity (Borel et al., 2014; Grimm et al., 2006; van Gestel et al., 2014). This toxicity likely stems from saturation of the miRNA processing pathway and disruption of the normal cellular miRNA population (Grimm et al., 2006; Valdmanis et al., 2016; van Gestel et al., 2014). Similarly, it will be important to assess any potential toxicity resulting from circRNA vectors in mice. Investigating cellular splicing pattern variations upon circGFP overexpression could evaluate potential oversaturation of splicing machinery.

While many clinical trials utilizing rAAV vectors for therapeutic applications have proved promising, dose-related toxicity still must be considered. Therefore, it would be beneficial to reduce the effective vector dose for rAAV vector-based therapeutics. CircRNAs are uniquely poised to address these issues. With a wide breadth of potential applications ranging from miRNA sponging to transcription regulation, circRNAs could prove useful in a myriad of situations. Combined with the well-defined tropisms of rAAV

vectors, rAAV-circRNA cassettes could allow tissue specific, stable expression of therapeutic agents. Further, though it remains to be directly tested, the stability of circRNAs could provide a mechanism for long term expression of therapeutic cassettes at a lower effective dosage. Ultimately, rAAV-delivered circRNA cassettes present a stable platform for *in vivo* gene expression.

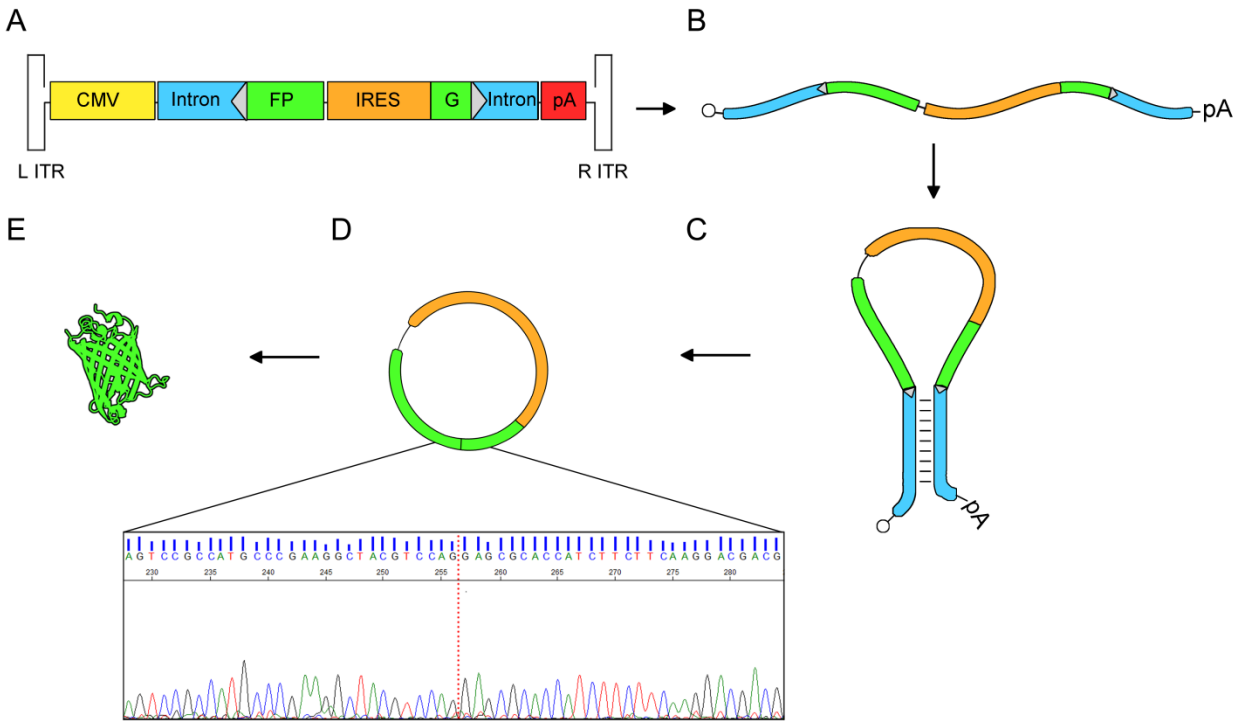

Figure 14: Schematic of circGFP expression. (A) Schematic of the circGFP rAAV vector genome encoding a split-GFP cassette. **(B)** An mRNA encoding the split-GFP cassette is expressed from the circGFP rAAV genome. **(C)** Circularization is facilitated by hybridization of engineered IGF2BP1 introns to bring splice sites (grey triangles) into closer proximity. **(D)** Circularization reconstitutes the full length GFP ORF. A chromatogram depicting sequencing results of a GFP PCR product is depicted below the circRNA. Template cDNA was generated from RNA isolated out of circGFP transfected HEK293 cells. The splicing junction is indicated by a dashed red line. **(E)** GFP protein is expressed from the reconstituted GFP ORF encoded by the circRNA.

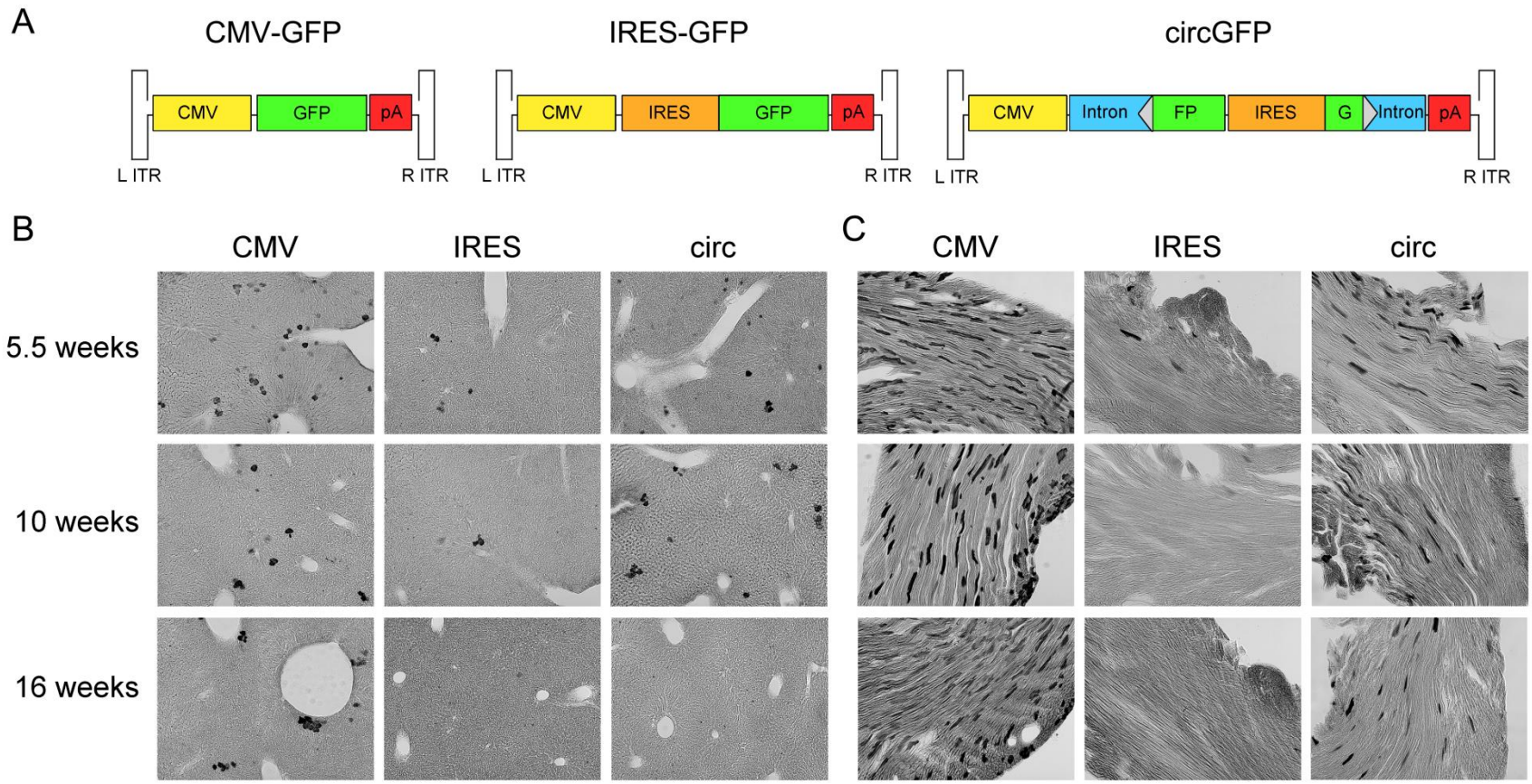

Figure 15: rAAV mediated delivery of circGFP expression cassette in mice. (A) Schematic representations of the rAAV vector genomes encoding each cassette used in this study. **(B, C)** 3,3'-Diaminobenzidine (DAB) stained heart **(B)** and liver **(C)** sections detecting GFP expression from mice expressing each rAAV-delivered cassette at 5.5, 10 and 16 weeks post-injection.

