

DEVELOPING A CRISPR-ASSISTED RNA PURIFICATION, PROTEOMIC
METHOD TO IDENTIFY RNA- SPECIFIC PROTEIN INTERACTIONS

DEVELOPING A CRISPR-ASSISTED RNA PURIFICATION, PROTEOMIC
METHOD TO IDENTIFY RNA- SPECIFIC PROTEIN INTERACTIONS

By PIYANKA SIVARAJAH, B.Sc. (Honours)

A Thesis Submitted to the School of Graduate Studies in Partial Fulfillment of the
Requirements for the Degree Master of Science

McMaster University © Copyright by Piyanka Sivarajah, August 2018

McMaster University MASTER OF SCIENCE (2018) Hamilton, Ontario (Biochemistry)

TITLE: Developing a CRISPR-assisted RNA purification, proteomic method to identify
RNA- specific protein interactions

AUTHOR: Piyanka Sivarajah, B.Sc. (McMaster University)

SUPERVISOR: Dr. Yu Lu

NUMBER OF PAGES: xii, 100

ABSTRACT

RNA regulation involves a complex interplay between RNA-sequence modulators and RNA-binding proteins (RBPs), many of which have yet to be identified for specific RNA species. The current methods to identify RBPs, however, are limited by their specificity, affinity and/or their ability to capture the dynamic protein interactions involved in RNA regulation. To overcome these limitations in the current RNA-RBP enrichment methods, we developed an *in vitro* and *in vivo* clustered regularly interspaced short palindromic repeats (CRISPR)-assisted proteomic method (CARP) to enrich specific RNA and therefore, its associated RBPs. In the *in vitro* CARP approach, we found there was ~3-6-fold enrichment of glyceraldehyde 3-phosphate dehydrogenase (GAPDH) messenger RNA (mRNA), though this was accompanied by the retention of non-specific mRNA. Attempts to remove non-specifically bound species by the use of both potent and mild detergents resulted in a corresponding loss of GAPDH mRNA. In the *in vivo* CARP approach, we created and validated constructs expressing CRISPR components. Furthermore, we established a stable cell-line expressing halo-deactivated Cas9 and conditions for gRNA and PAMmer delivery in cells. We did not find a considerable enrichment of GAPDH mRNA with the *in vivo* CARP system. This is the first study to establish conditions for an *in vivo* CRISPR/dCas9 RNA-RBP enrichment tool and to examine, quantitatively, the RNA-specificity of the CRISPR/Cas9 system. Our findings suggest although there was enrichment of the target RNA by the *in vitro* CARP system, there are concerns of specificity to this approach, as dCas9 appears to promiscuously bind to RNA.

ACKNOWLEDGMENTS

First and foremost, I would like to thank my graduate supervisor, Dr. Yu Lu. When I first met Dr. Lu, I was immediately inspired by his passion for science, in particular mass spectrometry, and dedication to creating a strong and friendly lab environment. Dr. Lu was willing to take a chance on me as his first graduate student and this emulated throughout my degree in his trust in my scientific inquiry and constant words of reassurance. His door was always open if I had questions or an obstacle. I could not have imagined a better supervisor and mentor for my studies. I am forever grateful for my time in the Lu Lab, as it fueled both my personal and academic growth. I will take the skills I have learned during this time and undoubtedly apply them in the future.

I would next like to thank my committee members, Dr. Yingfu Li and Dr. Giuseppe Melacini, for their helpful discussions and encouragement. The advice and support I received from you throughout my degree kept me engaged and importantly, helped drive and shape me as growing scientist. I would also like to thank Meg Rashev and Susan McCusker, as they have taught me pivotal skills required to complete this thesis. The time you took to teach me and your expertise is very appreciated.

A special thanks to all my fellow lab members, Michelle, Paige, Lina, Sansi, Jane and Ruilin, for their endless support, experimental contributions, words of encouragements and moments of laughter. I am honoured to have worked with a group of people who are kind, passionate about research and great collaborators. I am grateful for the life-long friendships we all will have. I especially wish to thank Ruilin, Sansi and Lina. Ruilin, we

began our journey together in the lab and your positivity and constant willingness to help never wavered once. I am grateful for the support and help you have provided me throughout my degree. Sansi, you are an amazing collaborator and a fantastic troubleshooter! You have provided me with many words of encouragement in crucial moments and I am grateful for that. Last but not least, Lina. Lina, you are a great scientist, your knowledge about different topics and papers is both inspiring and was so helpful during the obstacles I encountered. You have listened to my constant complaining throughout two years and provided me with the great advice and endless words of support, which I am so thankful for.

Finally, I would like to thank all my family members and friends for their support and understanding. In particular, I would like to give a special thanks to my parents and my brother. They have consistently supported my passions, listened to my complaining and provided me with the best support and love throughout the years. I am both grateful and honoured to have them in my life.

TABLE OF CONTENTS

Abstract.....	iii
Acknowledgements.....	iv
Table of Contents.....	vi
List of Figures.....	viii
List of Tables.....	ix
List of Abbreviations and Symbols.....	x
Declaration of Academic Achievement.....	xii

CHAPTER 1 INTRODUCTION

1.1 RNA Regulation.....	1
1.2 Identifying specific RNA-sequence modulators with CLIP.....	3
1.3 Identifying specific RBPs with MS based proteomics.....	5
1.3.1 Targeting specific RBPs with MS2 and lambda N22.....	6
1.3.2 Oligonucleotide probe approach: ChIRP-MS, RAP-MS and CHART.....	8
1.4 The CRISPR/Cas system.....	12
1.4.1 Structure of Cas9.....	15
1.4.2 CRISPR/Cas9 mechanism.....	16
1.4.3 The CRISPR/Cas9 system and its interaction with RNA.....	18
1.5 Thesis rationale.....	21
1.5.1 Thesis objectives and goal.....	22

CHAPTER 2: MATERIALS AND METHODS

2.1 dCas9 plasmid amplification and sequencing.....	24
2.2 Expression of dCas9.....	24
2.3 Purification of dCas9.....	25
2.4 His- MBP tag removal of dCas9 using TEV protease.....	25
2.5 Biotinylation of dCas9 and streptavidin-bead binding assay.....	26
2.6 Design of gRNA to target GAPDH mRNA.....	27
2.7 Strategy for designing dCas9 and gRNA constructs for expression <i>in vivo</i>	28
2.8 Testing PAMmer transfection efficiency.....	31
2.9 <i>In vitro</i> GAPDH mRNA pull down with dCas9.....	32
2.10 <i>In vivo</i> GAPDH mRNA pull down with dCas9.....	34
2.11 RAP.....	35
2.12 RT-qPCR.....	35
2.13 Cross-linking HEK 293FT cells.....	36
2.13.1 HEK 293FT cell culturing.....	36
2.13.2 UV cross-linking.....	36
2.13.3 Formaldehyde cross-linking.....	37
2.13.4 Cell lysis.....	37
2.13.5 Protein analysis.....	38

CHAPTER 3: RESULTS

Test *in vitro* CRISPR/dCas9 system for GAPDH mRNA enrichment

3.1 Prepare <i>in vitro</i> CRISPR/dCas9 components for GAPDH mRNA enrichment.....	39
3.1.1 Overexpressing dCas9.....	39
3.1.2 Purification of dCas9.....	40
3.1.3 Biotinylation of dCas9.....	44
3.1.4 Design and produce gRNA and PAMmers targeting GAPDH mRNA.....	45
3.1.5 Comparing relative expression of GAPDH and Beta-actin in HEK 293FT cells.....	48
3.2 Enriching GAPDH mRNA with the <i>in vitro</i> dCas9 system.....	49
3.2.1 Enrich GAPDH mRNA by the <i>in vitro</i> dCas9 system with gRNA 2 and 3....	51
3.2.2 Testing ChIP wash conditions for the <i>in vitro</i> dCas9 system.....	53
3.2.3 Testing NP-40 wash conditions for the <i>in vitro</i> dCas9 system.....	58
3.2.4 Testing UV and NP-40 conditions for the <i>in vitro</i> dCas9 system.....	60
3.2.5 Examining RNA specificity of <i>in vitro</i> dCas9 system with NP-40 buffers....	62
3.2.6 Examining RNA specificity of <i>in vitro</i> dCas9 system with UV.....	64
3.3 Optimizing UV cross-linking CL in HEK 293 FT cells for RAP.....	66
3.3.1 Enrich GAPDH mRNA by RAP.....	68

CHAPTER 4: RESULTS

Test *in vivo* CRISPR/dCas9 system for GAPDH mRNA enrichment

4.1 Create and verify constructs for dCas9 expression <i>in vivo</i>	69
4.2 Testing binding efficiency and elution conditions for halo-dCas9.....	70
4.3 Create and verify constructs for gRNA expression <i>in vivo</i>	71
4.4 Determining PAMmer transfection efficiency in HEK 293FT cells.....	72
4.5 Optimizing formaldehyde CL conditions for <i>in vivo</i> dCas9 system.....	74
4.6 Enriching GAPDH mRNA with the <i>in vivo</i> dCas9 system.....	75

CHAPTER 5: CONCLUDING REMARKS AND FUTURE DIRECTIONS

5.1 The use of specific mild and potent detergents is incompatible with the <i>in vitro</i> dCas9 system.....	78
5.2 UV Stabilization of dCas9 complexes may enhance capturing of non-specific RNA...	79
5.3 Assessing the RNA-specificity of the <i>in vitro</i> dCas9 system.....	80
5.4 Assessing the gRNA and PAMmer specificity to GAPDH mRNA.....	82
5.5 Cleavage versus binding specificity of Cas9.....	86
5.6 The effects of mutations introduced into the <i>in vitro</i> dCas9 protein.....	87
5.7 <i>In vivo</i> dCas9 system fails to capture the majority of GAPDH mRNA.....	88
5.8 Future Directions.....	89
5.9 Concluding remarks.....	92

REFERENCES.....	93
-----------------	----

LIST OF FIGURES

Figure 1.1 Schematic of MS2-GFP and lambda N22-GFP reporter systems.....	8
Figure 1.2 Workflow of RAP-MS, ChIRP-MS and CHART.....	12
Figure 1.3 CRISPR/Cas immunity in bacteria and archaea.....	15
Figure 1.4 Mechanism of CRISPR/Cas9.....	18
Figure 1.5 Cas9 cleaves ssRNA targets.....	21
Figure 3.1 Overexpression of his-MBP tagged dCas9.....	40
Figure 3.2 Purification of his-tagged dCas9.....	42
Figure 3.3 dCas9 his-MBP tag removal.....	43
Figure 3.4 Biotinylation of dCas9.....	45
Figure 3.5 Confirmation of <i>in vitro</i> gRNA production for GAPDH mRNA enrichment...47	
Figure 3.6 Relative expression of GAPDH mRNA in HEK 293FT cells.....	49
Figure 3.7 GAPDH mRNA enrichment by <i>in vitro</i> dCas9 system using gRNA1 and PAMmer v1.1 or PAMmer v1.2.....	51
Figure 3.8 GAPDH mRNA enrichment by <i>in vitro</i> dCas9 system using gRNA2/PAMmer 2 and gRNA 3/PAMmer 3.....	53
Figure 3.9 GAPDH mRNA enrichment by <i>in vitro</i> dCas9 system using ChIP wash buffers.....	56
Figure 3.10 GAPDH mRNA enrichment by <i>in vitro</i> dCas9 system using ChIP wash buffers with no SDS.....	58
Figure 3.11 GAPDH mRNA enrichment by <i>in vitro</i> dCas9 system using NP-40 buffer..59	
Figure 3.12 GAPDH mRNA enrichment by <i>in vitro</i> dCas9 system using UV.....	62
Figure 3.13 NP-40 buffers: GAPDH or β -actin mRNA enrichment by <i>in vitro</i> dCas9 system using U2 snRNA as a reference.....	64
Figure 3.14 UV cross-linking: GAPDH or β -actin mRNA enrichment by <i>in vitro</i> dCas9 system using U2 snRNA as a reference.....	66
Figure 3.15 Optimizing UV cross-linking conditions in HEK 293FT cells.....	67
Figure 3.16 GAPDH mRNA enrichment by RAP.....	68
Figure 4.1 Verifying halo-tagged dCas9 expression in HEK 293FT cells.....	70
Figure 4.2 Optimizing elution conditions for halo-dCas9 and the Halo-Link™ resin.....	71
Figure 4.3 Verifying gRNA expression in HEK 293FT cells.....	72
Figure 4.4 Testing PAMmer transfection efficiency in HEK 293FT cells.....	74
Figure 4.5 Optimizing formaldehyde cross-linking conditions in HEK 293FT cells.....	75
Figure 4.6 GAPDH mRNA enrichment by <i>in vivo</i> dCas9 system.....	77
Figure 5.1 Alignment of 5 bp seed region of gRNA 1 and β -actin mRNA, GAPDH mRNA and U2 snRNA to determine sequence similarity.....	84
Figure 5.2 Alignment of 5 bp seed region of gRNA 2 and β -actin mRNA, GAPDH mRNA and U2 snRNA to determine sequence similarity.....	84
Figure 5.3 Alignment of 5 bp seed region of gRNA 3 and β -actin mRNA, GAPDH mRNA and U2 snRNA to determine sequence similarity.....	85

LIST OF TABLES

Table 2.1 gBlock Sequences used for GAPDH gRNA synthesis.....	28
Table 2.2 Primers used for designing dCas9 constructs.....	30
Table 2.3 Primers used for designing gRNA constructs.....	31
Table 2.4 Sequence of PAMmer 1 used to test transfection efficiency.....	32
Table 2.5 Primers used to detect GAPDH and β -actin mRNA by RT-qPCR.....	36
Table 3.1 gRNA sequences for targeting GAPDH mRNA.....	46
Table 3.2 PAMmer sequences for targeting GAPDH mRNA.....	48

LIST OF ABBREVIATIONS AND SYMBOLS

°C	Degrees Celsius
6-FAM	5' 6-Carbofluorescein tag
4-SU	4-thiouridine
6-SG	6-thioguanosine
A	Alanine
AML	Acute myeloid leukemia
β-actin	Beta-actin
bp	Base pairs
C	Cysteine
CARP	CRISPR-assisted proteomic method
cDNA	Complementary DNA
CHART	Capture hybridization analysis of RNA targets
ChIP	Chromatin immunoprecipitation
ChIRP-MS	Comprehensive identification of RBPs by mass spectrometry
CLIP	Cross-linking immunoprecipitation
CRISPR	Clustered regularly interspaced short palindromic repeats
Cas	CRISPR- associated
crRNA	CRISPR RNA
CTD	C-terminal domain
CT	Cycle threshold
CL	Cross-linking
D	Aspartic acid
dCas9	Deactivated Cas9
DMEM	Dulbecco's Modified Eagle Medium
dsDNA	Double-stranded DNA
DTT	Dithiothreitol
<i>E. Coli</i>	<i>Escherichia Coli</i>
EDTA	Ethylenediaminetetraacetic acid
FBS	Fetal bovine serum
FITC	Fluorescein isothiocyanate
iCLIP	Individual nucleotide resolution CLIP
IP	Immunoprecipitation
GAPDH	Glyceraldehyde 3-phosphate dehydrogenase
GFP	Green fluorescent protein
gRNA	guide RNA
HEK	Human embryonic kidney
H	Histidine
IRES	Internal ribosome entry sequences
IPTG	Isopropyl β-D-1-thiogalactopyranoside
Kd	Dissociation constant
kDa	Kilo-Dalton
LC	Liquid chromatography

LDS	Lithium dodecyl sulfate
MBP	Maltose Binding Protein tag
MS	Mass spectrometry
mRNA	Messenger RNA
O.D. ₆₀₀	Optical density (600 nm)
PAGE	Polyacrylamide gel electrophoresis
PAM	Protospacer-adjacent motifs
PAMmers	PAM presenting oligonucleotides
PAR-CLIP	Photoactivatable ribonucleoside-enhanced
PEI	Polyethylenimine
PFS	Protospacer flanking sequence
PBS	Phosphate buffered saline
RBP	RNA-binding proteins
RAP-MS	RNA antisense purification with MS
REC	Recognition lobe
RT-qPCR	Real-time quantitative polymerase chain reaction
S	Serine
SDS	Sodium dodecyl sulfate
SEM	Standard error mean
Seq	Sequencing
sgRNA	crRNA-tracrRNA complex
snRNA	Small nuclear RNA
snRNP	Small nuclear ribonucleoprotein
SR proteins	Serine/arginine-rich family of phosphoproteins
ssRNA	Single-stranded RNA
<i>S. Pyogenes</i>	<i>Streptococcus Pyogenes</i>
TEV	Tobacco Etch Virus
UV	Ultraviolet
UTR	Untranslated regions

DECLARATION OF ACADEMIC ACHEIVEMENT

This thesis was mainly completed by Piyanka Sivarajah, with the following contributions from members of the Yu Lu lab.

- i. Sansi Xing performed RAP and generated the real-time quantitative polymerase chain reaction (RT-qPCR) results in Figure 3.16.
- ii. Paige Darville O'Quinn helped clone the halo-tag for the *in vivo* dCas9 expression construct.
- iii. Jane Wang helped establish UV and formaldehyde cross-linking conditions for human embryonic kidney 293FT (HEK 293FT) cells in Figure 3.15 and Figure 4.5.

CHAPTER 1

INTRODUCTION

1.1 RNA regulation

RNA is a highly diverse nucleic acid, often thought of as the centre of molecular biology. As a multi-functional molecule, RNA plays various, important roles in several cellular processes. For example, three types of RNA perform different but cooperative functions to direct protein synthesis. These are, messenger RNA (mRNA) that carries information from DNA in three letter ‘codes’ to specify an amino acid, transfer RNA that helps decipher the code in mRNA and ribosomal RNA that associates with proteins to form ribosomes, the key enzyme involved in protein translation (Lodish et al., 2000). RNA is also known to help regulate gene expression. Short non-coding RNA like micro-RNA and small-interfering RNA often act to inhibit gene expression by binding to a complementary transcript to target them for degradation (Morris & Mattick, 2014). Other types of ‘non-coding’ RNA include long non-coding RNA, thought to play a role in both gene inhibition or activation, and small nucleolar RNA involved ribosomal RNA processing (Rinn & Chang, 2012; Stepanov et al., 2015).