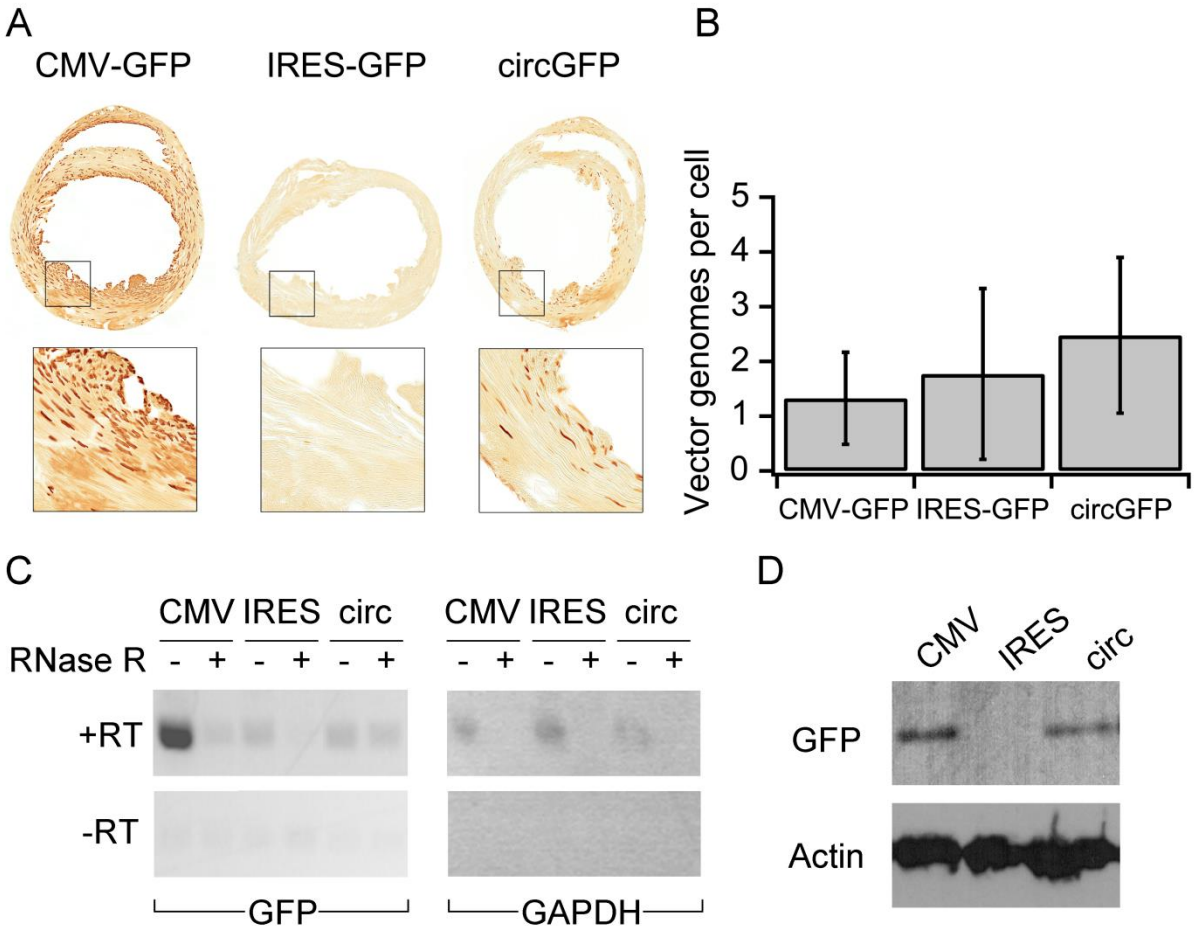

Figure 16: Molecular characterization of circGFP expression products in mice. (A) 3,3'-Diaminobenzidine (DAB) stained heart sections detecting GFP expression from mice expressing each rAAV-delivered cassette 10 weeks post-injection. Slide scans (top) of whole sections display widespread expression patterns while selected regions at higher magnification (bottom) reveal GFP expression by individual cells. **(B)** rAAV vector genomes per cell measured by qPCR in heart tissue for each rAAV cassette. Data are expressed as an average of three biological replicates. Statistical significance was calculated using a two-tailed Student's t-test. No significance was detected**. (C)** PCR products detecting either GFP (left panel) or GAPDH (right panel) from cDNA. cDNA was generated from RNA isolated from mouse heart. RNA was treated with RNase R as indicated prior to reverse transcription. Mice were treated with rAAV delivering either CMV-GFP, IRES-GFP, or circGFP and RNA was isolated 10 weeks

post injection for analysis. **(B)** Western blot detecting GFP or Actin loading control. Cell lysates were used for loading were generated from mouse heart treated with rAAV packaging either CMV-GFP, IRES-GFP, or circGFP. Heart was harvested 10 weeks post-injection.

CHAPTER 5: SYNOPSIS AND FUTURE DIRECTIONS

5.1 Summary

The field of gene therapy has successfully driven the development of effective therapeutics for diseases including Leber's congenital amaurosis and lipoprotein lipase deficiency. However, therapeutic gene expression requires careful regulation to protect against potential adverse effects. This dissertation focused on developing new approaches to regulate RNA stability for potential applications in gene therapy. To this end, I explored the use of the bacterial CRISPR endoribonuclease, Csy4, for manipulating RNA stability, and demonstrated the potential of rAAV-delivered circular RNA (circRNA) cassettes for protein expression *in vivo*. First, I established the functionality of Csy4 in human cells, as well as the ability of Csy4 to stabilize or destabilize an RNA based on target hairpin location. I found that Csy4 cleavage in the 3' UTR of a poly(A)-deficient RNA can support the stabilization and translation of that RNA. Further, I developed an inducible circRNA cassette which switches from a linear isoform to a circular isoform upon Csy4 co-expression. I demonstrated that Csy4 cleavage can remove competing downstream splicing signals to subsequently favor backsplicing of the RNA into a circRNA. Lastly, I delivered recombinant AAV (rAAV) encoding translatable circRNA cassettes in mice. I found that protein can be expressed from circRNA at a level nearing that of protein expressed from the linear isoform *in vivo*, despite the circRNA being present at lower relative abundance. Additionally, I found that

IRES-driven expression from these circRNA cassettes is more efficient *in vivo* compared to in cell culture.

CircRNAs are particularly persistent and stable, offering an ideal potential platform for long-term gene expression. However, the circRNA expression cassettes presented here will require further engineering to develop optimal candidates for gene therapy applications. Additionally, due to the bacterial origin of Csy4, it will be important to characterize Csy4 further in a eukaryotic context. It may also be worthwhile to investigate the functionality of other Csy4 orthologs in human cells. In this chapter, I discuss approaches for characterizing and improving the novel technologies described in this dissertation.

5.2 Further Evaluation and Regulation of Csy4 in Mammalian Cells

Chapters 2 and 3 of this dissertation investigated the use of the CRISPR endoribonuclease, Csy4, in mammalian cells. While functional in human cells, Csy4 originates from a bacterial system and certain aspects of Csy4 function in the eukaryotic environment remain to be determined.

For example, Csy4 recognizes the substrate hairpin through a number of sequence and structure specific interactions (Haurwitz et al., 2010; Sternberg et al., 2012) (Figure 2). It is possible that Csy4 is capable of interacting with, and potentially cleaving RNA hairpins present in the human transcriptome. Studies identifying these RNAs could yield insight into the safety of Csy4 as a potential therapeutic. Further, if identified, understanding common sequence and structure characteristics of these endogenous Csy4-targeted RNAs would also be beneficial. This information could

additionally provide a starting point for application of Csy4 cleavage on endogenously expressed RNAs. In Chapter 2 of this dissertation, I demonstrated the interaction between Csy4 and the target hairpin by RNA immunoprecipitation (RNA-IP) (Figure 6). A similar strategy, sequencing purified RNA products of the RNA-IP, could reveal additional RNA binding partners.

It would also be worthwhile to investigate any potential interactions between Csy4 and cellular proteins. I observed translation of RNAs lacking poly(A) tails after Csy4 binding and cleavage of the RNA. The mechanism of translation after Csy4 cleavage has not yet been elucidated, and could stem from a number of possibilities. For example, the remaining 3' phosphate and Csy4 binding after cleavage could stabilize the RNA long enough for translation to occur. Further, Csy4 could interact with translational initiation factors to recruit translation machinery. Hijacking of host translation machinery by exogenous protein factors is not unheard of. Indeed, influenza NS1 protein interacts with the eukaryotic translation initiation factor eIF4GI and poly-A binding protein (PABP) (Aragon et al., 2000; Burgui et al., 2003). This interaction recruits translational machinery to the viral RNA for efficient translation (Aragon et al., 2000; Burgui et al., 2003). Therefore, elucidating Csy4 interactions with cellular proteins may provide insight into the mechanism behind translation of Csy4 cleaved mRNAs.

It may also be useful to develop methods for regulating Csy4 activity to exercise tight control over Csy4-mediated effects. Small molecule screens may reveal drugs capable of inhibiting Csy4 catalytic activity. Further, taking advantage of Csy4’s tight interaction with the target hairpin, expression of a decoy hairpin may dilute the effective Csy4 population.

5.3 Future Applications and Engineering of Csy4

Csy4 as a "safety switch"

In this dissertation, I demonstrated new applications of Csy4 for regulating transgene expression and circRNA formation. These findings may serve as a starting point for the development of new technologies and applications of Csy4. For example, Csy4 could act as an effective "safety" switch, for turning off gene expression. Safety switches are becoming increasingly valuable in gene and cell therapy as precautions for unintended reactions to treatment (Jones et al., 2014). The use of cell-killing genes (ie: Caspase-9 and herpes simplex virus thymidine kinase (HSV-TK)) as safety switches has been described previously (Bonini et al., 1997; Di Stasi et al., 2011). In one rendition of a Csy4-based safety switch, the circGFP-CD cassette could be modified for expression of a cell-killing gene in the presence of Csy4. The cell-killing gene would be arranged in a split orientation and replace the split-GFP cassette such that it only expresses in a circular configuration. Meanwhile, the dsRed sequence would be replaced with a therapeutic gene. Thus, in the absence of Csy4, the therapeutic protein would be expressed from a linear RNA. Should the therapy prove detrimental, Csy4 could be delivered for switching off the therapeutic protein and inducing circRNA formation. Subsequent expression of the cell-killing protein would cause cell death to prevent further expression of the detrimental therapeutic.

Csy4 orthologs and recognition of new targets

Exploring the function of Csy4 orthologs from other Class 1 Type I CRISPR

systems in a mammalian context may also prove beneficial (Brouns et al., 2008; Carte et al., 2008; Hochstrasser and Doudna, 2014; Nam et al., 2012). If catalytic activity is retained in a mammalian context, it would be interesting to explore whether these orthologs are capable of gene regulation in a manner similar to Csy4. Like Csy4, many Class I Type I, pre-crRNA processing enzymes remain associated with the cleaved crRNA product (Carte et al., 2010; Nam et al., 2012; Sashital et al., 2011). However, the binding affinities of each enzyme for the cleavage product vary and most generate a 2',3'-cyclic phosphate terminal end (Jore et al., 2011; Nam et al., 2012). This differs from the tight association ($K_d \approx 50$ pM) between Csy4 and the 3'-phosphate-containing cleavage product (Sternberg et al., 2012; Wiedenheft et al., 2011). Therefore, it would be worthwhile to investigate whether these orthologs also stabilize 3' ends and support mRNA translation.

It may also be possible to engineer Csy4 for binding and cleaving new substrates. Substrate recognition is carried out through a series of sequence- and structure-specific contacts between Csy4 and the target RNA hairpin (Haurwitz et al., 2010; Sternberg et al., 2012). Structure-specific contacts are mediated by an Arginine rich helix which docks into the major groove of the target RNA hairpin (Haurwitz et al., 2010; Sternberg et al., 2012) (Figure 2). Meanwhile, Arg102, Gln104, and Arg115 are responsible for sequence-specific recognition of G20, A19 and G11 of the hairpin respectively (Sternberg et al., 2012) (Figure 2). Lastly, Phe155 orients G20 in the active site through base-stacking interactions (Haurwitz et al., 2010; Sternberg et al., 2012) (Figure 2). Approaches for engineering Csy4 for new substrate binding might include randomization of codons encoding residues involved in hairpin recognition. A proof-of-

concept strategy could involve restricting randomization to the residues recognizing specific nucleotides of the target hairpin (Arg102, Gln104, Arg115). This approach would potentially yield Csy4 variants capable of binding hairpins with differing nucleotides at positions 11, 19, and 20 of the hairpin. However, perhaps future efforts could be directed at engineering Csy4 to bind RNA structures differing from the canonical short hairpin.

5.4 Improving Current CircRNA Expression Cassettes

In Chapter 4 of this dissertation, I demonstrated that rAAV vectors can be utilized for delivering circRNA expression cassettes in mice. The circRNAs produced from these cassettes can be translated *in vivo* and persist until at least 16 weeks post-injection. While these vectors are a promising start towards circRNA based therapeutics, the efficiency of circRNA biogenesis and translation of circRNAs can likely be improved. To this end, I propose two strategies for optimizing AAV circRNA vectors.

First, translation efficiency can be optimized through the incorporation of various Internal Ribosome Entry Site (IRES) elements. The current vectors utilized in Chapter 4 rely on the widely expressed encephalomyocarditis virus (EMCV) IRES for protein translation. However, exchanging this IRES for others with alternate translation specificities and efficiencies may help fine tune expression for therapeutics of interest. For example, the Fibroblast Growth Factor-2 (FGF-2) IRES exhibits tissue specific translation activity with highest expression in mouse testis and brain (Creancier et al., 2000). Thus, exchanging the EMCV IRES for the FGF-2 IRES in circRNA-based protein expression systems may add a layer of tissue specificity.

A second strategy for improving AAV circRNA vectors is through enhancing circRNA biogenesis. CircRNA formation requires canonical splicing machinery and splicing signals such as the polypyrimidine tract and branch point (Starke et al., 2015). However, additional factors have been implicated in circRNA biogenesis. For example, Alu repeats have been identified flanking exons known to circularize (Ashwal-Fluss et al., 2014; Ivanov et al., 2015; Jeck et al., 2013; Kramer et al., 2015; Liang and Wilusz, 2014; Zhang et al., 2014). It is hypothesized that these repeat sequences base pair to bring splicing signals into close proximity for backsplicing. Indeed, mutations disrupting base pairing between these flanking sequences reduces the efficiency of circRNA formation (Kramer et al., 2015; Liang and Wilusz, 2014). The current circGFP cassette utilizes a set of engineered IGF2BP1 introns which were designed to base pair to promote circRNA biogenesis. However, the use of naturally occurring, Alu repeat containing introns may provide a more effective option for promoting circularization. Furthermore, as circRNAs often display tissue specific expression patterns, these introns could potentially contribute to tissue specificity of circRNA biogenesis (Liang and Wilusz, 2014; Salzman et al., 2013; Zheng et al., 2016).

I sought to explore the benefit of Alu repeat-containing introns on circRNA biogenesis *in vivo*. Towards this end, I cloned portions of introns flanking ZKSCAN1 exons two and three and those flanking HIPK3 exon two into GFP reporter rAAV plasmid backbones (Liang and Wilusz, 2014) (Figure 17A). I expect that, for therapeutic applications, maximal protein expression would be attained from translating protein from both linear and circular RNAs. Therefore, initial characterization of these vectors was carried out using an un-split GFP reporter exon. However, it will be important to also

investigate reporter expression of these cassettes using split-GFP cassettes in the future. I also utilized the control IRES-GFP cassette, which lacks flanking intron sequences and corresponding splicing signals (IRES-GFP) (Figure 17A).

Reporter cassettes were packaged into rAAV and characterized by infecting HEK293 cells at 50,000 MOI (multiplicity of infection). GFP expression for each vector was observed at 48 hours post-transfection by fluorescence microscopy (Figure 17B). However, I cannot definitively conclude that GFP signal is a result of expression from circRNA over linear mRNA. To demonstrate the presence of circRNAs, I isolated total cellular RNA and treated with RNase R. Linear RNAs are susceptible to this 3' to 5' endoribonuclease while circRNAs are resistant due to their lack of exposed ends (Suzuki et al., 2006). GFP RNA was detected from each condition at similar levels in the absence of RNase R (Figure 17C). As expected for a linear RNA, IRES-GFP RNA was depleted upon treatment with RNase R. Conversely, ZKSCAN1-GFP and HIPK3-GFP RNA demonstrated resistance to RNase R, consistent with circRNA formation (Figure 17C). Together, these data suggest that the ZKSCAN1-GFP and HIPK3-GFP rAAV are capable of producing circRNA and can express GFP protein.

I next delivered each rAAV vector into mice, in order to understand any tissue or cell-type specific differences in expression conferred by each set of introns. I utilized AAV serotype 9, which demonstrates widespread expression throughout mouse tissues (Pulicherla et al., 2011b). Adult BALB/c mice were injected intravenously with rAAV packaging either IRES-GFP, ZKSCAN1-GFP, or HIPK3-GFP (five mice per cassette). Six weeks post-injection, heart and liver were harvested from each mouse and GFP was detected using 3,3'-Diaminobenzidine (DAB) staining. In the heart, ZKSCAN1-GFP

displayed the most abundant GFP levels relative to the other tested cassettes (Figure 18A). HIPK3-GFP also expressed GFP in the heart, though, not to the levels observed with ZKSCAN1 introns (Figure 18A). The IRES-GFP control displayed minimal GFP expression, highlighting the importance of circularization for GFP expression (Figure 18A). In contrast, expression from all vectors in the liver was consistently low (Figure 18B). This pattern may be a consequence of rAAV vector genome dilution resulting from liver proliferation. However, further studies will need to be completed to fully understand this observation.

CircRNAs are particularly abundant in the mammalian brain (Rybak-Wolf et al., 2015; You et al., 2015). Therefore, I tested expression of our reporter cassettes in the mouse brain. rAAV vectors expressing each reporter construct were separately administered into P1 (postnatal day 1) pups via intracerebroventricular (ICV) injection. Brains were harvested six weeks post-injection, sectioned, and GFP expression was detected using DAB staining. Each cassette displayed unique expression patterns characterized by both regional and cell type expression. GFP positive cells for control IRES-GFP were detected occasionally in the cortex and the lining of the lateral ventricle (Figure 18C). Cortical expression was primarily neuronal with occasional astrocytic staining. In contrast, ZKSCAN1-GFP mediated GFP expression was more widespread. GFP positive cells were detected in the olfactory bulb, cortex, and occasionally in the hippocampus and cerebellum (Figure 17C). Cortical and hippocampus expression was primarily astrocytic while expression in the olfactory bulb and cerebellum was largely neuronal. HIPK3-GFP, in contrast, was expressed mainly in cortical astrocytes, with

occasional neuronal staining in the olfactory bulb, and expression in the lining of the ventricle (Figure 18C).

As the only difference between these cassettes is the flanking introns, these expression patterns are likely a conferred by the introns. Indeed, expression for certain circRNAs in the brain is region specific, which may imply regulation by regional and/or cell-type specific factors (Rybak-Wolf et al., 2015; Veno et al., 2015).

5.5 CircRNAs Containing an EMCV IRES are Efficiently Translated in Mice

There was a striking difference between protein expression from the same circRNA expression vectors in cell culture versus in mouse tissues. For example, when IRES-GFP, ZKSCAN1-GFP and HIPK3-GFP were delivered into HEK293 cells via rAAV, GFP fluorescence was similar between cells expressing each construct (Figure 17B). However, in mouse tissue (both in the heart and the brain) GFP production from each cassette was starkly different compared to HEK293 cells (Figure 18). IRES-GFP generated minimal GFP signal while the ZKSCAN1 and HIPK3 intron flanked cassettes produced strong GFP signal (Figure 18). This difference in expression pattern between cultured cells and mouse tissue reflects improved EMCV IRES translation efficiency in the background of mouse cardiac and brain tissue compared to HEK293 cells. Furthermore, these differences also suggest that the context of a linear RNA vs a circular RNA also plays a role in translation efficiency from the IRES. Therefore, it would be worthwhile to conduct additional studies to evaluate the contribution of the IRES on protein expression from these cassettes *in vivo* and *in vitro*, and in a linear versus

circular RNA context. These studies would allow for more informed design of therapeutic cassettes.

5.6 Future Evaluation of rAAV CircRNA Expression Cassettes

Further studies will need to be completed to confirm the presence of circRNAs in the tissues tested here. RNase R digestion of RNA harvested from mice expressing each rAAV vector would determine the ratio of circular to linear RNA present in each tissue. Further, use of split-GFP cassettes (as used in Chapter 4) flanked by each intron pair would more accurately reflect circRNA-specific translation products. Here, identification of neurons and astrocytes was determined by visual inspection of cell morphology. Immunostaining for cell-type specific markers would confirm cell identity. In addition, it would be informative to quantify vector genomes from each tissue to validate that each mouse received a similar dose of AAV. This would rule out expression differences attributed to lower effective vector delivery. Last, evaluation of toxicity and changes in overall cellular splicing patterns will be important for establishing the safety of circRNA platforms for gene expression.