Due to its versatility and contribution to several important functions, RNA is highly regulated. The two key elements involved in RNA regulation are RNA-sequence modulators and RNA-binding proteins (RBPs). RNA-sequence modulators are non-coding sequences that regulate regions within the same RNA molecule. There are different flavors of RNA-sequence modulators, for example, those that act independently of accessory

proteins, like the hairpin or ‘higher-order’ elements. These regulators form intramolecular mRNA structures that can inhibit translation independently of RBPs when positioned in the 5’ untranslated regions (UTR) near the cap structure (Gebauer, Preiss, & Hentze, 2012). The internal ribosome entry sequences (IRES) are another class of modulators found in 5’UTR of mRNA that can permit translation in a 5’ cap-independent manner, typically during cell stress or apoptosis (Gebauer et al., 2012). Other types of RNA-sequence modulators are largely dependent on their interaction with RBPs. For example, splice sites work in concert with RBPs to direct inclusion or excision of exons and introns in mRNA (Wang et al., 2015).

RBPs often work by binding to the target RNA itself or to the RNA-sequence modulators. For example, the splicing factors like the serine/arginine-rich family of phosphoproteins (SR proteins), contain two distinct regions to simultaneously interact with RNA (RNA binding motif) and other protein components (the SR region; Shepard & Hertel, 2009). Specifically, when binding to its corresponding exon splicing enhancer, the SR protein may recruit the splicing machinery to the adjacent intron, thereby ‘enhancing’ the splicing event (Shepard & Hertel, 2009). Other RBP regulators include polyadenylation factors that initiate the production of the polyA tail, and 5’ capping proteins that help form the 5’ 7-methylguanylate cap for mRNA stability (Shyu, Willikinson, & van Hoof, 2008; Wang et al., 2015). There are also RBPs (polypyrimidine tract-binding protein complexes) which may initiate or repress translation when bound to well-defined structures within the 5’ or 3’ UTR or coding regions, thus potentially directing cytoplasmic gene expression (Harvey et al., 2018; King et al., 2014).

Abnormal RNA regulation leads to disruption of normal cellular functions and thus, is a hallmark in several diseases. For example, in certain subtypes of acute myeloid leukemia (AML), it has been shown dysregulated splicing of specific transcripts (for example, AML1-ETO) contributes to leukemogenesis (Yan et al., 2006). Moreover, mis-splicing of tau pre-mRNA has been consistently identified in several tauopathies and shown to contribute to isoform-specific impairment of normal, physiological tau function (Park, Ahn, & Gallo, 2016). In AML, tauopathies and several other diseases, though it is clear that RNA-sequence modulators and RBPs play a pivotal role in disease progression, these regulators have yet to be identified for the target RNA. Thus, there is an increasing need for highly efficient, specific tools that can identify these RNA modulators.

1.2 Identifying specific RNA-sequence modulators with cross-linking immunoprecipitation

To identify RNA-sequence modulators, there are many different technologies available. Notably, there is cross-linking immunoprecipitation (CLIP), which exposes cells or tissues to ultraviolet (UV) light in order to form a covalent bond between RNA and proteins that are in close contact (Yang et al., 2015; Stork & Zheng, 2016). When the cells are lysed, RNA-protein complexes are captured by immunoprecipitation (IP), where a specific antibody is used to target a known, unique protein interactor on the RNA of interest. Reverse transcription of the enriched RNA, followed by high-throughput sequencing of the complementary DNA (cDNA) library is used to identify the specific sequence the protein of interest is bound to (Yang et al., 2015).

Since its discovery, CLIP has been improved to address caveats from the original approach. For example, the use of UV cross-linking can introduce mutations at the cross-linked sites and potentially cDNA truncations. This creates issues when aligning the cDNA library with the reference genome and attempting to identify the specific regions that are cross-linked. In photoactivatable ribonucleoside-enhanced (PAR)-CLIP, photoactive ribonucleoside analogs (4-thiouridine (4-SU) and 6-thioguanosine (6-SG)) are introduced into cells in hopes they will be incorporated into the cell's RNA (Hafner et al., 2010). Although the overall procedure is nearly identical to CLIP, the use of photoactivatable ribonucleosides means mutations induced by cross-linking can be easily identified. For example, when using 4-SU, cross-linking results in thymidine to cytidine transition, whereas using 6-SG results in guanosine to adenosine mutations (Hafner et al., 2010). Thus, this approach increases resolution of the 'traditional' CLIP technique, since sequences with these mutations can be separated from background cellular RNA and can be more thoroughly matched to the reference genome.

Another caveat of CLIP is the potential truncation that occurs during reverse transcription, at the site where the protein of interest is bound. Individual nucleotide resolution CLIP (iCLIP) exploits this by using a 3' exonuclease to digest the enriched RNA up to the cross-linked region (König, et al., 2010). An adaptor is then ligated to this site and after reverse transcription and sequencing, the precise RBP binding site can be identified at nucleotide resolution (König, et al., 2010). Therefore, in comparison to the 'traditional' CLIP approach, there is greater precision to iCLIP. This is particularly useful for studying

RNA-sequence modulators that are RBP dependent, as they only occupy a small fraction of the RNA sequence.

1.3 Identifying specific RBPs with mass spectrometry based proteomics

The most efficient approach to identify specific RBPs would be mass spectrometry (MS) based proteomics, as it provides an accurate mass determination and therefore, characterization of a specific protein. There are two MS techniques for protein identification, they are the top-down or bottom-up approach. The “top-down” approach is typically used to identify an intact protein. It begins by purifying a given protein using a separation method, for example, 1D electrophoresis or reverse-phase liquid chromatography (LC) (Catherman, Skinner, & Kelleher, 2015). The protein is then introduced to the high performance MS for mass analysis and fragmentation. The fragmentation spectra that is generated is searched against the appropriate database to identify and characterize the proteoforms (Catherman et al., 2015). This approach has the potential to identify all existing modifications on the protein in one spectrum and also saves time, since protein digestion prior to MS can be eliminated. However, traditionally, this method has been limited to simple mixtures or individual proteins.

In the “bottom-up” proteomic approach, crude protein extract is digested by a protease into smaller peptides, which are then separated by LC (Mendelsohn, Howley, Isarel, Gray, & Thompson, 2004). The LC is coupled to a tandem mass spectrometer, so peptides can be mass analyzed and fragmented as they elute from the LC (Mendelsohn et al., 2004). The fragmentation or tandem mass spectra of peptide precursors are searched

against a genome-wide protein database to identify these peptides and their parent proteins. The smaller and more uniform peptides produced from the bottom-up approach are easier to analyze and also can be determined with higher accuracy than the top-down approach (Zhang, Fonslow, Shan, Baek, & Yates, 2013). Therefore, typically this method is favored.

Although MS provides a unique tool to identify specific proteins, this method relies on the starting protein extract. Specifically, to use MS to study proteins involved in RNA regulation, it is critical to isolate specific RBPs from the RNA of interest. Therefore, MS requires a method to isolate and enrich a RNA target with its RBPs.

1.3.1 Targeting specific RBPs with MS2 and lambda N22

Traditionally to enrich specific RNA-RBP complexes researchers resorted to the use of overexpression constructs. For example, Bertrand et al. (1998) developed a technique that uses the specific binding interaction between the bacteriophage capsid protein, MS2 and its binding sequence, a 19 nucleotide stem-loop structure found in viral RNA. Specifically, they designed two overexpression constructs, one with MS2 fused to a green fluorescent protein (GFP) tag and one for the mRNA transcript of interest fused to six MS2 binding sequences (to provide increase signal for multiple bound GFP) (**Figure 1.1 A**). When transfected into cells, they could specifically determine the location of or target the mRNA of interest within the cell. One of difficulties of this approach is inserting repeats of the long, highly repetitive stem-loop MS2 binding sequence into the RNA of interest. To overcome the problem, Daigle and Ellenberg (2007) developed a technique centered on the interaction between the protein lambda N22 and its binding sequence referred to as “box

b” (**Figure 1.1 B**). Box b is shorter than the MS2 binding sequence (15 nucleotides) and due to its high affinity to lambda N22, less ‘repeats’ of the stem loop structure are added to achieve a higher signal (Daigle & Ellenberg, 2007). Thus, reducing the amount of exogenous RNA within the RNA of interest. The MS2 and lambda N22 approaches have the potential to carry an alternative tag to capture the RNA of interest and therefore, its associated proteins.

Both techniques provide several advantages when attempting to target specific RNA and its RBPs. MS2 and lambda N22 have high affinity to their target binding sequences (Dissociation constant (K_d) values: 0.4 nM and 22 nM respectively), therefore providing a higher signal to noise ratio than previous attempts (Bann & Parent, 2012). Additionally, these methods can be applied to live cells which avoids concerns from fixation artifacts. However, there are drawbacks. Mainly, the use of overexpression constructs can result in toxicity and thus, protein degradation can occur. Also, this system is not an accurate physiological representation of normal RNA regulation. It is difficult to predict whether or not the addition of the binding sequences may alter RBP binding. Strategic placement of the tag requires detailed knowledge of the RNA of interest, which may not always be available depending on the target.

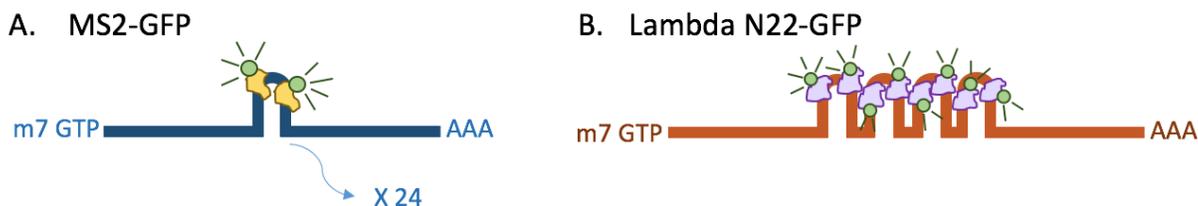


Figure 1.1 Schematic of MS2-GFP and lambda N22-GFP reporter systems. MS2-GFP (Bertrand et al., 1998, **A**) and lambda N22-GFP (Daigle and Ellenberg, 2007, **B**) are tagging systems developed to target a specific RNA of interest. The target RNA is tagged with a characteristic stem-loop binding sequence whereas the MS2 or lambda N22 are tagged with GFP. When MS2 or lambda N22 bind to their corresponding sequence, the RNA of interest can be localized or enriched.

1.3.2 Oligonucleotide probe approach: comprehensive identification of RBPs by MS (ChIRP-MS), RNA antisense purification with MS (RAP-MS) and capture hybridization analysis of RNA targets (CHART)

In 2015, two papers demonstrated physiologically relevant methods that targeted specific RNAs and their interacting RBPs; these methods are the comprehensive identification of RBPs by MS (ChIRP-MS) and RNA antisense purification with MS (RAP-MS) (Chu et al., 2015; McHugh et al., 2015). Both methods use cross-linking techniques (ChIRP-MS using formaldehyde and RAP-MS using UV) to preserve naturally occurring RNA- RBP complexes that are then targeted by DNA oligonucleotide probes (ChIRP with probes ranging 20 base pairs (bp) in length and RAP with probes 90 bp in length) (Chu et al., 2015; McHugh et al., 2015). Typically, the DNA probes are tagged with biotin so probe/RNA/RBP complexes can be captured by streptavidin beads. After eluting, the proteins candidates are then identified by MS (**Figure 1.2**). So far, ChIRP-MS and RAP-

MS have been successful in identifying key protein regulators associated with *Xist* RNA, a long non-coding RNA involved in X-chromosomal inactivation.

The challenge with oligonucleotide approaches is predicting whether or not probes will target a region occupied by RBPs, leading to inefficient enrichment. Both RAP-MS and ChIRP-MS obviate this problem by using a ‘tiling’ strategy, where probes are designed to cover the entire RNA sequence, so all potential hybridization spots can be covered. Capture hybridization analysis of RNA targets (CHART), another oligonucleotide based RNA-RBP capture technique, uses a different approach (Sexton, Maychna, & Simon, 2016). CHART is nearly identical to ChIRP, using biotinylated probes that range 20 bp in length and linking RNA-RBP complexes by formaldehyde cross-linking. However, the choice of targeting probes is based on a RNase H assay. Here, candidate probes are mixed with chromatin extract in the presence of RNase H, which cleaves RNA-DNA hybrids (Sexton et al., 2016). Oligonucleotides that hybridize to available RNA sites, therefore, will be cleaved by RNase H. Although this can be time-consuming, CHART ensures fewer probes are used and also produces a higher signal to noise ratio.

The current oligonucleotide techniques overcome several concerns from the over-expression construct approach. Mainly, the use of probes permits RNA regulation to be studied under its native, physiological context. Moreover, as in the case with *Xist* RNA, longer RNAs often require physical disruption, for example sonication, to solubilize the RNA. However, sonication shears the RNA into multiple fragments, making the use of tags, like the MS2 and the box b binding sequences, useless (Chu, Spitale, & Chang, 2015.). ChIRP-MS and RAP-MS both address this concern with their use of cross-linking, which

perseveres interactions, and the tiling strategy, which allows for the sheared fragments of the long RNA to be enriched (Chu, et al., 2015). Additionally, all of the oligonucleotide approaches use denaturing conditions during the probe hybridization and washing steps. This ensures removal of non-specifically bound RNA, thus reducing the background signal. There is also added confidence that the RBPs identified are indeed associated with the RNA of interest. The denaturing conditions also aid in targeting RNA that are closely associated with the nuclear matrix, for example, *Xist* RNA, which is difficult to extract under native conditions (Chu et al., 2015; McHugh et al., 2015).

Since their discovery, however, several key limitations have been identified with the oligonucleotide approaches. Notably, the use of the oligonucleotides means DNA fragments with sequence similarities and therefore, non-specific proteins can potentially be targeted (Chu et al., 2015). There is also the possibility that the oligonucleotides may bind to and enrich off-target RNA sequences. The specificity and affinity of nucleic acid hybridization are anti-correlated, thus an increase in affinity is accompanied with a decline in specificity (Demidov & Frank-Kamenetskii, 2004). This occurs because the large thermodynamic benefit from many correctly paired bases largely offsets the small penalty from a few mismatches. Thus, RAP-MS with its long oligonucleotide probes have high affinity for the target RNA sequence, due to increased binding (Machyna & Simon, 2018). Although, this also allows for more stringent washing conditions compared to ChIRP-MS and CHART, there is more potential for mismatch, and as a result there is a risk of off-target binding. Additionally, these longer probes are costlier to synthesize (Machyna & Simon, 2018). The shorter probes used in ChIRP-MS and CHART provide more specificity

and discrimination against off-target binding, although they do not bind with the same degree of affinity (Chu et al., 2015).

Additionally, RAP-MS uses high temperatures during hybridization, limiting it to being an *in vitro* approach. As an *in vitro* approach, the cells are only being studied in a specific instance and therefore, the identified proteins are only a snapshot of RNA regulation. This is not favourable as RNA regulation is a complex and dynamic process. With an *in vivo* approach, it would be possible to get an overall picture and therefore, a more accurate representation of the dynamics of RNA regulation. Furthermore, because of the temperature conditions, RAP-MS uses UV cross-linking, meaning only directly interacting RBPs will be linked to RNA. As mentioned, RNA regulation involves a complex network of interacting RBPs and RAP-MS fails to capture this by being limited to UV cross-linking. To overcome these limitations identified in RAP and ChIRP, we propose to develop a method targeting RNA and their associated proteins with a deactivated clustered regularly interspaced short palindromic repeats (CRISPR)/Cas9 system.

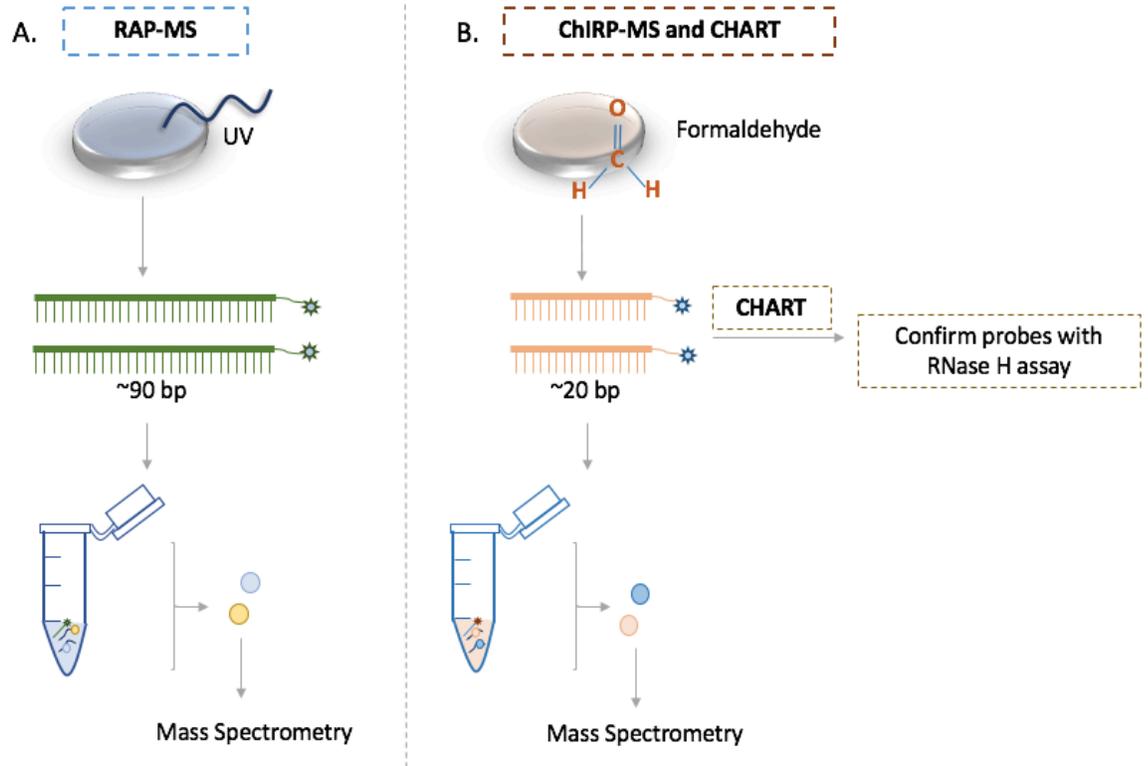


Figure 1.2 Workflow of RAP-MS, ChIRP-MS and CHART

Schematic of workflow for RAP-MS (McHugh et al., 2015 A), ChIRP-MS and CHART (B; Chu et al., 2015; Sexton et al., 2016). In RAP-MS, cells are UV cross-linked, in order to link any RBPs that are directly interacting with the RNA of interest. The target RNA is then captured with DNA oligonucleotide probes that are approximately 90 bp in length. The probes are biotinylated so any RNA-RBP complexes targeted can be enriched with streptavidin beads. Once the the target RNA-RBP complexes are eluted, the enriched RBPs are identified by MS. ChIRP-MS and CHART follow a nearly identical work-flow. The key differences are that the biotinylated DNA probes are 20 bp in length and formaldehyde cross-linking is used. Formaldehyde links both RBPs directly interacting with the target RNA and also RBP-RBP complexes. In CHART, DNA probes are validated by a RNase H assay prior to use.

1.4 The CRISPR/Cas system

Bacteria and archaea have long acquired an adaptive immune response known as the CRISPR/CRISPR- associated (Cas) systems, to protect against pathogenic viruses and

plasmids. The CRISPR/Cas system is made up of *cas* genes that are organized into operons and CRISPR sequences, or the sequences that target foreign genetic material, that are interspaced with repeats (Jinek et al., 2012). The CRISPR/Cas immunity occurs in three stages. In the first or adaptation stage, the host recognizes, excises and incorporate short sequences of foreign genetic material (protospacers) into the host chromosome at the proximal end of the CRISPR array (Rath, Amlinger, Rath, & Lundgren, 2015). In the expression stage, the host expresses the *cas* genes and transcribes the CRISPR into a longer precursor CRISPR RNA (crRNA, Jinek et al., 2012). Cas proteins and additional factors assist in cleaving the crRNA to yield a shorter version that can pair with complementary invading nucleic acids (Rath et al., 2015; Jinek et al., 2012). In the last stage, also known as interference, the foreign genetic material is recognized and destroyed by Cas proteins and crRNA (**Figure 1.3**).

The Cas proteins are a highly variant group. The Cas1 and Cas2 proteins are typically universally shared among the CRISPR/Cas systems. Other Cas proteins are classified as either Type I, Type II or Type III systems (Rath et al., 2015). In the Type I system, the main nuclease is Cas3. It often works as a cascade complex with Cas1 and Cas2, so it can both target the foreign nucleic acid with crRNA, and in most cases also processes crRNA into its mature form (Rath et al., 2015). The Type II system comprises of Cas1, Cas2 and the signature protein Cas9. Cas9 is capable of participating in the adaptation phase, the processing of crRNA and in targeting foreign DNA with the assistance of crRNA and tracrRNA, a conserved stem-loop structure required for Cas9 targeting (Jinek et al., 2012). Type III systems are characterized by the protein Cas10,

though its function is not completely clear. It has been shown Cas10 can form Cms or Cmr complexes similar to the Cascade complex formed by Cas3 (Liu, Pan, Li, Peng, & She, 2018). It has also been shown Type III systems can target both DNA and/or RNA, though the exact mechanism is uncertain (Liu et al., 2018). The Type I and Type III systems can be found in bacteria and archaea while the Type II systems are found exclusively in bacteria (Rath et al., 2015). Of the Cas proteins, the structure, function and mechanism of Cas9 is most known. Cas9 is also of particular interest because it is thought to be solely responsible for crRNA mediated silencing of foreign double-stranded DNA (dsDNA; Jinek et al., 2012).

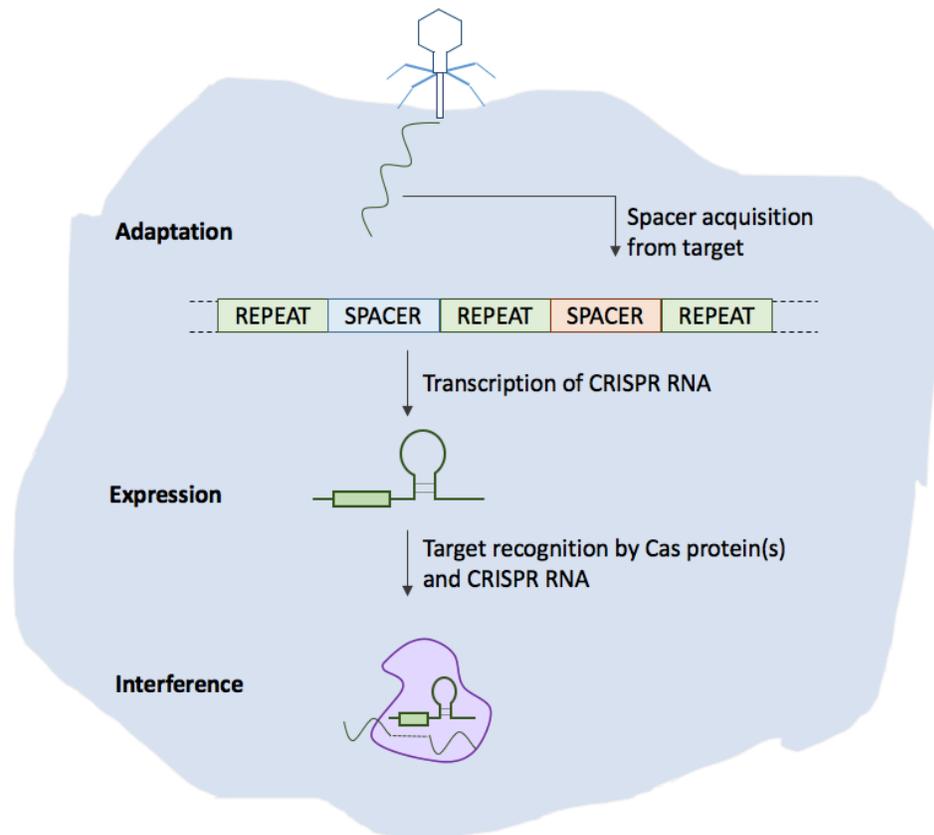


Figure 1.3 *CRISPR/Cas immunity in bacteria and archaea*

CRISPR/Cas immunity occurs in three stages. During adaptation, short fragments of foreign genetic material (protospacers) are acquired into the host's CRISPR array. During expression, the crRNA is transcribed by the host and processed by enzymatic cleavage into mature crRNA. Finally, during interference, the crRNA and Cas protein work together to recognize and cleave the foreign genetic material.

1.4.1 Structure of Cas9

In the inactive state, structures of Cas9 consists of two lobes: the alpha helical recognition (REC) lobe and the nuclease lobe, which contains the conserved HNH and the split RuvC nuclease domain as well as a variable C-terminal domain (CTD) (Jiang & Doudna, 2017). The two lobes are connected by two linker bridges: an arginine-rich bridge helix and a disordered linker (Jiang & Doudna, 2017). The REC lobe contains three helical domains

(Hel-I, Hel-II, Hel-III) and is structurally unique when compared to known proteins. It is responsible for binding to the tracrRNA portion of the crRNA-tracrRNA complex (sgRNA). The HNH domain of Cas9 closely resembles that of other HNH endonucleases. It is responsible for cleaving the dsDNA strand that is complementary to the crRNA, also known as the ‘target’ strand (Nishimasu et al., 2014). In contrast, the RuvC domain is responsible for cleaving the dsDNA strand that is opposite to the complementary strand, or the ‘non-target’ strand (Nishimasu et al., 2014). It has been demonstrated that this domain shares structural similarity to the retroviral integrase superfamily members that is characterized by a RNase H fold (Jiang & Doudna, 2017). The CTD is unique to Cas9, containing protospacer-adjacent motifs (PAM) interacting sites for PAM surveying. It is important to note the CTD is largely disordered in the ‘inactive’ Cas9, so that Cas9 is incapable of recognizing DNA when it is not bound to its sgRNA (Jiang & Doudna, 2017). Therefore, upon binding to sgRNA, Cas9 undergoes a conformational shift from the inactive state to a DNA recognition competent state. This has been demonstrated in previous studies where Cas9’s structure was analyzed with and without its sgRNA (Jiang & Doudna, 2017).

1.4.2 CRISPR/Cas9 mechanism

Although, as mentioned, Cas9 is multi-functional enzyme, its main role is as a DNA nuclease that targets and cleaves dsDNA with the assistance sgRNA. Specifically, once bound to the sgRNA, Cas9 begins to survey DNA. However, in order to survey and target dsDNA, Cas9 requires complementary base-pairing between the crRNA portion of sgRNA

and the target DNA and the presence of a PAM sequence adjacent to target region (Jiang & Doudna, 2017). The PAM exists on the non-target DNA strand and is only found in foreign DNA, distinguishing the target from the host. PAM sequences vary between different bacterial species (Rath et al., 2015). It has been shown that PAM recognition proceeds assessing complementarity between sgRNA and the target dsDNA region (Jiang & Doudna, 2017).

Once a PAM has been recognized, it initiates helicase activity at the PAM adjacent site. The sgRNA then forms a RNA-DNA hybrid with the DNA target to determine complementarity. For cleavage to occur, perfect complementarity is required at the seed region of the sgRNA (8-12 bases within crRNA region; Jiang & Doudna, 2017). Mismatches in non-seed regions are much more tolerated (Wu, Kriz, & Sharp, 2015). Upon PAM recognition and proper RNA-DNA duplex formation, Cas9 is activated for cleavage (Jiang & Doudna, 2017). As previously indicated, the RuvC and HNH domain are responsible for strand cleavage. Each domain cleaves the strand three base pairs upstream of the PAM sequence, generating a blunt end double-stranded break.

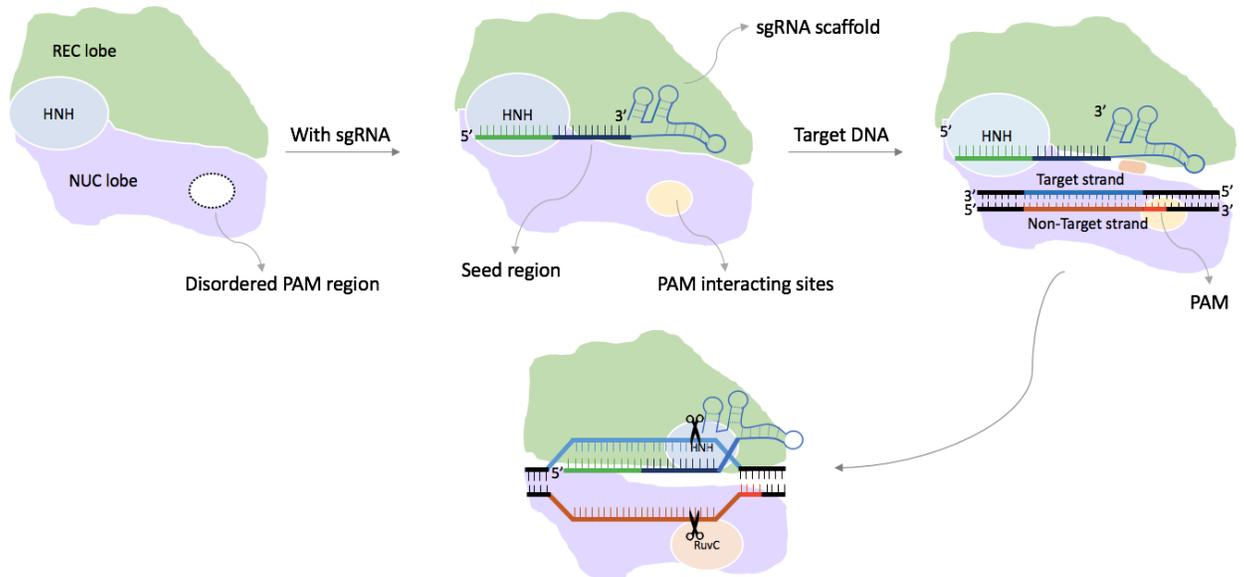


Figure 1.4 Mechanism of CRISPR/Cas9

Cas9 is a Type II variant of the CRISPR/Cas system. Cas9 exists in two states, the inactive state where the PAM interacting site is largely disordered and the active state or target recognition mode, where Cas9 upon binding to its sgRNA, undergoes conformational shift so that its PAM interacting sites are pre-structured for PAM recognition (Jiang & Doudna, 2017). Cas9 can then begin surveying DNA. It first recognizes PAM sequences, typically on the non-complementary DNA strand, adjacent to the region of complementarity. Once the PAM is recognized, the Cas9 initiates duplex melting in the PAM adjacent region. If the sgRNA matches, cleavage occurs.

1.4.3 The CRISPR/Cas9 system and its interaction with RNA

Traditionally, Cas9 has largely been studied in the context of dsDNA, while its interaction and activity with RNA was less explored. Over the years, however, studies have hinted at a potential interaction between Cas9 and RNA. For example, it was shown PAMs can be designed and delivered to prompt Cas9 to target and cleave single-stranded DNA, suggesting the potential to also prompt interaction with single-stranded RNA (ssRNA) (Sternberg, Redding, Jinek, Greene, & Doudna, 2014). Additionally, the catalytic domain

of Cas9, the HNH domain, closely resembles the catalytic domain of enzymes cleaving RNA substrates (Pommer et al., 2001). Therefore, in 2014, O’Connell et al. set to conclusively determine whether Cas9 was capable of specifically targeting and cleaving RNA.

To test if Cas9 could cleave RNA targets, O’Connell et al. (2014) set up an *in vitro* cleavage experiment with a panel of RNA and DNA targets. They found complementary, DNA-based PAM presenting oligonucleotides (PAMmers), could activate *Streptococcus pyogenes* (*S. Pyogenes*) Cas9 to cleave specific ssRNA, requiring either a 5’-NGG-3’ or 5’-GG-3’ PAM sequence, when supplemented with a complementary guide RNA (gRNA, with structure similar to sgRNA, **Figure 1.5**). Interestingly, ssRNA coupled with a complementary RNA-based PAMmer and double-stranded RNA could not be cleaved by Cas9. The authors hypothesized this may result from Cas9 recognizing the local helical structure and/or deoxyribose nucleotides within PAMs.

In order to determine whether Cas9 mediated ssRNA targeting was sequence specific, the authors prepared three distinct gRNAs, $\lambda 1$, $\lambda 2$, $\lambda 3$, and showed their corresponding ssRNA targets were only cleaved when supplemented with the complementary gRNA. Since no cross-reactivity was detected, the authors predicted ssRNA-gRNA base-pairing is critical for cleavage. Next, they tested whether nuclease activation by PAMmers requires base-pairing to the consequent ssRNA by performing an *in vitro* cleavage assay with matched and mismatched PAMmers. They demonstrated that Cas9 cleaves ssRNA regardless if the PAMmer is matched or not. Further, they demonstrated when ssRNA and a mismatched-PAMmer was incubated with dsDNA lacking a PAM sequence, only ssRNA

was cleaved by Cas9. This suggests that Cas9 can be manipulated to solely target ssRNA while eliminating possible targeting of dsDNA (O'Connell et al., 2014).

Finally, to further explore the binding dynamics of Cas9 to ssRNA, O'Connell et al. generated a catalytically deactivated Cas9 or 'dCas9', and tested whether dCas9 could target and enrich tag-less transcripts in HeLa cells. The dCas9 enzyme has two mutations: aspartic acid is substituted for alanine at position 10 (D10A), which inactivates the RuvC domain, and histidine is substituted for alanine at position 840 (H840A), which inactivates the HNH domain (Dominguez, Lim, & Qi, 2016). Therefore, although this enzyme lacks its nuclease activity, it still retains its targeting and surveying capabilities. They found, when supplemented with the biotinylated-dCas9, complementary sgRNAs and PAMmers, target mRNA was enriched compared to the control and dCas9 alone from total RNA. Compared to existing methods for enriching RNA, at that time, the authors showed this method works well in physiological conditions and does not require prior cross-linking. It has been suggested that this technique can be extended to total cell lysate as well (O'Connell et al., 2014).

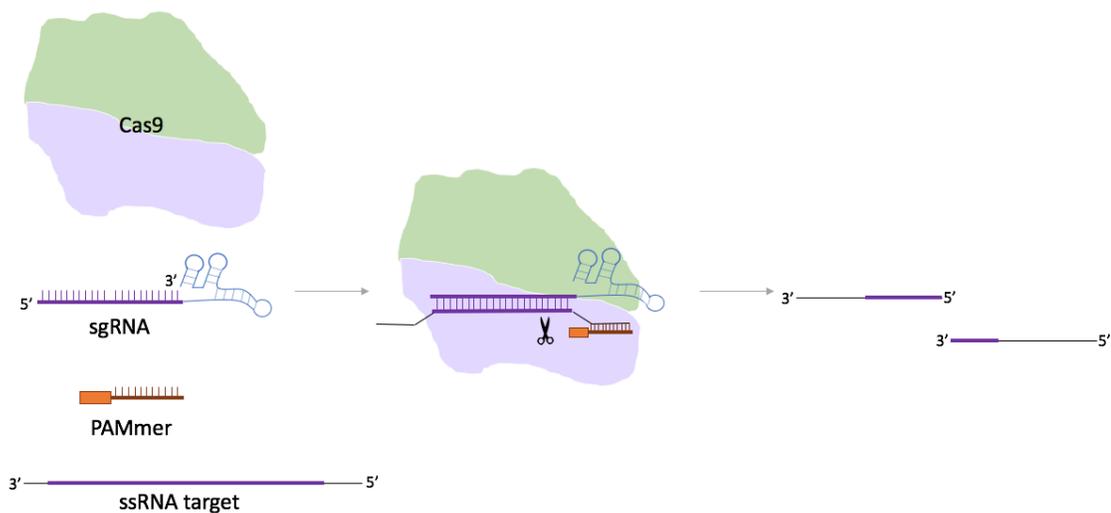


Figure 1.5 *Cas9 cleaves ssRNA targets*

Cas9 can target and cleave ssRNA targets when supplemented with a sgRNA complementary to the ssRNA target and a short PAM presenting DNA oligonucleotide (PAMmer).