5.7 Final Remarks

This dissertation described the development of new approaches to regulate RNA stability. I demonstrated that the bacterial CRISPR endoribonuclease Csy4/Cas6f is not only functional in human cells, but can also be applied to regulate RNA stability and circRNA biogenesis. In addition, I demonstrated delivery of translatable circRNA expression cassettes *in vivo* using rAAV vectors. The results described here add novel

tools to the array of CRISPR-derived technologies and raise new questions regarding translation of Csy4-bound RNAs. Further, these findings also provide a starting point for developing RNA-based therapeutic safety-switches. Finally, these studies lay the groundwork for expressing therapeutic circRNA cassettes via rAAV vectors.

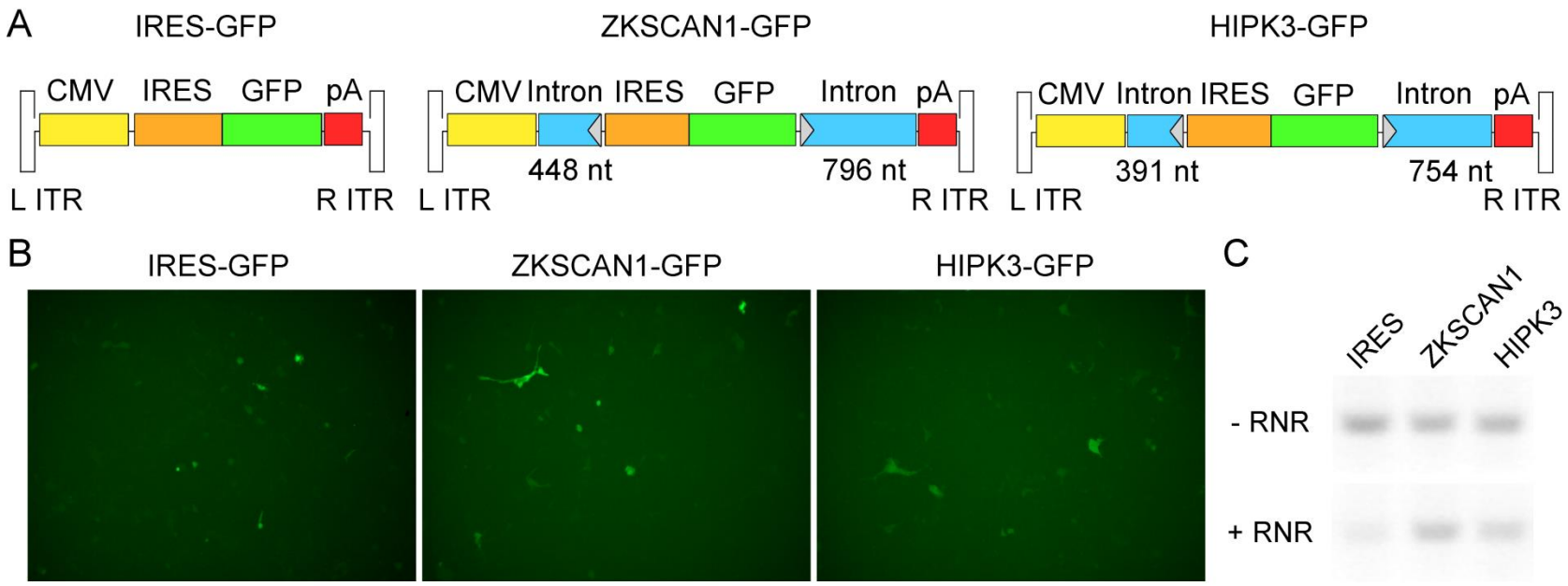

Figure 17: *In vitro* expression of circRNA reporters from viral vectors. (A) Schematic representations of rAAV genomes encoding reporter cassettes IRES-GFP, ZKSCAN1-GFP and HIPK3-GFP. The length of each intron segment is indicated below the respective intron. Splice sites are represented by grey triangles. **(B)** GFP expression detected by fluorescence microscopy from HEK293 cells. HEK293 cells were infected at 50,000 Multiplicity of Infection (MOI) with rAAV vectors expressing either IRES-GFP, ZKSCAN1-GFP, or HIPK3-GFP. **(C)** PCR products from cDNA template generated after RNase R treatment of RNA. Primers were designed to detect the GFP coding sequence. RNA was isolated from HEK293 cells infected with IRES-GFP, ZKSCAN1-GFP, or HIPK3-GFP at 50,000 MOI.

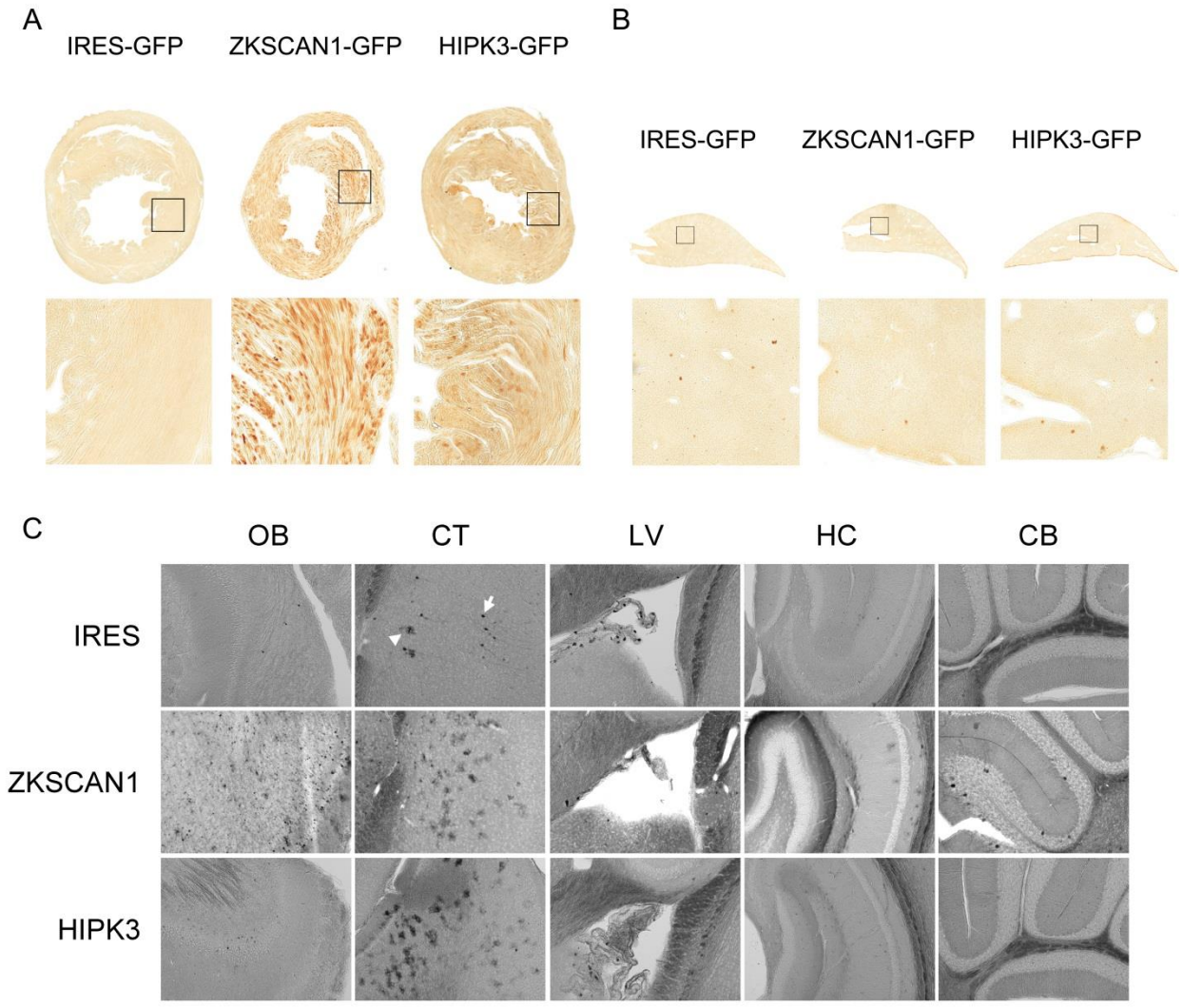

Figure 18: *in vivo* expression of circRNA reporters from viral vectors. (A, B) 3,3'-Diaminobenzidine (DAB) stained heart **(A)** and liver **(B)** sections detecting GFP expression from rAAV-delivered IRES-GFP, ZKSCAN1-GFP or HIPK3-GFP cassettes. Slide scans (top) of whole sections display global expression patterns while selected regions at higher magnification (bottom) reveal expression of individual cells. **(C)** DAB stained sagittal brain sections detecting GFP expression from rAAV-delivered IRES-GFP, ZKSCAN1-GFP or HIPK3-GFP cassettes. Various brain regions are indicated by the following abbreviations. OB: Olfactory bulb; CT: cortex; LV: lateral ventricle; HC: hippocampus; CB: cerebellum. Examples of astrocytic cells and neuronal cells are indicated by a filled triangle or arrowhead respectively.

REFERENCES

Aragon, T., de la Luna, S., Novoa, I., Carrasco, L., Ortin, J. and Nieto, A. 2000. Eukaryotic translation initiation factor 4GI is a cellular target for NS1 protein, a translational activator of influenza virus. *Mol. Cell. Biol.* **20**: 6259-6268.

Ashwal-Fluss, R., Meyer, M., Pamudurti, N.R., Ivanov, A., Bartok, O., Hanan, M., Evantal, N., Memczak, S., Rajewsky, N. and Kadener, S. 2014. circRNA biogenesis competes with pre-mRNA splicing. *Mol. Cell* **56**: 55-66.

Asokan, A. and Samulski, R.J. 2013. An emerging adeno-associated viral vector pipeline for cardiac gene therapy. *Hum. Gene Ther.* **24**: 906-913.

Asokan, A., Schaffer, D.V. and Samulski, R.J. 2012. The AAV vector toolkit: poised at the clinical crossroads. *Mol. Ther.* **20**: 699-708.