1.5 Thesis rationale

This RNA enrichment tool proposed by O’Connell et al. (2014) has several advantages for targeting RNA and its associated RBPs. For one, this methodology solely relies on the targeting ability of Cas9 and the complementary base-pairing of the gRNA and PAMmers to enrich ssRNA, thereby retaining physiological relevance as no exogenous delivery of modified RNA is required. Furthermore, the CRISPR/dCas9 system is stringent in how it interacts with nucleic acids. Specifically, even when gRNAs are present, dCas9 will not cleave single-stranded RNA unless a PAMmer is provided (O’Connell et al., 2014). The PAMmer is unique and can be designed specifically for each RNA sequence. It can also be designed in a way so Cas9 is unlikely to interact with the corresponding dsDNA (O’Connell et al., 2014). This can allow for a more thorough analysis of proteins coupled

to a RNA of interest compared to RAP-MS or ChIRP-MS. Notably, this technique also has the versatility to employ *in vitro* and *in vivo* approaches, for example, incubating dCas9, sgRNA and PAMmers with cross-linked cell lysate or transfecting dCas9, sgRNAs and PAMmers directly or through a construct into live cells, respectively. Additionally, formaldehyde cross-linking can be used with this approach. This means complex protein networks can be linked and captured. This provides the opportunity to explore the overall picture of RNA regulation and therefore, a potentially more effective RNA-RBP enrichment process compared to the available approaches.

1.5.1 Thesis objectives and goal

Despite its many potential advantages, the CRISPR/dCas9 system has yet to be tested quantitatively as a tool for RNA-RBP enrichment. For example, although cleavage appears to only occur on the target RNA, it has not been assessed whether, and the degree to which, dCas9 binds to and enriches non-specific RNA targets. Therefore, the goal of this thesis was to develop and compare the CRISPR/dCas9 system for enriching RNA, and therefore its RBPs, to one of the current standards for RNA-RBP enrichment, RAP-MS. The first objective of this thesis was to develop two strategies using the CRISPR/dCas9 system. The first approach is the '*in vitro*' approach, where individual CRISPR components, biotinylated-dCas9, gRNAs (sgRNA), and PAMmers, were incubated with total RNA under varying conditions. Any biotinylated-dCas9/gRNA/PAMmer/RNA complexes formed were captured with streptavidin beads. For the second approach or the '*in vivo*' approach, we developed expression constructs for tagged-dCas9 and gRNAs. The

constructs were transfected along with PAMmers into cells. Upon cross-linking cells, we hoped to link any tagged dCas9/sgRNAs/PAMmers/RNA/RBP complexes formed and capture them upon lysis. The second objective was to compare and determine the efficiency and specificity between the three approaches by assessing the RNA-fold enrichment using real-time quantitative polymerase chain reaction (RT-qPCR). Glyceraldehyde 3-phosphate dehydrogenase (GAPDH) mRNA was used as a target, due to its high abundance in cells and readily available information available about its different isoforms for gRNA design.

Comparing the fold enrichment of the target and non-target RNA sequences would determine whether the CRISPR/Cas9 system can generate specific protein lists by MS for a single RNA of interest. It would also provide insight on the off-target binding nature of Cas9 to RNA and therefore, help determine whether dCas9 is a suitable tool for RNA-RBP enrichment.

GOAL

To develop, determine the specificity and compare the *in vitro* and *in vivo* CRISPR/dCas9 system for enriching RNA and its associated RBPs.

CHAPTER 2

MATERIALS AND METHODS

2.1 dCas9 plasmid amplification and sequencing

The plasmid for dCas9 expression was obtained from Addgene (Addgene code: 60815), as used by the O'Connell et al (2014). The agar stab was streaked on a LB agar plate supplemented with ampicillin (100 µg/mL) and grown at 37 °C overnight. A single colony was selected and grown overnight at 37 °C in 200 mL LB media with ampicillin (100 µg/mL) and shaking at 200 rpm. The cells were harvested by centrifugation (4000 x g for 10 minutes) and the plasmid was purified using a PureLink™ HiPure Plasmid Filter Maxiprep Kit (Invitrogen). The purified dCas9 plasmid was confirmed by sanger sequencing (Mobix, McMaster).

2.2 Expression of dCas9

dCas9 expression plasmid was transformed into *Escherichia Coli* (*E. Coli*) B121 DE3 Rosetta (Novagen) as described by Novagen. The transformed *E. Coli* B121 DE3 Rosetta cells were plated on a LB agar plate with ampicillin (100 µg/mL) and grown at 37 °C overnight. Multiple colonies were selected and grown in 1L of 2xYT media with ampicillin (100 µg/mL) at 37 °C and shaking at 250 rpm to O.D.₆₀₀ of 0.7. Thereafter, it was grown at 18°C for 16 hours following induction with 0.5 mM isopropyl β-D-1-thiogalactopyranoside (IPTG). Cells were harvested by centrifugation (4000 x g for 10 minutes) and stored at -80 °C (Jinek et al., 2012). Protein expression was confirmed by

running pre and post induction samples on a sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) stained with coomassie-blue and a western blot probed with a Cas9-specific antibody as previously described (Gallgher & Sass, 2001; Mahmood & Yang 2012).

2.3 Purification of dCas9

Briefly, *E. Coli* BL21 DE3 Rosetta induced-cells were re-suspended in 50 mM Tris-HCl pH 8.0 and 5 mM dithiothreitol (DTT), supplemented with cOmplete™ Protease Inhibitor Cocktail (Sigma). Cells were disrupted using a homogenizer and spun at 15000 rpm for 30 minutes to obtain the cytoplasmic fraction. Supernatant from the spin was applied to a HisTrap column and washed with Buffer A (20 mM Tris-HCl pH 8.0, 0.5 M NaCl, 1.4 mM Beta-mercaptoethanol and 5% glycerol). Loaded protein was eluted with a linear gradient (0-100%) of Buffer B (20 mM Tris-HCl pH 8.0, 0.5 M NaCl, 1.4 mM Beta-mercaptoethanol and 5% glycerol, 0.3 M imidazole). Purification of dCas9 was confirmed by performing a SDS-PAGE stained with coomassie-blue and a western blot probed with a Cas9 antibody as previously described (Gallgher & Sass, 2001; Mahmood & Yang, 2012).

2.4 His-Maltose Binding Protein (MBP) tag removal of dCas9 using Tobacco Etch Virus (TEV) protease

Fractions of purified his-MBP tagged dCas9 were collected, pooled, concentrated using 100 kilo-dalton (kDa) Amicon filters (EMD Millipore) and dialyzed using 3.5 kDa Snakeskin dialysis membrane (ThermoScientific) to remove imidazole from the sample

(Dialysis buffer: 20mM Tris-HCl pH 8.0, 250 mM NaCl). His-MBP tag on dCas9 was cleaved with TEV protease (GeneScript) as suggested by the manufacturer. Cleaved dCas9 was applied to a HisTrap column and purified again with an imidazole step gradient (5, 10, 15, 20, 25, 100%) of Buffer B (Refer to ‘Purification of dCas9’). Purified-cleaved dCas9 was confirmed on a coomassie-blue stained SDS-PAGE and a western blot probed with a Cas9-specific antibody as previously described (Gallgher & Sass, 2001; Mahmood & Yang, 2012).

2.5 Biotinylation of dCas9 and streptavidin-bead binding assay

Purified-cleaved dCas9 fractions were collected, pooled and dialyzed using 3.5 kDa Snakeskin dialysis membrane (Dialysis buffer: 20mM Tris-HCl pH 8.0, 250 mM NaCl) to remove reducing agents that may interfere with the biotinylation reaction. dCas9 was biotinylated with EZ-Link™ Iodoacetyl-PEG2-Biotin as suggested by the manufacturer (ThermoScientific). Tagging was verified by western blot probed with a streptavidin antibody as previously described (Mahmood & Yang, 2012).

Binding efficiency of biotinylated-dCas9 was tested with a streptavidin-bead binding assay. Briefly, biotinylated and non-biotinylated dCas9 (8.5 pmol) were incubated with 25 μ L of Dynabeads MyOne Streptavidin C1 beads (ThermoScientific) and reaction buffer (20mM Tris-HCl pH 7.5, 75 mM KCl, 5mM MgCl₂, 1mM DTT, 5% glycerol) for two hours at 4°C. Following hybridization, beads were washed six times with 300 μ L of wash buffer (20mM Tris-HCl pH 7.5, 150 mM NaCl, 5mM MgCl₂, 1mM DTT, 5% glycerol, 0.1% TritonX-100) and then eluted with 4X laemmli loading buffer. Flow-

through, wash and elution fractions were collected and run on a western blot probed with either a Cas9-specific or streptavidin antibody to confirm binding (Mahmood & Yang, 2012).

2.6 Design of gRNA to target GAPDH mRNA

GAPDH gRNA sequences were ordered as gBlocks from IDT and synthesized into RNA using the HiScribe™ T7 High Yield RNA synthesis kit (NEB). GAPDH gRNA was designed using suggestions from O’Connell et al (2014), with the addition of a “GGG” sequence after the T7 promoter as suggested by the HiScribe™ T7 High Yield RNA synthesis kit (NEB). All gRNAs contain a unique 20-22 bp sequence complementary to GAPDH mRNA and a conserved 95 bp hairpin structure called “tracrRNA”, required for Cas9 targeting (**Table 2.1**). Sequences were inputted into the BLAST server to predict and minimize off target events. gRNA synthesis was confirmed by a urea-PAGE as previously described (Summer & Droge, 2009).

Table 2.1 *gBlock Sequences used for GAPDH gRNA synthesis*

Name	Sequence (5' → 3')
GAPDH gRNA 1	*TAATACGACTCACTATA*GGGGCAGAGATGATGACCCTGTT TAAGAGCTATGCTGGAAACAGCATAGCAAGTTTAAATAAGG CTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGC TTTTTTT
GAPDH gRNA 2	*TAATACGACTCACTATA*GGGGGCCAAAGTTGTCATGGATG ACGTTTAAGAGCTATGCTGGAAACAGCATAGCAAGTTTAAA TAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGT CGGTGCTTTTTTTT
GAPDH gRNA 3	*TAATACGACTCACTATA*GGGTGATGTCATCATATTTGGCA GGGTTTAAGAGCTATGCTGGAAACAGCATAGCAAGTTTAAA TAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGT CGGTGCTTTTTTTT
T7 Promoter- *TAATACGACTCACTATA* <u>GGG</u> start site addition	

2.7 Strategy for designing dCas9 and gRNA constructs for expression *in vivo*

For *in vivo* dCas9 expression, a construct was created using the plko.1 hygro backbone (Addgene code: 24150) due to its high copy number and availability of different promoters. dCas9 was amplified from a plasmid containing a variant of dCas9 (Addgene code: 83306) using the primers which contain BamHI and KpnI restriction enzyme sites (in addition to other restriction enzyme sites) for digestion and ligation into plko.1 hygro. Similarly, puromycin was amplified from a plasmid which contains a T2A specific promoter (Addgene code: 62987) using primers which contain XbaI and KpnI restriction enzyme sites for digestion and ligation. Finally, halo tag was amplified (Addgene #: 29644) using the primers which contain BamHI and NheI restriction enzyme sites for digestion and ligation and a NLS signal for localization into the nucleus. Digested and PCR amplicon products were separated on a 1% agarose gel, then purified using the E.Z.N.A Gel

Extraction Kit (Omega). Fragments were ligated using a T4 DNA ligase (NEB) as described by the manufacturer. Plasmid was verified by sanger sequencing (Mobix, McMaster). Primers used for creating the dCas9 construct are listed in **Table 2.2**.

A stable cell line expressing the halo-dCas9 construct was created as follows: first, human embryonic kidney (HEK) 293FT cells were grown as described in “2.13.1 HEK 293FT cell culturing”. At 80-90% confluency, dCas9 construct (2 µg) was transfected into cells with polyethylenimine (PEI, Sigma) as previously described (Lango, Kavran, Min-Sung, & Leahy, 2013). At 48 hours’ post-transfection, the media of cells are replaced (Refer to 2.13.1 HEK 293FT cell culturing), with puromycin (0.5 µg/mL) added, and cells are monitored for death. Cells are continued to be monitored and once they begin developing ‘colonies’, cells are carefully washed once with phosphate buffered saline (PBS, Sigma), dissociated with Trypsin (37°C, 5% CO₂, 2 minutes, Wisent Bioproducts), sedimented by centrifugation at 1,000 x g for two minutes and re-plated with a split ratio of one in two once every two days. Upon subsequent re-plating, a fraction of the cells (22 million) were harvested, i.e. after sedimentation, cells were flash freezed and stored at -80°C. Pellets were lysed with 8M Urea and 100 mM ammonium bicarbonate and vortexing (10 seconds on, 20 seconds on ice for 5 rounds). Lysate (20-40 µg) was used to confirm dCas9 expression by a western blot probed with a Cas9-specific antibody as previously described (Mahmood & Yang, 2012).

Table 2.2 *Primers used for designing dCas9 constructs*

Name	Sequence (5' → 3')
MCS1-dCas9-F	ATGCGGATCCACCGGTGCTAGCGACAAGAAGTACAGCATCGG
dCas9-MCS2-R	ATGCGGTACCGTTTAAACTCTAGAGTCGCCTCCCAGCTGAGACA
XbaI-T2A-Puro-F	ATGCTCTAGAGGCAGTGGAGAGGGCAGAGG
Puro-Stop-KpnI-R	ATGCGGTACCTCAGGCACCGGGCTTGCGGG
BamHI-Kozac-HaloTag-F	ATGCGGATCCACCATGGCAGAAATCGGTACTGG
HaloTag-NLS-NheI-R	ATGCGCTAGCACCGACCTTCCGCTTCTTCTTTGGGTTATCGCTCTG AAAGTACA

For gRNA expression, the gRNAs were cloned into a lentiviral U6-driven expression vector by amplifying the insertions using a common reverse primer and a unique forward primers containing the unique protospacer sequence, as previously described (Liu et al., 2017). Briefly, the forward primers were mixed with equal amount of reverse primer to PCR amplify gRNA fragments using pSLQ1651 vector (Addgene code: 100549) as the template. The PCR amplicon and the gRNA vector containing a mCherry reporter gene were digested by restriction enzymes BstXI and XhoI for three hours. The digestion PCR amplicon and gRNA vector were separated on a 1% agarose gel, then purified using the E.Z.N.A Gel Extraction Kit. Purified PCR amplicon was then ligated to the gRNA vector

using a T4 DNA ligase (NEB) as described by the manufacturer. Insertion of gRNA was validated by sanger sequencing (Mobix, McMaster). Primers used for gRNA constructs are listed in **Table 2.3**.

Table 2.3 *Primers used for designing gRNA constructs*

Name	Sequence (5'→ 3')
GAPDH gRNA1-F	GGAGAACCACCTTGTTGGGGGCAGAGATGATGACCCTGTTTAA GAGCTATGCTGGAAACAGCA
GAPDH gRNA2-F	GGAGAACCACCTTGTTGGGCCAAAGTTGTCATGGATGACGTTT AAGAGCTATGCTGGAAACAGCA
GAPDH gRNA3-F	GGAGAACCACCTTGTTGGTGATGTCATCATATTTGGCAGGGTTT AAGAGCTATGCTGGAAACAGCA
Common- gRNA-R	CTAGTACTCGAGAAAAAAGCACCGACTCGGTGCCAC

2.8 Testing PAMmer transfection efficiency

PAMmer 1 (**Table 2.4**) containing both 2'O methyl modifications and a 5' 6-Carbofluorescein (6-FAM) tag was ordered from IDT. HEK 293FT cells were grown in 6-well culture plates (Greiner Bio-One) at 37°C in Dulbecco's Modified Eagle Medium (DMEM) (Sigma) supplemented with 10% fetal bovine serum (FBS) (Wisent Bioproducts) and 10 M L-glutamine (Sigma). At 60-80% confluency, PAMmer 1 was transfected into HEK 293FT cells using varying amounts (0 µL, 3 µL, 6 µL, 9 µL) of Lipofectamine ® RNAiMAX (ThermoScientific), as recommended by the manufacturer. At 48 hours' post-transfection, cells were washed once with PBS (Sigma), dissociated with Trypsin (37°C, 5% CO₂, 2 minutes, Wisent Bioproducts) and sedimented by centrifugation at 1,000 x g

for two minutes. After two additional washes with PBS, cells were re-suspended in 1 mL of PEF (1X PBS, 1mM ethylenediaminetetraacetic acid (EDTA), 2% FBS). Percentage of cells that were ‘live’ and exhibited fluorescence was determined by the MACSQuant ® analyzer, as described by the manufacturer.

Table 2.4 *Sequence of PAMmer 1 used to test transfection efficiency*

Name	Sequence
PAMmer 1	/56-FAM/mAT mGAmC CmCT mAGG mGGmC TmCC mCCmC CmUG mCAmAA

/56-FAM/ 5' 6-Carbofluorescein
m- Nucleotide before is 2'O methylated

2.9 *In vitro* GAPDH mRNA pull down with dCas9

HEK 293FT cells were prepared as described in “2.13.1 HEK 293FT cell culturing”. At 80-90% confluency, cells were scraped from culture plates and washed twice with PBS. Cells were then aliquoted into 1.5 mL microcentrifuge tubes in five million or 25 million cell batches. Cells were pelleted by centrifugation at 1,600 x g for four minutes and frozen at -80°C. Either total RNA was isolated from HEK 293FT cells using RNeasy Plus Mini Kit (Qiagen) or pellets were lysed with lysis buffer (20mM Tris-HCl pH 7.5, 75 mM KCl, 5mM MgCl₂, 1mM DTT, 5% glycerol, 1% NP-40, 0.1% SDS) and vortexing (10 seconds on, 20 seconds on ice for 5 rounds). dCas9–gRNA complexes were reconstituted before pull-down experiments by incubating a twofold molar excess of Cas9 with gRNA for 10 min at 37°C in reaction buffer (Refer to ‘Biotinylation of dCas9 and streptavidin-bead binding assay’). HEK 293FT total RNA (40 µg) or lysate (5x10⁶ cells) was added to

reaction buffer with 40U RNase inhibitor (Promega), PAMmer (5 μ M) and the biotin-dCas9 (50nM)–gRNA (25 nM) in a total volume of 100 μ L and incubated at 37°C for 1 hour. This mixture was then added to 25 μ L magnetic streptavidin beads (Dynabeads MyOne Streptavidin C1; ThermoScientific) pre-equilibrated in reaction buffer and agitated at 4°C for 2 h. Beads were then washed with either with 300 μ L of wash buffer (Refer to ‘Biotinylation of dCas9 and streptavidin-bead binding assay’) or with NP-40 Wash Buffer (20 mM Tris-HCl pH 7.5, 150 mM NaCl, 5mM MgCl₂, 1% NP-40 and 5% glycerol) for six times or once with Low Salt Immune Complex Wash Buffer (with or without 0.1% SDS, 1% Triton X-100, 2mM EDTA, 20mM Tris-HCl, pH 8.1, 150mM NaCl), once with High Salt Immune Complex Wash Buffer (with or without 0.1% SDS, 1% Triton X-100, 2mM EDTA, 20mM Tris-HCl, pH 8.1, 500mM NaCl, and once with LiCl Immune Complex Wash Buffer (0.25M LiCl, 1% IGEPAL-CA630, 1% deoxycholic acid(sodium salt), 1mM EDTA, 10mM Tris, pH 8.1). Input, flow-through and wash fractions were taken for analysis. RNA was eluted by re-suspending beads in lithium dodecyl sulfate (LDS) buffer (150 mM Tris-HCl pH 7.5, 2 mM EDTA, 8% LDS). To perform buffer exchange, flow-through, wash and elution fractions were purified with RNeasy Micro Kit (Qiagen).

Note: for “cross-linked samples”, the same procedure was followed except after the 1-hour incubation at 37°C, samples were irradiated with either 0 J, 0.05, 0.1, 0.4 J/cm² UV light at 254 nm in a GS Gene Linker UV Chamber (Bio-Rad).

2.10 *In vivo* GAPDH mRNA pull down with dCas9

A stable HEK 293FT cell line expressing halo-dCas9 was generated as described in ‘Strategy for designing dCas9 and gRNA constructs for expression *in vivo*’. A lentiviral transfection approach was used to express GAPDH gRNA constructs in the stable cell line as follows: first, HEK 293FT cells were grown as described in “2.13.1 HEK 293FT cell culturing” in 15 cm plates. Three plates were used per gRNA construct. At 80-90% confluency, media was aspirated and replaced with ‘incomplete’ media, i.e. only DMEM. A mixture of the following plasmids was created: pspAX2 (20.25 µg, Addgene code: 12260), pMD2.G (8.1 µg, Addgene code: 12259), gRNA construct (27 µg) and topped to 2.5 mL with Opti-MEM media (FisherScientific). PEI (1mg/mL) was diluted by adding 150 µL of PEI to 2.5 mL of Opti-MEM media and incubated at room temperature for five minutes. Diluted PEI is added to the diluted DNA and incubated at room temperature for 20 minutes. The entire transfection mix is added to the HEK 293FT cells. At 8 hours’ post-transfection, media was replaced with complete DMEM media (as described in 2.13.1 HEK 293FT cell culturing). At 48-72 hours post transfection, the virus was harvested by collecting the supernatant. The supernatant was spun at 1500 x g for 5 minutes to collect cells that may have accidentally been captured.

At 30% confluency, in a 6-well plate, halo-dCas9 expressing HEK 293FT cells were infected with the viral supernatant in a 1:1 ratio with complete DMEM media and polybrene (10 µg/mL). The process was repeated again at 24 hours’ post-infection. At 48 hours’ post initial infection, PAMmer was added to cells using Lipofectamine RNAiMAX, as described by the manufacturer. At 24 hours’ post-addition of the PAMmer, cells were washed once

with cold 1X PBS. Formaldehyde (Sigma) was added to the culture dishes to a final concentration of 1%. Culture plates were incubated at room temperature for 20 minutes with gentle agitation. Cross-linking was quenched by adding glycine (Sigma) to a final concentration of 0.2 M. Cells were scraped from culture plates and washed twice with PBS. Cells were pelleted by centrifugation at 1,600 x g for four minutes. Pellets were lysed and complexes formed with halo-dCas9 were captured with Halo-Link™ Resin (Promega) as described by the manufacturer. Samples were eluted by incubating the Halo-Link™ Resin with recommend amount of TEV Protease (GeneScript) overnight at 4 °C with agitation. Input, flow-through and elution fractions were reverse cross-linked for 6 hours at 65°C and then treated with Proteinase K (1mg/mL, ThermoScientific) for 1 hour at 55°C. To perform buffer exchange, flow-through, wash and elution fractions were purified with RNeasy Micro Kit.

2.11 RAP

Probes for targeting GAPDH mRNA and protocol for enriching GAPDH mRNA were followed as indicated by McHugh et al. (2015).

2.12 RT-qPCR

Samples from *in vitro* and *in vivo* dCas9 pull-down and RAP were analyzed by RT-qPCR. Briefly, samples underwent reverse-transcription using a iScript cDNA synthesis kit (Bio-Rad) as described by the manufacturer. Samples for qPCR were prepared using 3 µL of cDNA and SsoAdvanced™ Universal SYBR® Green Supermix (Bio-Rad). Beta-actin (β -

actin) or U2 was used as the reference gene. Relative gene expression was determined using the delta-delta Ct method in Bio-Rad's CFX96 software. Primer sequences to detect GAPDH and β -actin mRNA are listed in **Table 2.5**.

Table 2.5 Primers used to detect GAPDH and β -actin mRNA by RT-qPCR

Name	Sequence (5' -> 3')
GAPDH mRNA F	GATTGGTTCGTATTGGGCGC
GAPDH mRNA R	TTCCCGTTCTCAGCCTTGAC
B-actin mRNA F	TGAAGTGTGACGTGGACATC
B-actin mRNA R	GGAGGAGCAATGATCTTGAT
U2 snRNA F	TGGAGCAGGGAGATGGAATA
U2 snRNA R	CGTTCCTGGAGGTACTGCAA

2.13 Cross-linking HEK 293FT cells

2.13.1 HEK 293FT cell culturing

HEK 293FT cells were grown in either 10 cm or 15 cm cell culture plates (Greiner Bio-One) at 37°C in DMEM supplemented with 10% FBS and 10 M L-glutamine. Cells were routinely passaged by washing once with PBS (Sigma), dissociating with Trypsin (37°C, 5% CO₂, 2 minutes, Wisent Bioproducts), subsequent diluting with an equal amount of DMEM with FBS and L-glutamine and then sedimenting by centrifugation at 1,000 x g for two minutes, and re-plating approximately once every two days, with a split ratio of one in two.

2.13.2 UV cross-linking

HEK 293FT cells were grown to 80-90% confluency in cell culturing conditions. Cells were washed once with cold 1X PBS. Cell culture plates were placed on ice and irradiated

with either 0.2 J, 0.4, 0.8, 1.0 J/cm² UV light at 254 nm in a GS Gene Linker UV Chamber. Cells were scraped from culture plates, washed once with PBS. Cross-linked cells were aliquoted into 1.5 mL microcentrifuge tubes in one million cell batches. Cells were pelleted by centrifugation at 1,600 x g for four minutes and frozen at -80° C.

2.13.3 Formaldehyde cross-linking

HEK 293FT cells were grown to 80-90% confluency in cell culturing conditions. Cells were washed twice with cold 1X PBS. Formaldehyde was added to the culture dishes to a final concentration of 0.1, 0.5, 1, or 3%. Culture plates were incubated at room temperature for 20 minutes with gentle agitation. Formaldehyde cross-linking was quenched by adding glycine to a final concentration of 0.2 M. Cells were scraped from culture plates and washed twice with PBS. Cross-linked cells were aliquoted into 1.5 mL microcentrifuge tubes in one million cell batches. Cells were pelleted by centrifugation at 1,600 x g for four minutes and frozen at -80°C.

2.13.4 Cell lysis

Cross-linked and non-cross-linked HEK 293FT cell pellets were re-suspended in cold lysis/binding buffer (20 mM Tris-HCl pH 7.5, 500 mM LiCl, 0.5% LDS, 1 mM EDTA, 5 mM DTT, 200 U RNase inhibitor, 2X protease inhibitor). Cells were incubated at 4°C for 10 minutes with gentle agitation to allow lysis to proceed. The cell lysate was passed through a 21-gauge needle (VWR) attached to a 3 mL syringe (VWR) to decrease viscosity.

2.13.5 Protein analysis

To test cross-linking efficiency, total mRNA was targeted with oligo-dT beads (NEB) as suggested by the manual. Poly(A) RNA-complexes were treated with 100 µg of RNase A (BioBasics) and incubated for 1hour at 37° C. Protein samples were run on a SDS-PAGE (Gallgher & Sass, 2001). The gel was silver-stained according to Pierce Silver Stain for MS protocol (ThermoScientific) to visualize protein bands.

CHAPTER 3

RESULTS: Test *in vitro* CRISPR/dCas9 system for GAPDH mRNA enrichment

3.1 Prepare *in vitro* CRISPR/dCas9 components for GAPDH mRNA enrichment.

3.1.1 Overexpressing dCas9

To express dCas9 for the *in vitro* GAPDH mRNA enrichment, we first obtained the dCas9 plasmid used by O'Connell et al. (2014) from Addgene. This plasmid was designed to have mutations to inactivate Cas9 nuclease activity while retaining the enzyme's targeting abilities (D10A; H840A), a TEV protease cleavable N-terminal his-MBP tag for purification and solubility purposes respectively and an IPTG-inducible promoter for controlled protein expression. Using this plasmid, dCas9 was expressed in IPTG-induced *E. Coli* BL21 (DE3) Rosetta at O.D.₆₀₀ of 0.7. Protein expression was verified by comparing pre and post-induction samples on a SDS-PAGE stained with coomassie-blue. Post-induction samples have a strong band at approximately 210 kDa, which is the predicted size of the dCas9-his-MBP protein (**Figure 3.1A**). Minimal expression was detected in pre-induction samples. Protein identity was confirmed by western blot probed with a Cas9-specific antibody (**Figure 3.1B**). These results suggest dCas9 was successfully overexpressed.

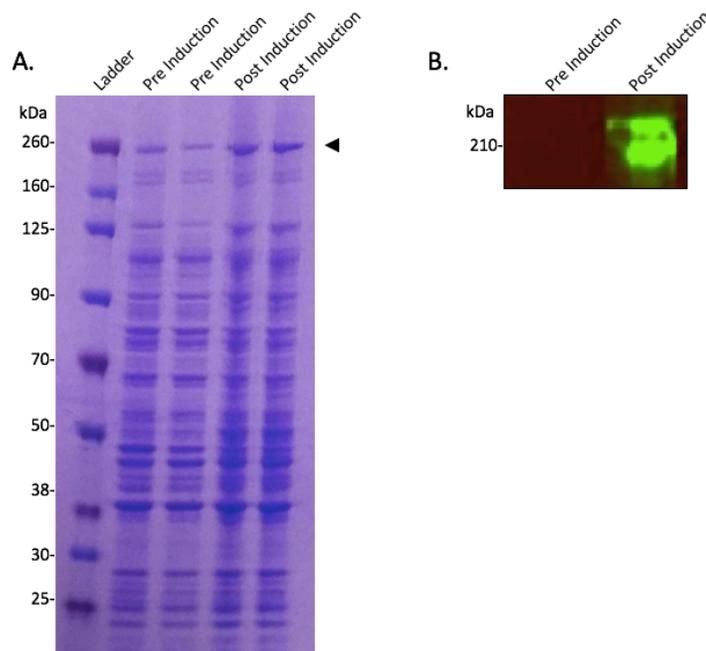


Figure 3.1 Overexpression of his-MBP tagged dCas9
dCas9 was overexpressed in IPTG-induced *E. Coli* BL21 (DE3) cultured in LB medium containing ampicillin. Pre (lane 2, 3) and Post-induction (lane 4, 5; sixteen hours after IPTG induction) samples were analyzed on a coomassie-stained SDS-PAGE (A) and western blot probed with Cas9-specific antibody (B). Black triangle indicates dCas9 at 210 kDa expressed after IPTG addition.

3.1.2 Purification of dCas9

To purify dCas9, we used a HisTrap column to specifically target the N-terminal his-MBP tag on dCas9. Flow-through, wash and elution fractions were run on a coomassie stained SDS-PAGE to determine dCas9 purity. Elution fractions have a concentrated band at approximately 210 kDa on the SDS-PAGE stained with coomassie-blue (Figure 3.2A/B). At higher imidazole concentrations, more dCas9 was eluted. There were some lower, fainter bands detected in the elution which were contaminant proteins retained from purification (around 40 kDa). Flow-through and wash fractions have minimal dCas9

detected. The protein in elution fractions was confirmed to be dCas9 by western blot probed with a Cas9-specific antibody (**Figure 3.2C**).

To remove the his-MBP tag, which may interfere with dCas9 function, we used TEV protease that would target the TEV recognition site between the his-MBP tag and dCas9 sequence. TEV-cleaved dCas9 was detected by a coomassie-blue stained SDS-PAGE as displayed by the band at ~160 kDa, the predicted band size of cleaved dCas9 (**Figure 3.3A**). Notably, more concentrated samples retained more un-cleaved dCas9 and all TEV-cleaved dCas9 samples have more contaminate bands post-cleavage. To remove contaminate proteins and un-cleaved dCas9 retained from the cleavage reaction, TEV-cleaved samples underwent an additional round of HisTrap purification with a step gradient. Purified TEV-cleaved dCas9 was isolated from the majority of contaminants as seen by the band at 160 kDa in elution fractions 1D1-1D12 (**Figure 3.3B/C**), which corresponds to approximately 60 mM of imidazole. Contaminate proteins and un-cleaved dCas9 were eluted at higher imidazole concentrations (250-300 mM). Cleaved dCas9 identity was confirmed by western blot probed with a Cas9-specific antibody (**Figure 3.3D**). Collectively, these results suggest dCas9 was successfully purified.

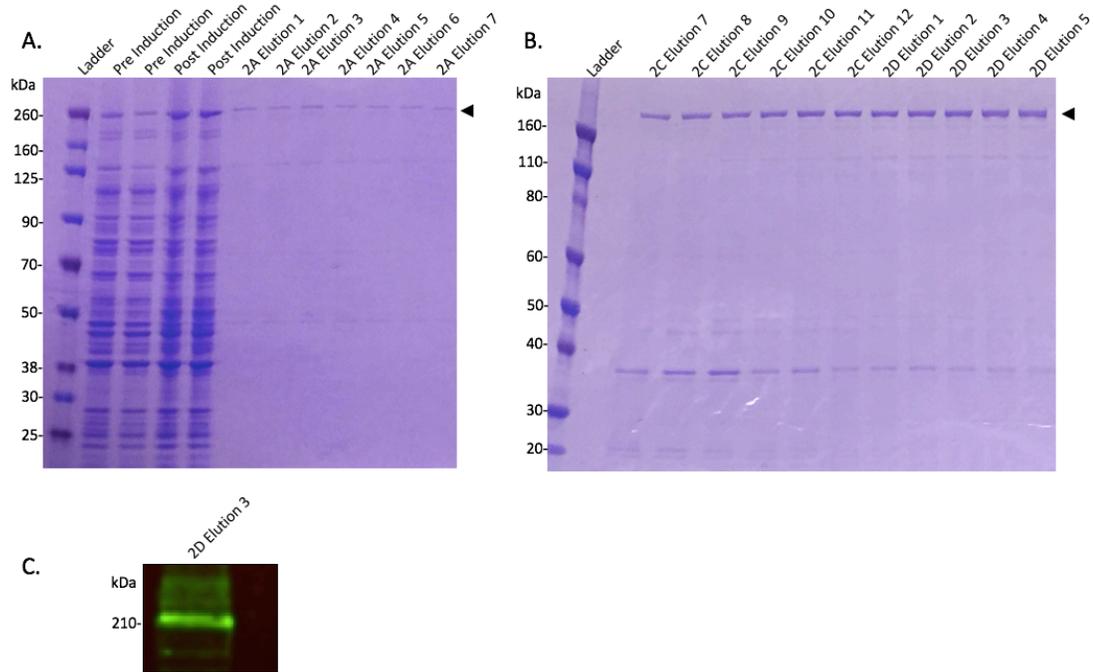


Figure 3.2 Purification of his-tagged dCas9

dCas9 was first overexpressed in IPTG-induced *E. Coli* BL21 (DE3) cultured in LB medium containing ampicillin. Cultures were harvested 16 hours after IPTG induction and lysed by a homogenizer to obtain the cytoplasmic fraction, which was then loaded onto a HisTrap column. Pre-Induction (lane 2,3), post-induction (lane 4,5) and elution fractions (A: lane 6-12, B: 2-12) from the HisTrap purification were analyzed on a coomassie-stained SDS-PAGE (A, B). Elution fraction 2D3 was blotted with a Cas9-specific antibody to check protein identity (C). Black triangles indicate dCas9 at 210 kDa in post-induction and elution fractions.