Babendure, J.R., Babendure, J.L., Ding, J.H. and Tsien, R.Y. 2006. Control of mammalian translation by mRNA structure near caps. *RNA* **12**: 851-861.

Barrangou, R., Fremaux, C., Deveau, H., Richards, M., Boyaval, P., Moineau, S., Romero, D.A. and Horvath, P. 2007a. CRISPR provides acquired resistance against viruses in prokaryotes. *Science* **315**: 1709-1712.

Barrangou, R., Fremaux, C., Deveau, H., Richards, M., Boyaval, P., Moineau, S., Romero, D.A. and Horvath, P. 2007b. CRISPR provides acquired resistance against viruses in prokaryotes. *Science* **315**: 1709-1712.

Bolotin, A., Quinquis, B., Sorokin, A. and Ehrlich, S.D. 2005. Clustered regularly interspaced short palindrome repeats (CRISPRs) have spacers of extrachromosomal origin. *Microbiology* **151**: 2551-2561.

Bonini, C., Ferrari, G., Verzeletti, S., Servida, P., Zappone, E., Ruggieri, L., Ponzoni, M., Rossini, S., Mavilio, F., Traversari, C. et al. 1997. HSV-TK gene transfer into donor lymphocytes for control of allogeneic graft-versus-leukemia. *Science* **276**: 1719-1724.

Borchardt, E.K., Vandoros, L.A., Huang, M., Lackey, P.E., Marzluff, W.F. and Asokan, A. 2015. Controlling mRNA stability and translation with the CRISPR endoribonuclease Csy4. *RNA* **21**: 1921-1930.

Borel, F., Kay, M.A. and Mueller, C. 2014. Recombinant AAV as a platform for translating the therapeutic potential of RNA interference. *Mol. Ther.* **22**: 692-701.

Brimble, M.A., Reiss, U.M., Nathwani, A.C. and Davidoff, A.M. 2016. New and improved AAVenues: current status of hemophilia B gene therapy. *Expert Opin. Biol. Ther.* **16**: 79-92.

Brouns, S.J., Jore, M.M., Lundgren, M., Westra, E.R., Slijkhuis, R.J., Snijders, A.P., Dickman, M.J., Makarova, K.S., Koonin, E.V. and van der Oost, J. 2008. Small CRISPR RNAs guide antiviral defense in prokaryotes. *Science* **321**: 960-964.

Brown, J.A., Valenstein, M.L., Yario, T.A., Tycowski, K.T. and Steitz, J.A. 2012. Formation of triple-helical structures by the 3'-end sequences of MALAT1 and MENbeta noncoding RNAs. *Proc. Natl. Acad. Sci. U. S. A.* **109**: 19202-19207.

Bryant, L.M., Christopher, D.M., Giles, A.R., Hinderer, C., Rodriguez, J.L., Smith, J.B., Traxler, E.A., Tycko, J., Wojno, A.P. and Wilson, J.M. 2013. Lessons learned from the clinical development and market authorization of Glybera. *Hum. Gene Ther. Clin. Dev.* **24**: 55-64.

Burgui, I., Aragon, T., Ortin, J. and Nieto, A. 2003. PABP1 and eIF4GI associate with influenza virus NS1 protein in viral mRNA translation initiation complexes. *J. Gen. Virol.* **84**: 3263-3274.

Cady, K.C. and O'Toole, G.A. 2011. Non-identity-mediated CRISPR-bacteriophage interaction mediated via the Csy and Cas3 proteins. *J. Bacteriol.* **193**: 3433-3445.

Capel, B., Swain, A., Nicolis, S., Hacker, A., Walter, M., Koopman, P., Goodfellow, P. and Lovell-Badge, R. 1993. Circular transcripts of the testis-determining gene Sry in adult mouse testis. *Cell* **73**: 1019-1030.

Carte, J., Pfister, N.T., Compton, M.M., Terns, R.M. and Terns, M.P. 2010. Binding and cleavage of CRISPR RNA by Cas6. *RNA* **16**: 2181-2188.

Carte, J., Wang, R., Li, H., Terns, R.M. and Terns, M.P. 2008. Cas6 is an endoribonuclease that generates guide RNAs for invader defense in prokaryotes. *Genes Dev.* **22**: 3489-3496.

Chang, A.L., Wolf, J.J. and Smolke, C.D. 2012. Synthetic RNA switches as a tool for temporal and spatial control over gene expression. *Curr. Opin. Biotechnol.* **23**: 679-688.

Chen, C.Y. and Sarnow, P. 1995. Initiation of protein synthesis by the eukaryotic translational apparatus on circular RNAs. *Science* **268**: 415-417.

Choudhury, R., Tsai, Y.S., Dominguez, D., Wang, Y. and Wang, Z. 2012. Engineering RNA endonucleases with customized sequence specificities. *Nat. Commun.* **3**: 1147.

Cideciyan, A.V., Aleman, T.S., Boye, S.L., Schwartz, S.B., Kaushal, S., Roman, A.J., Pang, J.J., Sumaroka, A., Windsor, E.A., Wilson, J.M. et al. 2008. Human gene therapy for RPE65 isomerase deficiency activates the retinoid cycle of vision but with slow rod kinetics. *Proc. Natl. Acad. Sci. U. S. A.* **105**: 15112-15117.

Coller, J.M., Gray, N.K. and Wickens, M.P. 1998. mRNA stabilization by poly(A) binding protein is independent of poly(A) and requires translation. *Genes Dev.* **12**: 3226-3235.

Cong, L., Ran, F.A., Cox, D., Lin, S., Barretto, R., Habib, N., Hsu, P.D., Wu, X., Jiang, W., Marraffini, L.A. et al. 2013. Multiplex genome engineering using CRISPR/Cas systems. *Science* **339**: 819-823.

Conn, S.J., Pillman, K.A., Toubia, J., Conn, V.M., Salmanidis, M., Phillips, C.A., Roslan, S., Schreiber, A.W., Gregory, P.A. and Goodall, G.J. 2015. The RNA binding protein quaking regulates formation of circRNAs. *Cell* **160**: 1125-1134.

Conrad, N.K., Mili, S., Marshall, E.L., Shu, M.D. and Steitz, J.A. 2006. Identification of a rapid mammalian deadenylation-dependent decay pathway and its inhibition by a viral RNA element. *Mol. Cell* **24**: 943-953.

Creancier, L., Morello, D., Mercier, P. and Prats, A.C. 2000. Fibroblast growth factor 2 internal ribosome entry site (IRES) activity ex vivo and in transgenic mice reveals a stringent tissue-specific regulation. *J. Cell Biol.* **150**: 275-281.

Danan, M., Schwartz, S., Edelheit, S. and Sorek, R. 2012. Transcriptome-wide discovery of circular RNAs in Archaea. *Nucleic Acids Res.* **40**: 3131-3142.

Datsenko, K.A., Pougach, K., Tikhonov, A., Wanner, B.L., Severinov, K. and Semenova, E. 2012. Molecular memory of prior infections activates the CRISPR/Cas adaptive bacterial immunity system. *Nat. Commun.* **3**: 945.

Deltcheva, E., Chylinski, K., Sharma, C.M., Gonzales, K., Chao, Y., Pirzada, Z.A., Eckert, M.R., Vogel, J. and Charpentier, E. 2011. CRISPR RNA maturation by trans-encoded small RNA and host factor RNase III. *Nature* **471**: 602-607.

Deveau, H., Barrangou, R., Garneau, J.E., Labonte, J., Fremaux, C., Boyaval, P., Romero, D.A., Horvath, P. and Moineau, S. 2008. Phage response to CRISPR-encoded resistance in Streptococcus thermophilus. *J. Bacteriol.* **190**: 1390-1400.

Di Stasi, A., Tey, S.K., Dotti, G., Fujita, Y., Kennedy-Nasser, A., Martinez, C., Straathof, K., Liu, E., Durett, A.G., Grilley, B. et al. 2011. Inducible apoptosis as a safety switch for adoptive cell therapy. *N. Engl. J. Med.* **365**: 1673-1683.

DiCarlo, J.E., Norville, J.E., Mali, P., Rios, X., Aach, J. and Church, G.M. 2013. Genome engineering in Saccharomyces cerevisiae using CRISPR-Cas systems. *Nucleic Acids Res.* **41**: 4336-4343.

Dong, R., Zhang, X.O., Zhang, Y., Ma, X.K., Chen, L.L. and Yang, L. 2016. CircRNA-derived pseudogenes. *Cell Res.*

Du, P., Miao, C., Lou, Q., Wang, Z. and Lou, C. 2016a. Engineering Translational Activators with CRISPR-Cas System. *ACS Synth. Biol.* **5**: 74-80.

Du, W.W., Yang, W., Liu, E., Yang, Z., Dhaliwal, P. and Yang, B.B. 2016b. Foxo3 circular RNA retards cell cycle progression via forming ternary complexes with p21 and CDK2. *Nucleic Acids Res.*

Ebbesen, K.K., Kjems, J. and Hansen, T.B. 2015. Circular RNAs: Identification, biogenesis and function. *Biochim. Biophys. Acta*.

Enuka, Y., Lauriola, M., Feldman, M.E., Sas-Chen, A., Ulitsky, I. and Yarden, Y. 2015. Circular RNAs are long-lived and display only minimal early alterations in response to a growth factor. *Nucleic Acids Res.*

Friedland, A.E., Tzur, Y.B., Esvelt, K.M., Colaiacovo, M.P., Church, G.M. and Calarco, J.A. 2013. Heritable genome editing in C. elegans via a CRISPR-Cas9 system. *Nat. Methods* **10**: 741-743.

Garneau, J.E., Dupuis, M.E., Villion, M., Romero, D.A., Barrangou, R., Boyaval, P., Fremaux, C., Horvath, P., Magadan, A.H. and Moineau, S. 2010a. The CRISPR/Cas bacterial immune system cleaves bacteriophage and plasmid DNA. *Nature* **468**: 67-71.