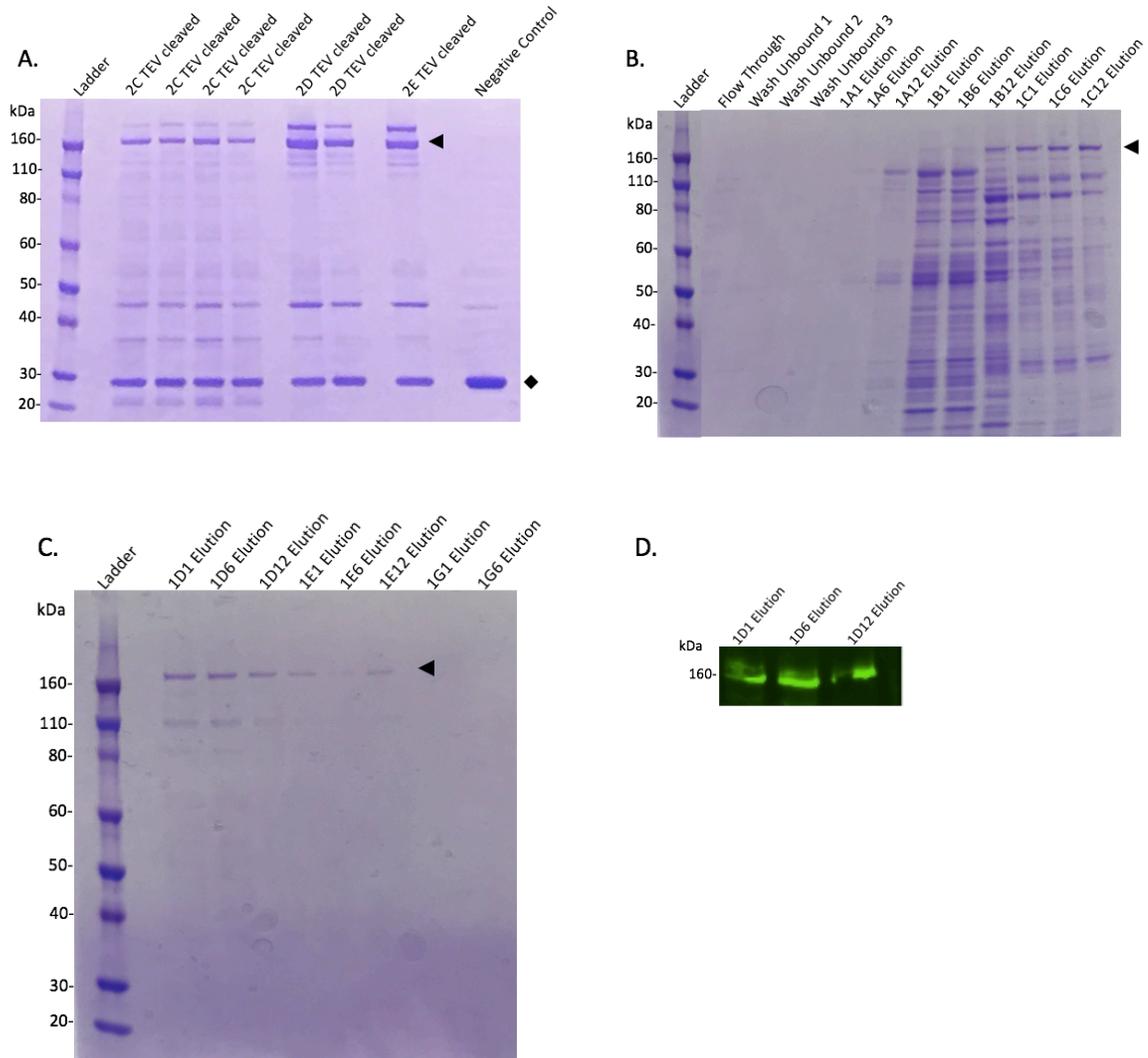


Figure 3.3 *dCas9 his-MBP tag removal*

Different fractions of purified dCas9 samples (2C, 2D from **Figure 3.2**) were treated with TEV protease to remove the N-terminal his-MBP tag on dCas9. Cleaved dCas9 samples (lane 2-8) and a negative control with only TEV protease (lane 9) were analyzed on a coomassie-blue stained SDS-PAGE to determine cleavage efficiency (**A**). Cleaved dCas9 samples were then pooled and loaded onto a HisTrap column for step gradient purification. Flow-through (Lane 2), wash (Lane3-5) and elution fractions (**B**: lane 6-14, **C**: lane 2-9) were analyzed on a coomassie-blue stained SDS-PAGE (**B/C**). Elution fractions 1D1,1D6 and 1D12 were run on a western-blot and probed with a Cas9-specific antibody to check protein identity (**D**). Black triangle indicates cleaved dCas9 at 160 kDa in post-TEV cleaved and elution samples. Black diamond indicates TEV protease at 27 kDa.

3.1.3 Biotinylation of dCas9

To capture dCas9/gRNA/PAMmer/RNA/RBP complexes formed during GAPDH mRNA enrichment, we biotinylated the dCas9 protein. O'Connell et al. (2014) have modified the original dCas9 sequence in their plasmid to contain only one cysteine residue and therefore, one thiol group that can be targeted by EZ-Link™ Iodoacetyl-PEG2-Biotin labelling reagent (ThermoScientific). Biotin-labelling of dCas9 was confirmed using a western blot probed with a streptavidin antibody (**Figure 3.4A**).

To test whether the biotinylated dCas9 binds specifically to streptavidin beads, we performed a streptavidin-bead binding assay where we incubated non-biotinylated and biotinylated dCas9 with streptavidin beads. Input, flow-through, wash and elution fractions from biotinylated and non-biotinylated dCas9 were run on western blot probed with either a Cas9-specific or streptavidin antibody. When probed with a Cas9-specific antibody, the western blot showed that the majority of the biotinylated dCas9 binds to the streptavidin beads, as it is only detected in the elution (**Figure 3.4B**). Non-biotinylated dCas9 was detected mainly in the flow-through and minimally in the elution (**Figure 3.4B**). When probed with a streptavidin antibody, it can be seen that the majority of the dCas9 detected in the elution by the Cas9-specific antibody (**Figure 3.4C**) was indeed biotinylated. Non-biotinylated dCas9 was not detected here. These results suggest dCas9 was successfully biotinylated and binds specifically to the streptavidin beads.

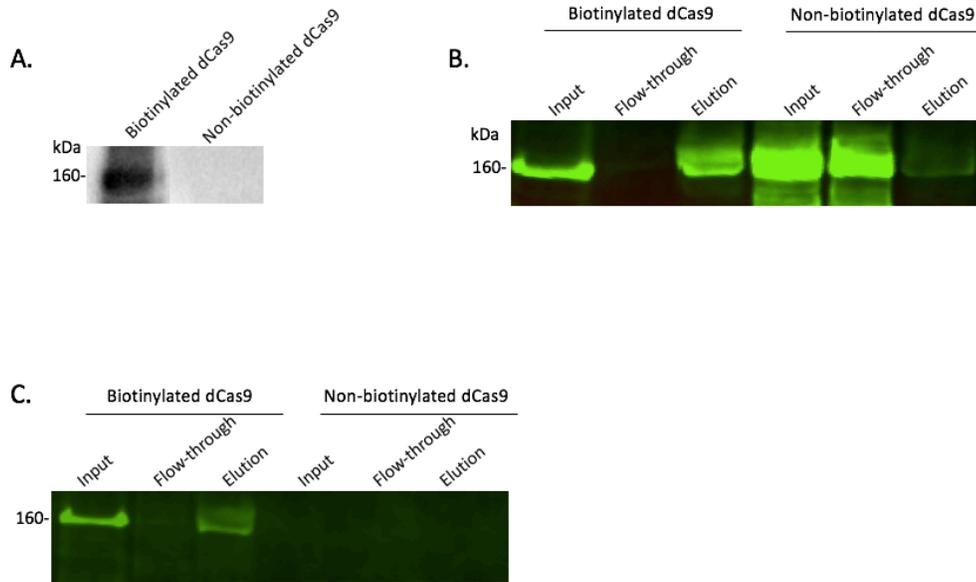


Figure 3.4 *Biotinylation of dCas9*

Purified cleaved dCas9 fractions (**Figure 3.3**) were pooled, dialyzed and labelled with EZ-Link™ Iodoacetyl-PEG2-Biotin labelling reagent. Biotinylation of dCas9 was verified by a western blot probed with a streptavidin antibody (EMD Millipore) (**A**). Biotinylated (lane 1-3) and non-biotinylated (lane 4-6) dCas9 samples were incubated with Dynabeads MyOne Streptavidin C1 beads to test binding specificity. Input, flow-through and elution fractions from this assay were analyzed on a western blot probed with a Cas9-specific antibody (**B**) and a streptavidin-specific antibody (**C**).

3.1.4 Design and produce gRNA and PAMmers targeting GAPDH mRNA

To help dCas9 recognize and specifically capture GAPDH mRNA, we designed and synthesized gRNAs. Each gRNA has two regions: first, there is the “protospacer sequence” which has a unique twenty-two base-pair sequence complementary to different exonic regions on GAPDH mRNA. Specifically, gRNA 1 targets exon 5, gRNA 2 targets exon 6 and gRNA 3 targets exon 7 on GAPDH mRNA. These exon regions were selected as they are common amongst all GAPDH mRNA isoforms (O’Connell et al., 2014). gRNAs also

have a common ninety-five base pair scaffold sequence called the “tracRNA” so it can also hybridize with dCas9 (**Table 3.1**). *In vitro* transcribed gRNA for GAPDH mRNA was confirmed by urea-PAGE at the estimated size of 117 bps (**Figure 3.5**), indicating successful production of gRNAs. The size of the positive control, used as a marker here, is 128 bps.

Table 3.1 *gRNA sequences for targeting GAPDH mRNA*

Name	Sequence (5' → 3')
gRNA 1	<u>GGG</u> GCAGAGATGATGACCCTGTTTAAGAGCTATGCTGGAAACAG <u>CATACAAG</u> TTTAAATAAGGCTAGTCCGTTATCAACTTGAAAAAG TGGCACCGAGCGGTGCTTTTTTT
gRNA 2	<u>GGC</u> CAAAGTTGTCATGGATGACGTTTAAGAGCTATGCTGGAAAC <u>AGCATAGCAAG</u> TTTAAATAAGGCTAGTCCGTTATCAACTTGAAA AAGTGGCACCGAGTCGGTGCTTTTTTT
gRNA 3	<u>TGATG</u> TCATCATATTTGGCAGGGTTTAAGAGCTATGCTGGAAAC <u>AGCATAGCAAG</u> TTTAAATAAGGCTAGTCCGTTATCAACTTGAAA AAGTGGCACCGAGTCGGTGCTTTTTTT

Underlined is the unique GAPDH targeting sequence

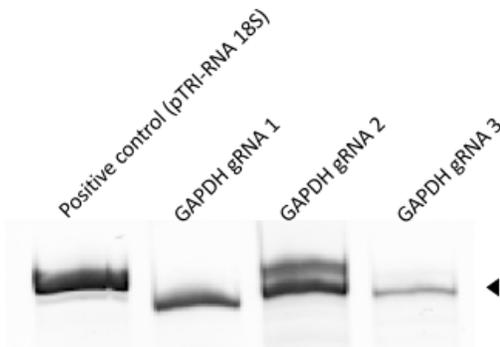


Figure 3.5 Confirmation of *in vitro* gRNA production for GAPDH mRNA enrichment

Three gRNAs were designed and synthesized, each containing a unique 22 bp sequence complementary to different exonic regions of GAPDH mRNA and a common 95 bp scaffold sequence. Designed gRNAs sequences were ordered as gBlocks and then transcribed into RNA using an *in vitro* T7 transcription kit (NEB). Transcribed gRNAs (lane 2-4) and control RNA (lane 1) were detected by urea-PAGE. Black triangle indicates gRNAs at 117 bp (the predicted size of gRNAs).

In addition to each gRNA, we ordered “PAMmers” or PAM presenting oligonucleotides since dCas9 will not target single-stranded RNA without a “PAM” sequence to differentiate between the target and non-target RNA sequences. As initially reported by O’Connell et al. (2014), we designed PAMmers to have a “PAM” with a 5’-NGG-3’ sequence. The PAM was then surrounded with ‘extension’ sequences complementary to GAPDH mRNA directly upstream and slightly overlapping their corresponding gRNA. The PAMmers were ordered to have a 2’O methyl modification to avoid any RNase H activity that may occur in cell lysate (**Table 3.2**).

Table 3.2 *PAMmer sequences for targeting GAPDH mRNA*

Name	Sequence
PAMmer v1.1	mATmGAmCCmCTm AGG mGGmCTmCCmCCmCCmTGmCAmAA
PAMmer v1.2	mATmGAmCCmCTm AGmGG mGCmTCmCCmCCmCTmGCmAAmA
PAMmer 2	mTGmGAmTGmACm CGG mGGmCCmAGmGGmGTmGCmTAmAG
PAMmer 3	mTTmGGmCAmGGm TGG mTTmCTmAGAmCGGmCAmGGmTC

NGG is the PAM sequence

m- Nucleotide before is 2'O methylated

3.1.5 Comparing the relative expression of GAPDH mRNA and β -actin mRNA in HEK 293FT cells

To test the relative expression of GAPDH and β -actin mRNA in HEK 293FT cells, we isolated total RNA from cells and performed reverse transcription to obtain cDNA, which was probed for both GAPDH and β -actin mRNA using RT-qPCR. This would confirm the high abundance of the GAPDH mRNA and provide re-assurance that the transcript we intend to target was present. It would also allow us to compare expression levels of GAPDH and β -actin mRNA. We found there a 1:1 detection of GAPDH mRNA to β -actin mRNA in HEK 293FT cells (**Figure 3.6**)

During RT-qPCR, a positive reaction is determined by the accumulation of a fluorescence signal. The cycle threshold (CT) value is defined as the number of cycles required for the fluorescence signal to pass the threshold, or the background level, to achieve a 'real' signal (Deepak et al., 2007). The CT value is inversely proportional to the amount of nucleic acid in the sample, so the lower the CT value the higher the amount of target nucleic acid in the sample. Typically, CT values below 29 indicate abundant amount of nucleic acids while CT values above 38 indicate minimum amounts (Deepak et al.,

2007). We found the CT value for GAPDH and β -actin mRNA to be 17.60 and 17.89, respectively, therefore indicating abundant amounts of each transcript in HEK 293FT cells.

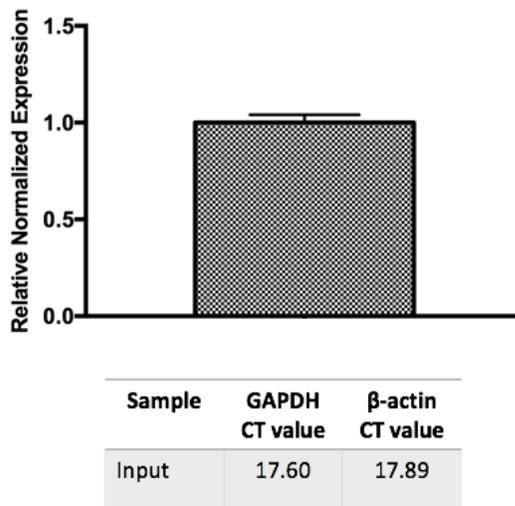


Figure 3.6 *Relative expression of GAPDH mRNA in HEK 293FT cells*

Total RNA from HEK 293FT cells was isolated and synthesized into cDNA by reverse transcription. cDNA was probed for GAPDH mRNA and β -actin mRNA by RT-qPCR. Top half is the relative expression plot of GAPDH mRNA compared to the reference, β -actin mRNA. Bottom half is the raw CT values from RT-qPCR. Error bars are standard error mean (SEM).

3.2 Enriching GAPDH mRNA with the *in vitro* dCas9 system.

To test GAPDH mRNA enrichment using the *in vitro* CRISPR/dCas9 system, biotinylated-dCas9 and total RNA from HEK 293FT cells were incubated with either gRNA 1 and PAMmer v1.1, gRNA 1 and PAMmer v1.2 or with no gRNA/PAMmer and the resulting complexes were captured with streptavidin beads. Input, flow-through, wash and elution fractions from the enrichment assay were analyzed by RT-qPCR in order to detect and quantify GAPDH mRNA. Results were normalized to β -actin mRNA expression and the

input was used as the reference sample. No enrichment of GAPDH mRNA was detected by dCas9 alone (**Figure 3.7 A**). A ~6-fold enrichment of GAPDH mRNA was detected in elution fractions by the biotinylated-dCas9, gRNA 1 and PAMmer v1.1 (**Figure 3.7 B**). Some GAPDH mRNA was detected in the wash fraction. The gRNA 1 and PAMmer v1.2 was only able to enrich GAPDH mRNA ~2-fold (**Figure 3.6 C**). No GAPDH mRNA was detected in the wash fraction here.

For dCas9 alone, the CT values of the flow-through fractions for both GAPDH and β -actin (15.38, 15.42 respectively) were lower than that of gRNA 1 with PAMmer v1.1 (16.07, 16.38 respectively) and v1.2 (17.38, 16.13 respectively). This suggest that both GAPDH and β -actin mRNA were being captured to varying degrees by the dCas9 gRNA1, PAMmer v1.1 and PAMmer v1.2. Though the elution fractions of the gRNA 1 systems suggest there was more GAPDH mRNA retained than β -actin compared to dCas9 alone, the CT value of β -actin in both cases (25.51 for v1.1 and 21.26 for v1.2) was relatively low. This could indicate that either dCas9 was non-specifically binding to β -actin mRNA or that the wash conditions were not stringent enough to remove non-specifically bound RNA.