Garneau, J.E., Dupuis, M.E., Villion, M., Romero, D.A., Barrangou, R., Boyaval, P., Fremaux, C., Horvath, P., Magadan, A.H. and Moineau, S. 2010b. The CRISPR/Cas bacterial immune system cleaves bacteriophage and plasmid DNA. *Nature* **468**: 67-71.

Gaudet, D., de Wal, J., Tremblay, K., Dery, S., van Deventer, S., Freidig, A., Brisson, D. and Methot, J. 2010. Review of the clinical development of alipogene tiparvovec gene therapy for lipoprotein lipase deficiency. *Atheroscler. Suppl.* **11**: 55-60.

Gilbert, L.A., Larson, M.H., Morsut, L., Liu, Z., Brar, G.A., Torres, S.E., Stern-Ginossar, N., Brandman, O., Whitehead, E.H., Doudna, J.A. et al. 2013. CRISPR-mediated modular RNA-guided regulation of transcription in eukaryotes. *Cell* **154**: 442-451.

Gratz, S.J., Cummings, A.M., Nguyen, J.N., Hamm, D.C., Donohue, L.K., Harrison, M.M., Wildonger, J. and O'Connor-Giles, K.M. 2013. Genome engineering of Drosophila with the CRISPR RNA-guided Cas9 nuclease. *Genetics* **194**: 1029-1035.

Grimm, D., Streetz, K.L., Jopling, C.L., Storm, T.A., Pandey, K., Davis, C.R., Marion, P., Salazar, F. and Kay, M.A. 2006. Fatality in mice due to oversaturation of cellular microRNA/short hairpin RNA pathways. *Nature* **441**: 537-541.

Grissa, I., Vergnaud, G. and Pourcel, C. 2007. The CRISPRdb database and tools to display CRISPRs and to generate dictionaries of spacers and repeats. *BMC Bioinformatics* **8**: 172.

Grissa, I., Vergnaud, G. and Pourcel, C. 2009. Clustered regularly interspaced short palindromic repeats (CRISPRs) for the genotyping of bacterial pathogens. *Methods Mol. Biol.* **551**: 105-116.

Groher, F. and Suess, B. 2014. Synthetic riboswitches - A tool comes of age. *Biochim. Biophys. Acta* **1839**: 964-973.

Guo, J.U., Agarwal, V., Guo, H. and Bartel, D.P. 2014. Expanded identification and characterization of mammalian circular RNAs. *Genome Biol.* **15**: 409-014-0409-z.

Guo, R., Wang, H., Cui, J., Wang, G., Li, W. and Hu, J.F. 2015. Inhibition of HIV-1 Viral Infection by an Engineered CRISPR Csy4 RNA Endoribonuclease. *PLoS One* **10**: e0141335.

Hansen, T.B., Jensen, T.I., Clausen, B.H., Bramsen, J.B., Finsen, B., Damgaard, C.K. and Kjems, J. 2013. Natural RNA circles function as efficient microRNA sponges. *Nature* **495**: 384-388.

Hatoum-Aslan, A., Maniv, I., Samai, P. and Marraffini, L.A. 2014. Genetic characterization of antiplasmid immunity through a type III-A CRISPR-Cas system. *J. Bacteriol.* **196**: 310-317.

Haurwitz, R.E., Jinek, M., Wiedenheft, B., Zhou, K. and Doudna, J.A. 2010. Sequence- and structure-specific RNA processing by a CRISPR endonuclease. *Science* **329**: 1355-1358.

Heler, R., Samai, P., Modell, J.W., Weiner, C., Goldberg, G.W., Bikard, D. and Marraffini, L.A. 2015. Cas9 specifies functional viral targets during CRISPR-Cas adaptation. *Nature* **519**: 199-202.

Hinderer, C., Bell, P., Louboutin, J.P., Zhu, Y., Yu, H., Lin, G., Choa, R., Gurda, B.L., Bagel, J., O'Donnell, P. et al. 2015. Neonatal Systemic AAV Induces Tolerance to CNS Gene Therapy in MPS I Dogs and Nonhuman Primates. *Mol. Ther.* **23**: 1298-1307.

Hochstrasser, M.L. and Doudna, J.A. 2014. Cutting it close: CRISPR-associated endoribonuclease structure and function. *Trends Biochem. Sci.*

Horvath, P., Romero, D.A., Coute-Monvoisin, A.C., Richards, M., Deveau, H., Moineau, S., Boyaval, P., Fremaux, C. and Barrangou, R. 2008. Diversity, activity, and evolution of CRISPR loci in Streptococcus thermophilus. *J. Bacteriol.* **190**: 1401-1412.

Hutchinson, J.N., Ensminger, A.W., Clemson, C.M., Lynch, C.R., Lawrence, J.B. and Chess, A. 2007. A screen for nuclear transcripts identifies two linked noncoding RNAs associated with SC35 splicing domains. *BMC Genomics* **8**: 39.

Hwang, W.Y., Fu, Y., Reyon, D., Maeder, M.L., Tsai, S.Q., Sander, J.D., Peterson, R.T., Yeh, J.R. and Joung, J.K. 2013. Efficient genome editing in zebrafish using a CRISPR-Cas system. *Nat. Biotechnol.* **31**: 227-229.

Ivanov, A., Memczak, S., Wyler, E., Torti, F., Porath, H.T., Orejuela, M.R., Piechotta, M., Levanon, E.Y., Landthaler, M., Dieterich, C. et al. 2015. Analysis of intron sequences reveals hallmarks of circular RNA biogenesis in animals. *Cell. Rep.* **10**: 170-177.

Jeck, W.R., Sorrentino, J.A., Wang, K., Slevin, M.K., Burd, C.E., Liu, J., Marzluff, W.F. and Sharpless, N.E. 2013. Circular RNAs are abundant, conserved, and associated with ALU repeats. *RNA* **19**: 141-157.

Jiang, W., Bikard, D., Cox, D., Zhang, F. and Marraffini, L.A. 2013. RNA-guided editing of bacterial genomes using CRISPR-Cas systems. *Nat. Biotechnol.* **31**: 233-239.

Jinek, M., Chylinski, K., Fonfara, I., Hauer, M., Doudna, J.A. and Charpentier, E. 2012. A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science* **337**: 816-821.

Jinek, M., East, A., Cheng, A., Lin, S., Ma, E. and Doudna, J. 2013. RNA-programmed genome editing in human cells. *Elife* **2**: e00471.

Jones, B.S., Lamb, L.S., Goldman, F. and Di Stasi, A. 2014. Improving the safety of cell therapy products by suicide gene transfer. *Front. Pharmacol.* **5**: 254.

Jore, M.M., Lundgren, M., van Duijn, E., Bultema, J.B., Westra, E.R., Waghmare, S.P., Wiedenheft, B., Pul, U., Wurm, R., Wagner, R. et al. 2011. Structural basis for CRISPR RNA-guided DNA recognition by Cascade. *Nat. Struct. Mol. Biol.* **18**: 529-536.

Karvelis, T., Gasiunas, G., Miksys, A., Barrangou, R., Horvath, P. and Siksnys, V. 2013. crRNA and tracrRNA guide Cas9-mediated DNA interference in Streptococcus thermophilus. *RNA Biol.* **10**: 841-851.

Kozak, M. 1986. Influences of mRNA secondary structure on initiation by eukaryotic ribosomes. *Proc. Natl. Acad. Sci. U. S. A.* **83**: 2850-2854.

Kozak, M. 1989. Circumstances and mechanisms of inhibition of translation by secondary structure in eucaryotic mRNAs. *Mol. Cell. Biol.* **9**: 5134-5142.

Kramer, M.C., Liang, D., Tatomer, D.C., Gold, B., March, Z.M., Cherry, S. and Wilusz, J.E. 2015. Combinatorial control of Drosophila circular RNA expression by intronic repeats, hnRNPs, and SR proteins. *Genes Dev.* **29**: 2168-2182.

Kunin, V., Sorek, R. and Hugenholtz, P. 2007. Evolutionary conservation of sequence and secondary structures in CRISPR repeats. *Genome Biol.* **8**: R61.

Langevin, S.A., Bent, Z.W., Solberg, O.D., Curtis, D.J., Lane, P.D., Williams, K.P., Schoeniger, J.S., Sinha, A., Lane, T.W. and Branda, S.S. 2013. Peregrine: A rapid and unbiased method to produce strand-specific RNA-Seq libraries from small quantities of starting material. *RNA Biol.* **10**: 502-515.

Lasda, E. and Parker, R. 2014. Circular RNAs: diversity of form and function. *RNA* **20**: 1829-1842.

Lee, H.Y., Haurwitz, R.E., Apffel, A., Zhou, K., Smart, B., Wenger, C.D., Laderman, S., Bruhn, L. and Doudna, J.A. 2013. RNA-protein analysis using a conditional CRISPR nuclease. *Proc. Natl. Acad. Sci. U. S. A.* **110**: 5416-5421.

Li, D., Qiu, Z., Shao, Y., Chen, Y., Guan, Y., Liu, M., Li, Y., Gao, N., Wang, L., Lu, X. et al. 2013. Heritable gene targeting in the mouse and rat using a CRISPR-Cas system. *Nat. Biotechnol.* **31**: 681-683.

Li, Z., Huang, C., Bao, C., Chen, L., Lin, M., Wang, X., Zhong, G., Yu, B., Hu, W., Dai, L. et al. 2015. Exon-intron circular RNAs regulate transcription in the nucleus. *Nat. Struct. Mol. Biol.* **22**: 256-264.

Liang, D. and Wilusz, J.E. 2014. Short intronic repeat sequences facilitate circular RNA production. *Genes Dev.* **28**: 2233-2247.

Liang, J.C., Bloom, R.J. and Smolke, C.D. 2011. Engineering biological systems with synthetic RNA molecules. *Mol. Cell* **43**: 915-926.

Long, C., Amoasii, L., Mireault, A.A., McAnally, J.R., Li, H., Sanchez-Ortiz, E., Bhattacharyya, S., Shelton, J.M., Bassel-Duby, R. and Olson, E.N. 2016. Postnatal genome editing partially restores dystrophin expression in a mouse model of muscular dystrophy. *Science* **351**: 400-403.

Lu, Z., Filonov, G.S., Noto, J.J., Schmidt, C.A., Hatkevich, T.L., Wen, Y., Jaffrey, S.R. and Matera, A.G. 2015. Metazoan tRNA introns generate stable circular RNAs in vivo. *RNA* **21**: 1554-1565.

Maguire, A.M., Simonelli, F., Pierce, E.A., Pugh, E.N.,Jr, Mingozzi, F., Bennicelli, J., Banfi, S., Marshall, K.A., Testa, F., Surace, E.M. et al. 2008. Safety and efficacy of gene transfer for Leber's congenital amaurosis. *N. Engl. J. Med.* **358**: 2240-2248.

Makarova, K.S., Wolf, Y.I., Alkhnbashi, O.S., Costa, F., Shah, S.A., Saunders, S.J., Barrangou, R., Brouns, S.J., Charpentier, E., Haft, D.H. et al. 2015. An updated evolutionary classification of CRISPR-Cas systems. *Nat. Rev. Microbiol.* **13**: 722-736.

Mali, P., Aach, J., Stranges, P.B., Esvelt, K.M., Moosburner, M., Kosuri, S., Yang, L. and Church, G.M. 2013a. CAS9 transcriptional activators for target specificity screening and paired nickases for cooperative genome engineering. *Nat. Biotechnol.* **31**: 833-838.

Mali, P., Yang, L., Esvelt, K.M., Aach, J., Guell, M., DiCarlo, J.E., Norville, J.E. and Church, G.M. 2013b. RNA-guided human genome engineering via Cas9. *Science* **339**: 823-826.

Marraffini, L.A. 2015. CRISPR-Cas immunity in prokaryotes. *Nature* **526**: 55-61.

Marzluff, W.F. 2012. Novel 3' ends that support translation. *Genes Dev.* **26**: 2457-2460.

Marzluff, W.F., Wagner, E.J. and Duronio, R.J. 2008. Metabolism and regulation of canonical histone mRNAs: life without a poly(A) tail. *Nat. Rev. Genet.* **9**: 843-854.

Memczak, S., Jens, M., Elefsinioti, A., Torti, F., Krueger, J., Rybak, A., Maier, L., Mackowiak, S.D., Gregersen, L.H., Munschauer, M. et al. 2013. Circular RNAs are a large class of animal RNAs with regulatory potency. *Nature* **495**: 333-338.

Mitton-Fry, R.M., DeGregorio, S.J., Wang, J., Steitz, T.A. and Steitz, J.A. 2010. Poly(A) tail recognition by a viral RNA element through assembly of a triple helix. *Science* **330**: 1244-1247.

Mojica, F.J., Diez-Villasenor, C., Garcia-Martinez, J. and Almendros, C. 2009. Short motif sequences determine the targets of the prokaryotic CRISPR defence system. *Microbiology* **155**: 733-740.

Mojica, F.J., Diez-Villasenor, C., Garcia-Martinez, J. and Soria, E. 2005. Intervening sequences of regularly spaced prokaryotic repeats derive from foreign genetic elements. *J. Mol. Evol.* **60**: 174-182.

Murlidharan, G., Samulski, R.J. and Asokan, A. 2014. Biology of adeno-associated viral vectors in the central nervous system. *Front. Mol. Neurosci.* **7**: 76.

Nam, K.H., Haitjema, C., Liu, X., Ding, F., Wang, H., DeLisa, M.P. and Ke, A. 2012. Cas5d protein processes pre-crRNA and assembles into a cascade-like interference complex in subtype I-C/Dvulg CRISPR-Cas system. *Structure* **20**: 1574-1584.

Nathwani, A.C., Reiss, U.M., Tuddenham, E.G., Rosales, C., Chowdary, P., McIntosh, J., Della Peruta, M., Lheriteau, E., Patel, N., Raj, D. et al. 2014. Long-term safety and efficacy of factor IX gene therapy in hemophilia B. *N. Engl. J. Med.* **371**: 1994-2004.

Nekrasov, V., Staskawicz, B., Weigel, D., Jones, J.D. and Kamoun, S. 2013. Targeted mutagenesis in the model plant Nicotiana benthamiana using Cas9 RNA-guided endonuclease. *Nat. Biotechnol.* **31**: 691-693.

Nelles, D.A., Fang, M.Y., O'Connell, M.R., Xu, J.L., Markmiller, S.J., Doudna, J.A. and Yeo, G.W. 2016. Programmable RNA Tracking in Live Cells with CRISPR/Cas9. *Cell*.

Nelson, C.E., Hakim, C.H., Ousterout, D.G., Thakore, P.I., Moreb, E.A., Castellanos Rivera, R.M., Madhavan, S., Pan, X., Ran, F.A., Yan, W.X. et al. 2016. In vivo genome editing improves muscle function in a mouse model of Duchenne muscular dystrophy. *Science* **351**: 403-407.

Nissim, L., Perli, S.D., Fridkin, A., Perez-Pinera, P. and Lu, T.K. 2014. Multiplexed and programmable regulation of gene networks with an integrated RNA and CRISPR/Cas toolkit in human cells. *Mol. Cell* **54**: 698-710.

Nunez, J.K., Lee, A.S., Engelman, A. and Doudna, J.A. 2015. Integrase-mediated spacer acquisition during CRISPR-Cas adaptive immunity. *Nature* **519**: 193-198.

Pourcel, C., Salvignol, G. and Vergnaud, G. 2005. CRISPR elements in Yersinia pestis acquire new repeats by preferential uptake of bacteriophage DNA, and provide additional tools for evolutionary studies. *Microbiology* **151**: 653-663.

Pul, U., Wurm, R., Arslan, Z., Geissen, R., Hofmann, N. and Wagner, R. 2010. Identification and characterization of E. coli CRISPR-cas promoters and their silencing by H-NS. *Mol. Microbiol.* **75**: 1495-1512.

Pulicherla, N., Shen, S., Yadav, S., Debbink, K., Govindasamy, L., Agbandje-McKenna, M. and Asokan, A. 2011a. Engineering liver-detargeted AAV9 vectors for cardiac and musculoskeletal gene transfer. *Mol. Ther.* **19**: 1070-1078.

Pulicherla, N., Shen, S., Yadav, S., Debbink, K., Govindasamy, L., Agbandje-McKenna, M. and Asokan, A. 2011b. Engineering liver-detargeted AAV9 vectors for cardiac and musculoskeletal gene transfer. *Mol. Ther.* **19**: 1070-1078.

Qi, L., Haurwitz, R.E., Shao, W., Doudna, J.A. and Arkin, A.P. 2012. RNA processing enables predictable programming of gene expression. *Nat. Biotechnol.* **30**: 1002-1006.

Rybak-Wolf, A., Stottmeister, C., Glazar, P., Jens, M., Pino, N., Giusti, S., Hanan, M., Behm, M., Bartok, O., Ashwal-Fluss, R. et al. 2015. Circular RNAs in the Mammalian Brain Are Highly Abundant, Conserved, and Dynamically Expressed. *Mol. Cell* **58**: 870-885.

Salvail-Lacoste, A., Di Tomasso, G., Piette, B.L. and Legault, P. 2013. Affinity purification of T7 RNA transcripts with homogeneous ends using ARiBo and CRISPR tags. *RNA* **19**: 1003-1014.

Salzman, J., Chen, R.E., Olsen, M.N., Wang, P.L. and Brown, P.O. 2013. Cell-type specific features of circular RNA expression. *PLoS Genet.* **9**: e1003777.

Salzman, J., Gawad, C., Wang, P.L., Lacayo, N. and Brown, P.O. 2012. Circular RNAs are the predominant transcript isoform from hundreds of human genes in diverse cell types. *PLoS One* **7**: e30733.

Samai, P., Pyenson, N., Jiang, W., Goldberg, G.W., Hatoum-Aslan, A. and Marraffini, L.A. 2015. Co-transcriptional DNA and RNA Cleavage during Type III CRISPR-Cas Immunity. *Cell* **161**: 1164-1174.

Samulski, R.J. and Muzyczka, N. 2014. AAV-Mediated Gene Therapy for Research and Therapeutic Purposes. *Annu. Rev. Virol.* **1**: 427-451.

Sapranauskas, R., Gasiunas, G., Fremaux, C., Barrangou, R., Horvath, P. and Siksnys, V. 2011. The Streptococcus thermophilus CRISPR/Cas system provides immunity in Escherichia coli. *Nucleic Acids Res.* **39**: 9275-9282.

Sashital, D.G., Jinek, M. and Doudna, J.A. 2011. An RNA-induced conformational change required for CRISPR RNA cleavage by the endoribonuclease Cse3. *Nat. Struct. Mol. Biol.* **18**: 680-687.

Schindelin, J., Arganda-Carreras, I., Frise, E., Kaynig, V., Longair, M., Pietzsch, T., Preibisch, S., Rueden, C., Saalfeld, S., Schmid, B. et al. 2012. Fiji: an open-source platform for biological-image analysis. *Nat. Methods* **9**: 676-682.

Selle, K. and Barrangou, R. 2015. CRISPR-Based Technologies and the Future of Food Science. *J. Food Sci.* **80**: R2367-72.

Shah, S.A., Erdmann, S., Mojica, F.J. and Garrett, R.A. 2013. Protospacer recognition motifs: mixed identities and functional diversity. *RNA Biol.* **10**: 891-899.

Shen, S., Horowitz, E.D., Troupes, A.N., Brown, S.M., Pulicherla, N., Samulski, R.J., Agbandje-McKenna, M. and Asokan, A. 2013a. Engraftment of a galactose receptor footprint onto adeno-associated viral capsids improves transduction efficiency. *J. Biol. Chem.*

Shen, S., Horowitz, E.D., Troupes, A.N., Brown, S.M., Pulicherla, N., Samulski, R.J., Agbandje-McKenna, M. and Asokan, A. 2013b. Engraftment of a galactose receptor footprint onto adeno-associated viral capsids improves transduction efficiency. *J. Biol. Chem.* **288**: 28814-28823.