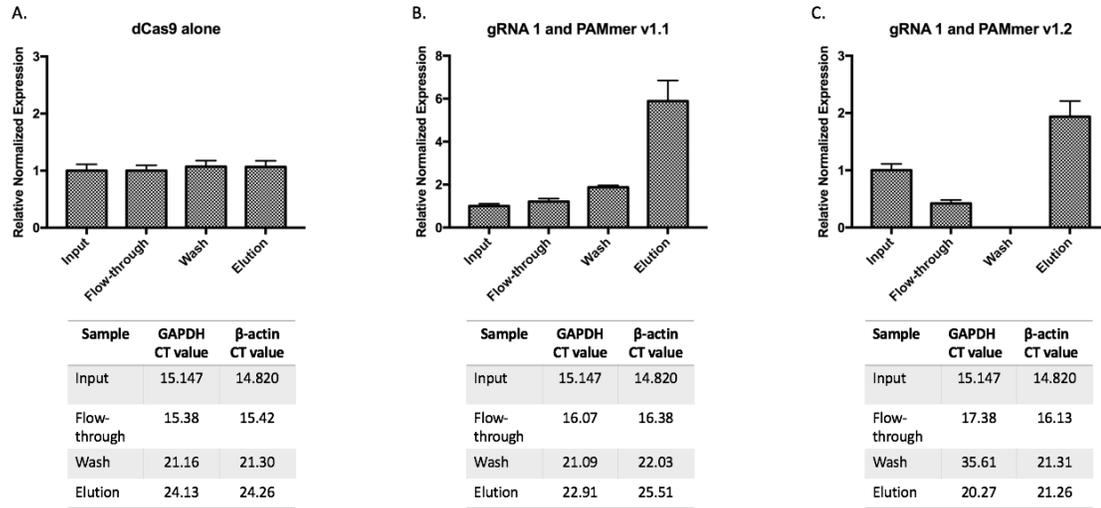


Figure 3.7 GAPDH mRNA enrichment by *in vitro* dCas9 system using gRNA1 and PAMmer v1.1 or PAMmer v1.2

Biotinylated dCas9 and total RNA from HEK 293FT cells were incubated with either with no gRNA or PAMmer (A), gRNA and PAMmer v1.1 (B), gRNA 1 and PAMmer v1.2 (C) then captured with Dynabeads MyOne Streptavidin C1 beads. Input, flow-through, wash and elution fractions from the enrichment assay were probed for GAPDH mRNA by RT-qPCR. Top half is the relative expression plots of GAPDH mRNA compared to the reference, β -actin mRNA. Bottom half is the raw CT values from RT-qPCR. Error bars are SEM.

3.2.1 Enrich GAPDH mRNA by the *in vitro* dCas9 system with gRNA 2 and 3

To determine whether the retention of β -actin mRNA was a phenomenon of gRNA 1 and PAMmer v1.1/v1.2, we tested two additional gRNA and PAMmer sequences targeting alternative regions on GAPDH mRNA (Refer to Chapter 2: Material and Methods, Section 2.6 Design of gRNA to target GAPDH mRNA). Biotinylated-dCas9 and total RNA from HEK 293FT cells were again incubated with either gRNA 2 and PAMmer 2, gRNA 3 and PAMmer 3 or with no gRNA/PAMmer and the resulting complexes were captured with streptavidin beads. Again, flow-through, wash and elution fractions from the enrichment assay were tested by RT-qPCR in order to detect and quantify GAPDH mRNA. Results

were normalized to β -actin mRNA expression. No enrichment of GAPDH mRNA was detected by dCas9 alone (**Figure 3.8 A**). A \sim 3-fold enrichment of GAPDH mRNA was detected in elution fractions by the biotinylated-dCas9, gRNA 2 and PAMmer 2 system (**Figure 3.8 B**), though there was nearly a \sim 3-fold detection of GAPDH mRNA in the wash. The gRNA 3 and PAMmer 3 was able to enrich GAPDH mRNA \sim 3.5-fold, with a similar loss of GAPDH mRNA to the wash fraction (\sim 3.6 fold, **Figure 3.8 C**).

Looking at the CT values of the elution fraction, we found for both gRNA 2/PAMmer 2 and gRNA 3/PAMmer 3, there was high retention of β -actin mRNA (CT for β -actin = 22.84, 22.98 respectively). This suggests that the capturing of non-specific RNA by our *in vitro* dCas9 system is not singularity of one specific gRNA.

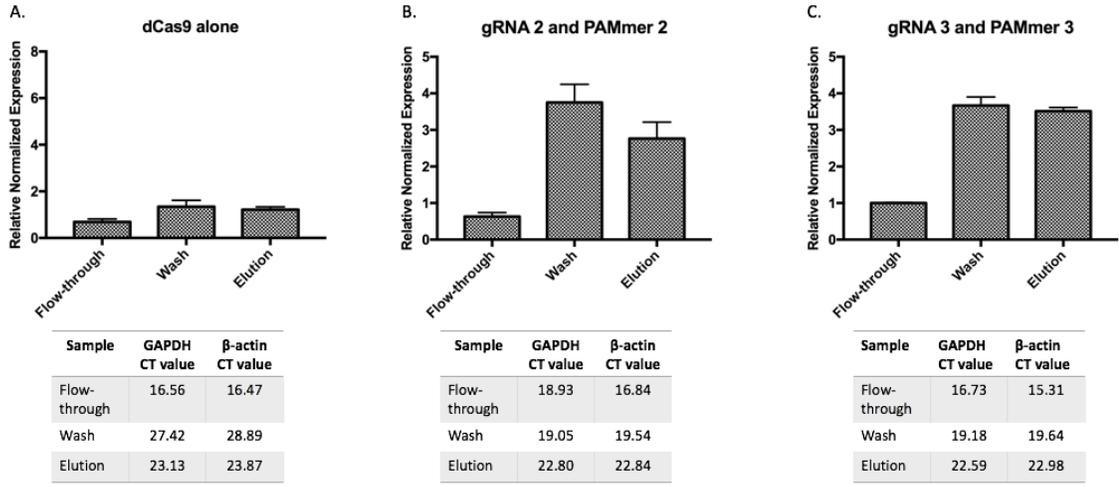


Figure 3.8 GAPDH mRNA enrichment by *in vitro* dCas9 system using gRNA2/PAMmer 2 and gRNA 3/PAMmer 3

Biotinylated dCas9 and total RNA from HEK 293FT cells were incubated with either with no gRNA or PAMmer (A), gRNA 2 and PAMmer 2 (B), gRNA 3 and PAMmer 3 (C) then captured with Dynabeads MyOne Streptavidin C1 beads. Flow-through, wash and elution fractions from the enrichment assay were probed for GAPDH mRNA by RT-qPCR. Top half is the relative expression plots of GAPDH mRNA compared to the reference, β -actin mRNA. Bottom half is the raw CT values from RT-qPCR. Error bars are SEM. Note: Flow-through fraction was used as the reference sample here.

3.2.2 Testing chromatin immunoprecipitation (ChIP) wash conditions for the *in vitro* dCas9 system

Since β -actin mRNA was retained by all three GAPDH gRNAs, we decided to test more stringent wash conditions to see if non-specifically bound RNA could be removed while still enriching the target GAPDH mRNA. We compared our current wash conditions to established protocols for IP, for example chromatin IP (ChIP), to gain insight on how non-specifically bound species were being accounted for. ChIP is a technique by which specific DNA regions are captured by targeting a known DNA-interacting protein with an antibody, which are then captured by its corresponding beads (Raha, Hong, & Snyder, 2010).

Antibodies and antigens are bound through weak non-covalent interactions, for example, hydrophobic interactions and van der waals forces (Sundberg, 2009). Wash buffers for ChIP, however, contains detergents such as SDS, which have the potential to disrupt the antigen-antibody binding. Thus, the use of small amounts of SDS in ChIP wash buffers indicates that the weak antibody-antigen interactions can still uphold under specific conditions of certain detergents.

In the case of the *in vitro* dCas9 system, the biotin-streptavidin interaction is considered to be stronger than antibody-antigen interactions due to the presence of multiple hydrogen bonds in addition to hydrophobic interactions (Diamandis & Christopoulos, 1991). Furthermore, interactions between dCas9-gRNA, dCas9-PAMmer and the gRNA-target RNA consist of strong hydrogen bonding (Jiang & Doudna, 2017). The current wash conditions, however, use 0.1 % Triton X-100, considered a ‘weak’ detergent, to remove non-specifically bound RNA. Therefore, we were interested in testing whether dCas9 could still enrich GAPDH mRNA and if β -actin mRNA could be effectively removed under ChIP wash conditions (Refer to Chapter 2: Material and Methods, Section 2.9 *In vitro* GAPDH mRNA pull down with dCas9). With dCas9 alone, there was no enrichment of GAPDH mRNA (**Figure 3.9 A**). The raw CT values, however, showed there was some retention of both GAPDH mRNA and β -actin mRNA (19.81, 20.47 respectively for low salt wash) though by the final wash, there was little of each mRNA present (32.02, 28.75 respectively for elution). This could indicate promiscuous binding by dCas9 to RNA.

Using gRNA 1 and PAMmer v1.1, we found that with each wash step of ChIP (Low Salt, High Salt and LiCl) there was an increase in CT values for β -actin mRNA, suggesting

the retained β -actin mRNA was effectively being removed (**Figure 3.9 B**). However, there was also a corresponding loss in GAPDH mRNA, as seen by the \sim 2-fold detection of GAPDH mRNA in each wash step (**Figure 3.9 B**). Due to the loss of GAPDH mRNA, there was no enrichment detected in the elution fraction. The raw CT values show that the first two wash steps resulted in a largest loss of GAPDH mRNA (19.90, 23.99 respectively). Though the CT values indicate there was more GAPDH mRNA compared to the dCas9 alone control (29.92, 32.02 respectively) in the elution, the high value indicates there was little GAPDH mRNA enriched overall. Moreover, this was comparable to the amount of β -actin retained in the elution fraction. Collectively, these findings suggest that though ChIP buffers help remove non-specifically bound RNA, using stronger conditions may be too stringent and therefore interfere with GAPDH mRNA enrichment by the *in vitro* dCas9 system.

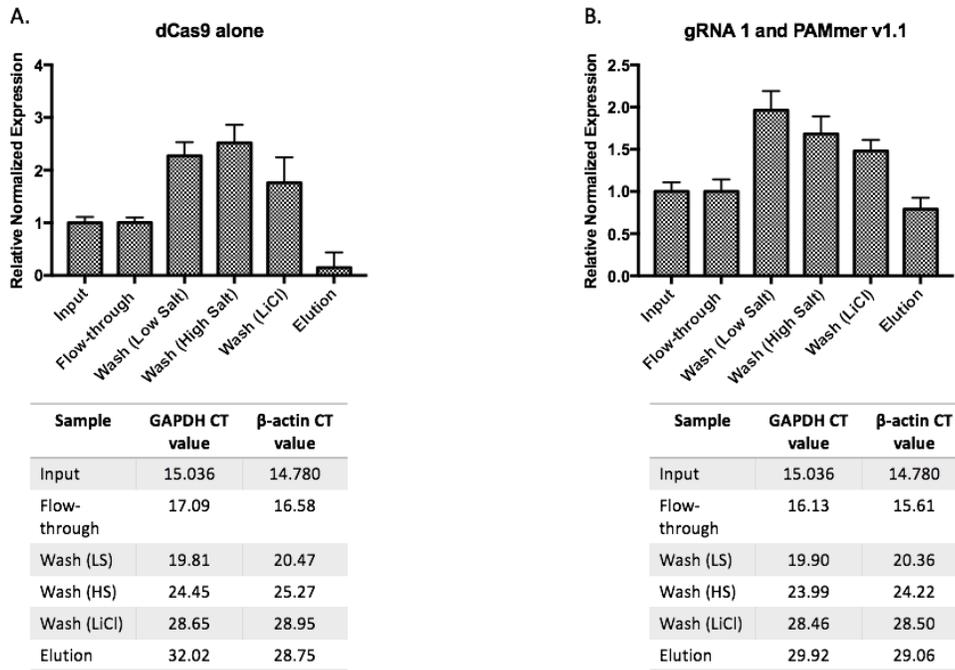


Figure 3.9 *GAPDH* mRNA enrichment by *in vitro* dCas9 system using ChIP wash buffers. Biotinylated dCas9 and total RNA from HEK 293FT cells were incubated with either with no gRNA or PAMmer (A) or gRNA and PAMmer v1.1 (B) then captured with Dynabeads MyOne Streptavidin C1 beads. Input, flow-through, wash (low salt (LS), high salt (HS) and LiCl wash (LiCl)) and elution fractions from the enrichment assay were probed for *GAPDH* mRNA by RT-qPCR. Top half is the relative expression plots of *GAPDH* mRNA compared to the reference, β -actin mRNA. Bottom half is the raw CT values from RT-qPCR. Error bars are SEM.

We hypothesized that potentially the addition of SDS, a very potent detergent in comparison to Triton X-100, could, even at low amounts, interfere with dCas9 binding to the target RNA or be an incompatible condition for the purified dCas9 protein. The ChIP buffers contain other, milder detergents (for example, 1% Triton X-100) and higher salt concentrations to help reduce non-specific binding. Therefore, we tested the ChIP conditions once again, this time eliminating SDS from the buffers. We found though there was no enrichment of *GAPDH* mRNA by the dCas9 alone group, there was still retention

of both GAPDH mRNA and β -actin mRNA as indicated by the CT values (19.43, 19.70 respectively for low salt wash, **Figure 3.10 A**). Again, indicating potential promiscuous binding by dCas9 to RNA. Additionally, in the elution, there was more GAPDH and β -actin mRNA retained using non SDS buffers (CT=24.79, 24.20 respectively) suggesting that these conditions were not stringent enough to remove non-specifically bound RNA.

Using gRNA 1 and PAMmer v1.1, we found nearly identical results to the previous SDS condition. Notably, there was a large loss of GAPDH mRNA during the first two wash steps (CT=19.52, 23.42 respectively), with a ~2-fold detection of GAPDH mRNA (**Figure 3.10 B**). Though the amount of GAPDH mRNA in the elution fraction was greater than that of the SDS condition and the dCas9 alone group (CT=24.63), there was still a nearly equal retention of beta actin mRNA (CT=24.52). This again re-iterates these conditions not being stringent enough to remove non-specifically bound mRNA completely. However, it also indicates these conditions were again too stringent to support GAPDH mRNA enrichment by the *in vitro* dCas9 system.

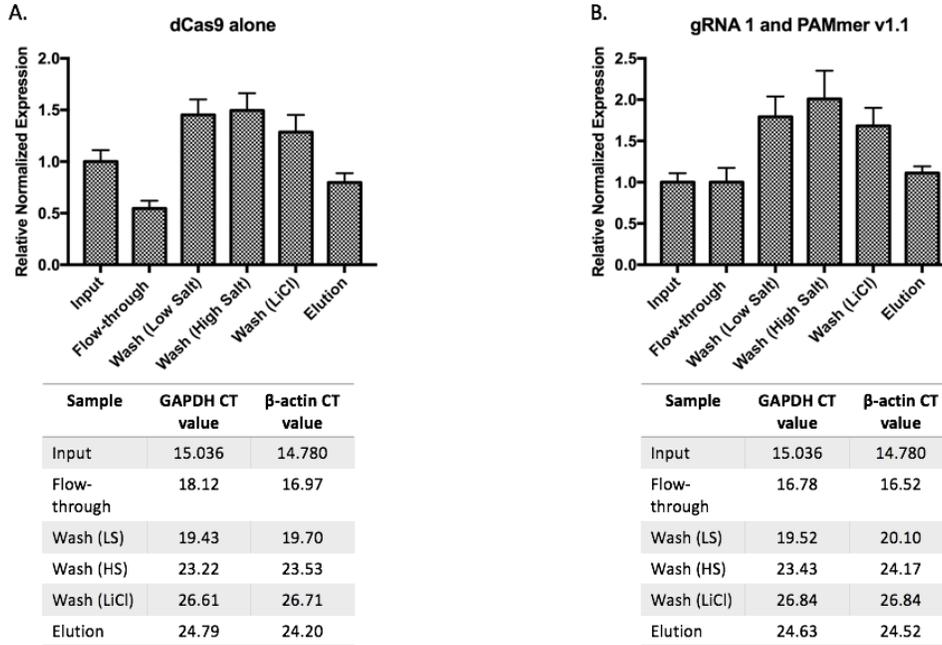


Figure 3.10 GAPDH mRNA enrichment by *in vitro* dCas9 system using ChIP wash buffers with no SDS

Biotinylated dCas9 and total RNA from HEK 293FT cells were incubated with either with no gRNA or PAMmer (A), gRNA and PAMmer v1.1 then captured with Dynabeads MyOne Streptavidin C1 beads. Input, flow-through, wash (low salt (LS), high salt (HS) and LiCl wash (LiCl)) and elution fractions from the enrichment assay were probed for GAPDH mRNA by RT-qPCR. Top half is the relative expression plots of GAPDH mRNA compared to the reference, β -actin mRNA. Bottom half is the raw CT values from RT-qPCR. Error bars are SEM.

3.2.3 Testing NP-40 wash conditions for the *in vitro* dCas9 system

Using ChIP buffers with and without SDS suggested there was a specific detergents and range in which dCas9 still retains GAPDH mRNA. We postulated that using a milder detergent to Triton X-100 at a higher amount, for example 1% NP-40, could potentially still support the interaction between dCas9-gRNA-PAMmer with GAPDH mRNA while effectively removing non-specifically bound RNA. In the dCas9 alone group, using the NP-40 buffer, there was no enrichment of GAPDH mRNA. We also found there was more

effective removal of GAPDH mRNA and β -actin mRNA compared to ChIP non-SDS buffers, with the elution CT values being 38.10, 36.92 respectively (**Figure 3.11 A**). However, the wash fraction suggests there was still initial retention of both GAPDH mRNA and β -actin mRNA (CT= 20.54, 21.32).

With gRNA 1 and PAMmer v1.1 we found the NP-40 buffer still results in loss of both β -actin and GAPDH mRNA (CT= 20.68, 21.13). This contributes to a lack enrichment of GAPDH mRNA in the elution fraction (CT value=30.38) (**Figure 3.11 B**). Therefore, though 1% NP-40 was stringent enough to remove non-specifically bound RNA, it also appears to interfere with the enrichment of the target RNA.