Sinkunas, T., Gasiunas, G., Fremaux, C., Barrangou, R., Horvath, P. and Siksnys, V. 2011. Cas3 is a single-stranded DNA nuclease and ATP-dependent helicase in the CRISPR/Cas immune system. *Embo j.* **30**: 1335-1342.

Sinkunas, T., Gasiunas, G., Waghmare, S.P., Dickman, M.J., Barrangou, R., Horvath, P. and Siksnys, V. 2013. In vitro reconstitution of Cascade-mediated CRISPR immunity in Streptococcus thermophilus. *Embo j.* **32**: 385-394.

Starke, S., Jost, I., Rossbach, O., Schneider, T., Schreiner, S., Hung, L.H. and Bindereif, A. 2015. Exon circularization requires canonical splice signals. *Cell. Rep.* **10**: 103-111.

Sternberg, S.H., Haurwitz, R.E. and Doudna, J.A. 2012. Mechanism of substrate selection by a highly specific CRISPR endoribonuclease. *Rna* **18**: 661-672.

Sternberg, S.H., Redding, S., Jinek, M., Greene, E.C. and Doudna, J.A. 2014. DNA interrogation by the CRISPR RNA-guided endonuclease Cas9. *Nature* **507**: 62-67.

Suzuki, H., Zuo, Y., Wang, J., Zhang, M.Q., Malhotra, A. and Mayeda, A. 2006. Characterization of RNase R-digested cellular RNA source that consists of lariat and circular RNAs from pre-mRNA splicing. *Nucleic Acids Res.* **34**: e63.

Swarts, D.C., Mosterd, C., van Passel, M.W. and Brouns, S.J. 2012. CRISPR interference directs strand specific spacer acquisition. *PLoS One* **7**: e35888.

Szczelkun, M.D., Tikhomirova, M.S., Sinkunas, T., Gasiunas, G., Karvelis, T., Pschera, P., Siksnys, V. and Seidel, R. 2014. Direct observation of R-loop formation by single RNA-guided Cas9 and Cascade effector complexes. *Proc. Natl. Acad. Sci. U. S. A.* **111**: 9798-9803.

Szymczak, A.L. and Vignali, D.A. 2005. Development of 2A peptide-based strategies in the design of multicistronic vectors. *Expert Opin. Biol. Ther.* **5**: 627-638.

Tabebordbar, M., Zhu, K., Cheng, J.K., Chew, W.L., Widrick, J.J., Yan, W.X., Maesner, C., Wu, E.Y., Xiao, R., Ran, F.A. et al. 2016. In vivo gene editing in dystrophic mouse muscle and muscle stem cells. *Science* **351**: 407-411.

Tang, T.H., Bachellerie, J.P., Rozhdestvensky, T., Bortolin, M.L., Huber, H., Drungowski, M., Elge, T., Brosius, J. and Huttenhofer, A. 2002. Identification of 86 candidates for small non-messenger RNAs from the archaeon Archaeoglobus fulgidus. *Proc. Natl. Acad. Sci. U. S. A.* **99**: 7536-7541.

Tang, T.H., Polacek, N., Zywicki, M., Huber, H., Brugger, K., Garrett, R., Bachellerie, J.P. and Huttenhofer, A. 2005. Identification of novel non-coding RNAs as potential antisense regulators in the archaeon Sulfolobus solfataricus. *Mol. Microbiol.* **55**: 469-481.

Tsai, S.Q., Wyvekens, N., Khayter, C., Foden, J.A., Thapar, V., Reyon, D., Goodwin, M.J., Aryee, M.J. and Joung, J.K. 2014. Dimeric CRISPR RNA-guided FokI nucleases for highly specific genome editing. *Nat. Biotechnol.* **32**: 569-576.

Tycowski, K.T., Shu, M.D., Borah, S., Shi, M. and Steitz, J.A. 2012. Conservation of a triple-helix-forming RNA stability element in noncoding and genomic RNAs of diverse viruses. *Cell. Rep.* **2**: 26-32.

Valdmanis, P.N., Gu, S., Chu, K., Jin, L., Zhang, F., Munding, E.M., Zhang, Y., Huang, Y., Kutay, H., Ghoshal, K. et al. 2016. RNA interference-induced hepatotoxicity results from loss of the first synthesized isoform of microRNA-122 in mice. *Nat. Med.*

van Gestel, M.A., van Erp, S., Sanders, L.E., Brans, M.A., Luijendijk, M.C., Merkestein, M., Pasterkamp, R.J. and Adan, R.A. 2014. shRNA-induced saturation of the microRNA pathway in the rat brain. *Gene Ther.* **21**: 205-211.

Veno, M.T., Hansen, T.B., Veno, S.T., Clausen, B.H., Grebing, M., Finsen, B., Holm, I.E. and Kjems, J. 2015. Spatio-temporal regulation of circular RNA expression during porcine embryonic brain development. *Genome Biol.* **16**: 245-015-0801-3.

Vergnaud, G., Li, Y., Gorge, O., Cui, Y., Song, Y., Zhou, D., Grissa, I., Dentovskaya, S.V., Platonov, M.E., Rakin, A. et al. 2007. Analysis of the three Yersinia pestis CRISPR loci provides new tools for phylogenetic studies and possibly for the investigation of ancient DNA. *Adv. Exp. Med. Biol.* **603**: 327-338.

Wang, H., Yang, H., Shivalila, C.S., Dawlaty, M.M., Cheng, A.W., Zhang, F. and Jaenisch, R. 2013. One-step generation of mice carrying mutations in multiple genes by CRISPR/Cas-mediated genome engineering. *Cell* **153**: 910-918.

Wang, P.L., Bao, Y., Yee, M.C., Barrett, S.P., Hogan, G.J., Olsen, M.N., Dinneny, J.R., Brown, P.O. and Salzman, J. 2014. Circular RNA is expressed across the eukaryotic tree of life. *PLoS One* **9**: e90859.

Wang, Y., Cheong, C.G., Hall, T.M. and Wang, Z. 2009. Engineering splicing factors with designed specificities. *Nat. Methods* **6**: 825-830.

Wang, Y. and Wang, Z. 2015. Efficient backsplicing produces translatable circular mRNAs. *RNA* **21**: 172-179.

Wang, Y., Wang, Z. and Tanaka Hall, T.M. 2013. Engineered proteins with Pumilio/fem-3 mRNA binding factor scaffold to manipulate RNA metabolism. *Febs j.* **280**: 3755-3767.

Wei, Y., Chesne, M.T., Terns, R.M. and Terns, M.P. 2015. Sequences spanning the leader-repeat junction mediate CRISPR adaptation to phage in Streptococcus thermophilus. *Nucleic Acids Res.* **43**: 1749-1758.

Westra, E.R., van Erp, P.B., Kunne, T., Wong, S.P., Staals, R.H., Seegers, C.L., Bollen, S., Jore, M.M., Semenova, E., Severinov, K. et al. 2012. CRISPR immunity relies on the consecutive binding and degradation of negatively supercoiled invader DNA by Cascade and Cas3. *Mol. Cell* **46**: 595-605.

Wiedenheft, B., van Duijn, E., Bultema, J.B., Waghmare, S.P., Zhou, K., Barendregt, A., Westphal, W., Heck, A.J., Boekema, E.J., Dickman, M.J. et al. 2011. RNA-guided complex from a bacterial immune system enhances target recognition through seed sequence interactions. *Proc. Natl. Acad. Sci. U. S. A.* **108**: 10092-10097.

Wilusz, J.E., Freier, S.M. and Spector, D.L. 2008. 3' end processing of a long nuclear-retained noncoding RNA yields a tRNA-like cytoplasmic RNA. *Cell* **135**: 919-932.

Wilusz, J.E., JnBaptiste, C.K., Lu, L.Y., Kuhn, C.D., Joshua-Tor, L. and Sharp, P.A. 2012. A triple helix stabilizes the 3' ends of long noncoding RNAs that lack poly(A) tails. *Genes Dev.* **26**: 2392-2407.

Wozniak, J., Wandtke, T., Kopinski, P. and Chorostowska-Wynimko, J. 2015. Challenges and Prospects for Alpha-1 Antitrypsin Deficiency Gene Therapy. *Hum. Gene Ther.* **26**: 709-718.

Wright, A.V., Nunez, J.K. and Doudna, J.A. 2016. Biology and Applications of CRISPR Systems: Harnessing Nature's Toolbox for Genome Engineering. *Cell* **164**: 29-44.

Xiao, X., Li, J. and Samulski, R.J. 1998. Production of high-titer recombinant adeno-associated virus vectors in the absence of helper adenovirus. *J. Virol.* **72**: 2224-2232.

Yosef, I., Goren, M.G. and Qimron, U. 2012. Proteins and DNA elements essential for the CRISPR adaptation process in Escherichia coli. *Nucleic Acids Res.* **40**: 5569-5576.

You, X., Vlatkovic, I., Babic, A., Will, T., Epstein, I., Tushev, G., Akbalik, G., Wang, M., Glock, C., Quedenau, C. et al. 2015. Neural circular RNAs are derived from synaptic genes and regulated by development and plasticity. *Nat. Neurosci.* **18**: 603-610.

Zaphiropoulos, P.G. 1996. Circular RNAs from transcripts of the rat cytochrome P450 2C24 gene: correlation with exon skipping. *Proc. Natl. Acad. Sci. U. S. A.* **93**: 6536-6541.

Zhang, B., Arun, G., Mao, Y.S., Lazar, Z., Hung, G., Bhattacharjee, G., Xiao, X., Booth, C.J., Wu, J., Zhang, C. et al. 2012. The lncRNA Malat1 is dispensable for mouse development but its transcription plays a cis-regulatory role in the adult. *Cell. Rep.* **2**: 111-123.

Zhang, X.O., Wang, H.B., Zhang, Y., Lu, X., Chen, L.L. and Yang, L. 2014. Complementary sequence-mediated exon circularization. *Cell* **159**: 134-147.

Zheng, Q., Bao, C., Guo, W., Li, S., Chen, J., Chen, B., Luo, Y., Lyu, D., Li, Y., Shi, G. et al. 2016. Circular RNA profiling reveals an abundant circHIPK3 that regulates cell growth by sponging multiple miRNAs. *Nat. Commun.* **7**: 11215.