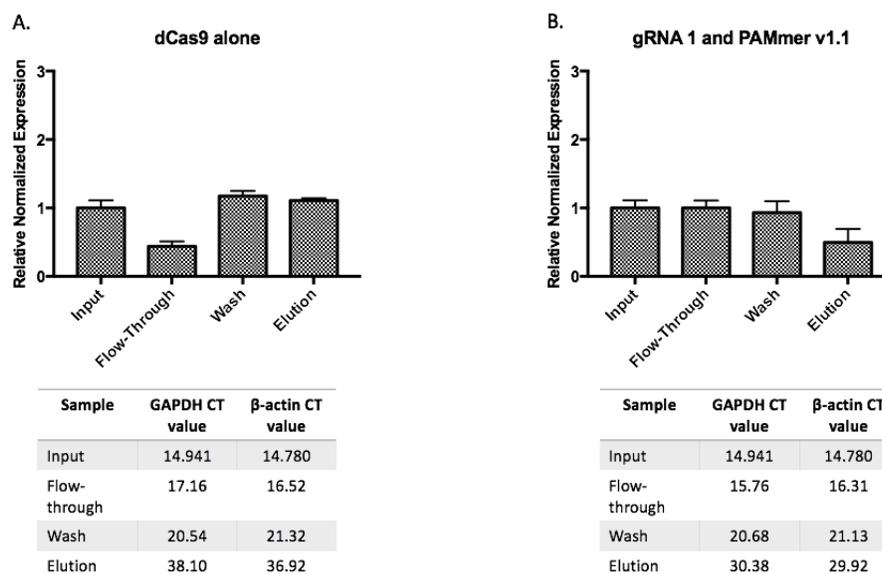


Figure 3.11 *GAPDH mRNA enrichment by in vitro dCas9 system using NP-40 buffer*
 Biotinylated dCas9 and total RNA from HEK 293FT cells were incubated with either with no gRNA or PAMmer (**A**), gRNA and PAMmer v1.1 then captured with Dynabeads MyOne Streptavidin C1 beads. Input, flow-through, wash and elution fractions from the enrichment assay were probed for GAPDH mRNA by RT-qPCR. Top half is the relative expression plots of GAPDH mRNA compared to the reference, β -actin mRNA. Bottom half is the raw CT values from RT-qPCR. Error bars are SEM.

3.2.4 Testing UV crosslinking and NP-40 conditions for the *in vitro* dCas9 system

Our tests with stringent wash conditions suggested that although the binding between the target RNA and the dCas9-gRNA-PAMmer complex involves hydrogen bonding, it was easily disrupted by both potent and mild detergents. We wondered whether strengthening the interaction between the dCas9-gRNA-PAMmer complex with the target RNA could help alleviate this issue. UV cross-linking introduces a covalent bond between two closely placed aromatic rings (one from a nucleic acid, the other, presumably from a RBP; Poria & Ray, 2017). This could help maintain the dCas9-target RNA complex while also permitting the rigorous removal of non-specifically bound RNA. We tested three different UV cross-linking conditions, 50, 100, 400 mJ/cm² post-incubation of the dCas9/gRNA/PAMmer complex with total RNA. These conditions were selected as they are on the lower range of the GS Gene Linker UV Chamber and would help minimize off-target UV cross-linking events, for example the potential cross-linking of a non-specific RNA in close proximity to dCas9.

Using the NP-40 conditions and 50 mJ/cm², we found that there was less GAPDH detected in the flow-through compared to non cross-linked conditions (CT=value 18.725, 15.76 respectively), though there appears to be ~2-fold more GAPDH mRNA in the flow-through in comparison to β -actin mRNA. We also found there was less loss of both β -actin and GAPDH mRNA to the wash fraction (CT= 25.698, 26.041 respectively) compared to non cross-linked samples (**Figure 3.12 A**). However, there was still an equal retention of both RNAs in the elution fraction, contributing to a lack of GAPDH mRNA enrichment (~1.2 fold).

With 100 mJ/cm², we again saw there was less loss of both β -actin and GAPDH mRNA to the wash fraction (CT= 25.196, 26.2610 respectively) compared to non cross-linked samples (**Figure 3.12 B**). However, there was still an equal retention of both RNA in the elution fraction, contributing to a lack of GAPDH mRNA enrichment. It is worth noting that the 100 mJ/cm² flow-through fraction has a ~4-fold more GAPDH mRNA than β -actin mRNA (CT=17.202, 19.725), thereby further contributing to the lack of GAPDH mRNA enrichment in the elution step.

Similar patterns were also observed for 400 mJ/cm² for the wash fractions. The flow-through fraction, however, contained less GAPDH mRNA (CT=18.092) while the amount of β -actin mRNA remained consistent between all cross-linking conditions. The 400 mJ/cm² elution fraction also contained slightly more GAPDH mRNA compared to lower mJ/cm² conditions (CT value= 24.190), while the amount of β -actin retained remained consistent. This contributed to ~2-fold enrichment of GAPDH mRNA. These findings indicate at higher UV intensities (**Figure 3.12 C**), there was potentially a more stable dCas9-gRNA-PAMmer complex formed with the target RNA, since more GAPDH mRNA was retained. Cross-linking did not result in a corresponding increase in β -actin mRNA retention, though there was still β -actin mRNA detected in the elution fraction comparable to GAPDH mRNA.

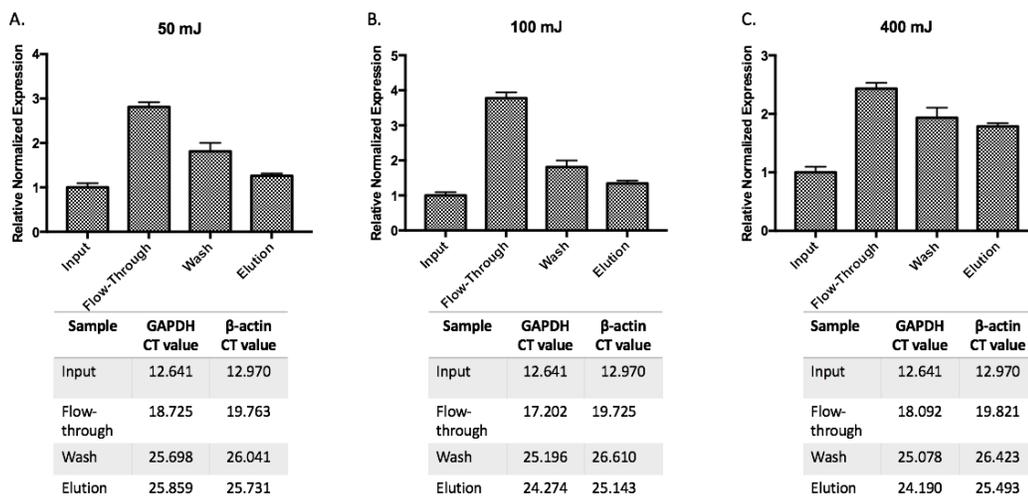


Figure 3.12 *GAPDH* mRNA enrichment by *in vitro* dCas9 system using UV cross-linking. Biotinylated dCas9 and total RNA from HEK 293FT cells were incubated gRNA 1 and PAMmer v1.1 then irradiated with either 50 (A), 100 (B), or 400 (C) mJ/cm² of UV light at 254 nm. Complexes formed with dCas9 were captured with Dynabeads MyOne Streptavidin C1 beads. Input, flow-through, wash and elution fractions from the enrichment assay were probed for GAPDH mRNA by RT-qPCR. Top half is the relative expression plots of GAPDH mRNA compared to the reference, β -actin mRNA. Bottom half is the raw CT values from RT-qPCR. Error bars are SEM.

3.2.5 Examining RNA specificity of *in vitro* dCas9 system with NP-40 buffers

Thus far, our results have consistently indicated that our *in vitro* dCas9 system was non-specifically capturing and enriching β -actin mRNA. We wondered whether this effect was specific to β -actin mRNA or if dCas9 was also capturing other RNA targets to a similar degree. Therefore, to further examine the RNA-specificity of the *in vitro* dCas9 system, we performed enrichment of GAPDH mRNA, using NP-40 buffer conditions, and probed for GAPDH and β -actin mRNA and U2 small nuclear RNA (snRNA) in the flow-through, wash and elution fractions. U2 is a component of the major spliceosome machinery. In its mature form, it forms complexes with polypeptides to form small nuclear

ribonucleoproteins (snRNPs). U2 snRNA is highly abundant in cells, making it a suitable reference gene to compare to GAPDH and β -actin mRNA. With gRNA and PAMmer v1.1, we found there ~7-fold more GAPDH mRNA than U2 snRNA in the flow-through fraction, indicating that a large portion of GAPDH mRNA was not captured by the dCas9-gRNA 1-PAMmer v1.1 system (**Figure 3.13 A**). There was also more GAPDH mRNA detected in the wash fraction, then the elution fraction, contributing to a lack of enrichment in the elution fraction (enrichment fold ~2.54, 1.738 respectively). CT values indicate that following washes, there was little GAPDH mRNA present in the elution fraction, though it was more than U2 snRNA (27.219, 28.218 respectively).

Interestingly, the enrichment of β -actin mRNA in comparison to U2 snRNA was similar to the results from GAPDH mRNA. There was less β -actin mRNA detected in the flow-through fraction than GAPDH mRNA (~4.5-fold, CT = 19.320), suggesting potentially more β -actin mRNA was being captured (**Figure 3.13 B**). Examining the elution fraction seemingly confirms this, as there was ~2-fold enrichment of β -actin mRNA with a lower CT-value than that of GAPDH mRNA (CT = 26.93). These results indicate that β -actin mRNA was being preferentially captured by the *in vitro* dCas9 system.

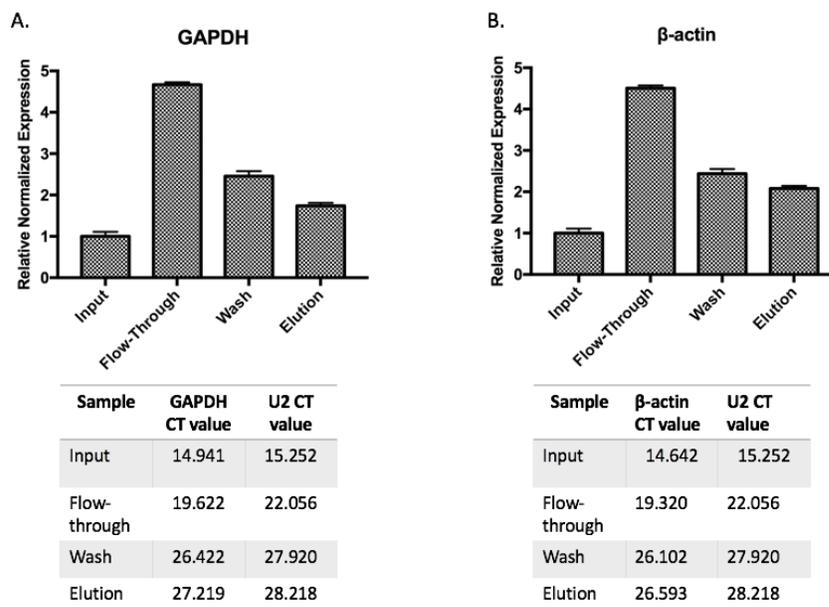


Figure 3.13 NP-40 buffers: GAPDH or β -actin mRNA enrichment by *in vitro* dCas9 system using U2 snRNA as a reference

Biotinylated dCas9 and total RNA from HEK 293FT cells were incubated with gRNA and PAMmer v1.1 and complexes formed with dCas9 were captured with Dynabeads MyOne Streptavidin C1 beads. Input, flow-through, wash and elution fractions from the enrichment assay were probed for GAPDH and β -actin mRNA by RT-qPCR. Top half is the relative expression plots of GAPDH (A) and β -actin (B) mRNA compared to the reference U2 RNA. Bottom half is the raw CT values from RT-qPCR. Error bars are SEM.

3.2.6 Examining RNA specificity of *in vitro* dCas9 system with UV cross-linking

As previously mentioned, a concern with using UV cross-linking to stabilize the dCas9-gRNA-PAMmer-target RNA complex is the corresponding stabilization of dCas9 complexes formed with non-specific RNA, leading to non-specific capturing of RNA and RBPs. To explore the effects of UV cross-linking and the capturing of non-specific RNA by the *in vitro* dCas9 system, we performed enrichment of GAPDH mRNA, using UV cross-linking and NP-40 buffer conditions, and probed for GAPDH and β -actin mRNA and U2 snRNA in the flow-through, wash and elution fractions. When probing for GAPDH

mRNA, we found there was ~7-fold more GAPDH mRNA than U2 snRNA in the flow-through fraction and there was ~ 4-fold enrichment of GAPDH mRNA in the elution fraction (**Figure 3.14 A**). Little GAPDH mRNA was lost to the wash fraction. This suggests UV cross-linking did not result in capturing of U2 snRNA and that NP-40 buffers could be used to minimize the amount of U2 snRNA in the elution fraction (CT = 28.013).

However, when probing for β -actin mRNA, we found similar results to non cross-linked samples. Mainly, in the flow-through fraction there was less β -actin mRNA detected than GAPDH mRNA (~2.5-fold). Though, in this case, the elution fraction had less enrichment of β -actin mRNA (~3.2 fold), the CT value indicates there were comparable retention between GAPDH and for β -actin mRNA (CT= 25.759, 25.962 respectively, **Figure 3.14 B**). This suggests UV cross-linking may also result in stabilization of the binding between dCas9 and β -actin mRNA, again contributing to non-specific targeting of β -actin mRNA by our *in vitro* dCas9 system.

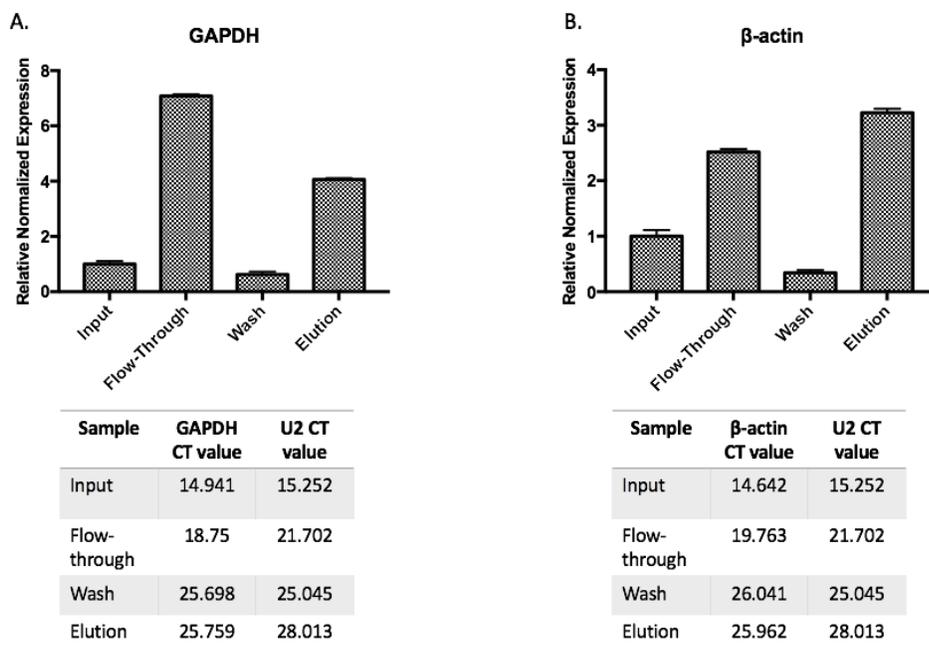


Figure 3.14 UV cross-linking: GAPDH or β -actin mRNA enrichment by *in vitro* dCas9 system using U2 snRNA as a reference

Biotinylated dCas9 and total RNA from HEK 293FT cells were incubated gRNA and PAMmer v1.1 then irradiated with 50 mJ/cm² of UV light at 254 nm. Complexes formed with dCas9 were captured with Dynabeads MyOne Streptavidin C1 beads. Input, flow-through, wash and elution fractions from the enrichment assay were probed for GAPDH and β -actin mRNA by RT-qPCR. Top half is the relative expression plots of GAPDH (A) and β -actin (B) mRNA compared to the reference U2 RNA. Bottom half is the raw CT values from RT-qPCR. Error bars are SEM.

3.3 Optimizing UV cross-linking in HEK 293 FT cells for RAP

To capture the RBPs associated with GAPDH mRNA with RAP, we cross-linked HEK 293FT cells to covalently link RNA to its interacting protein. Total mRNA-protein complexes, captured with oligo-dT beads, from UV cross-linked HEK 293FT cells were analyzed on a silver-stained SDS-PAGE to determine cross-linking conditions (Figure 3.15). As the dose or intensity of UV increases, more protein was recovered from the oligo-dT beads as displayed by the increase in protein band intensity (35, 50 kDa) from 0 mJ/cm²

to 1000 mJ/cm². There was an increase in protein band intensity from 400 mJ/cm² to 800 mJ/cm² while 800 mJ/cm² and 1000 mJ/cm² are nearly equivalent in intensity. Therefore, from these results, we concluded the cross-linking conditions for the GAPDH mRNA/protein enrichment would be 800 mJ/cm² for UV in HEK 293FT cells.

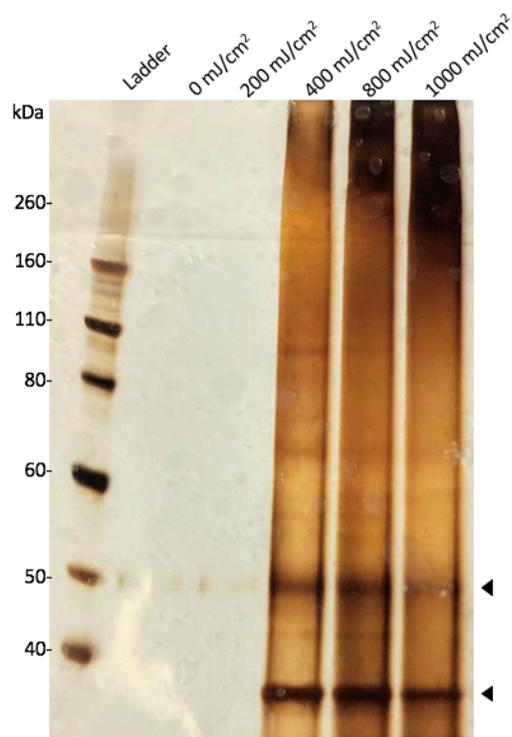


Figure 3.15 *Optimizing UV cross-linking conditions in HEK 293FT cells*

HEK 293FT cells were either irradiated with 0, 200, 400, 800, 1000 mJ/cm² (lane 2-6 respectively) UV light at 254 nm. Cells were harvested, lysed and mRNA-protein complexes were captured with oligo-dT beads. mRNA-protein complexes from cross-linking were analyzed on a silver-stained SDS-PAGE. Black triangles indicate protein bands (35, 50 kDa) that display different intensities at different cross-linking conditions.

3.3.1 Enrich GAPDH mRNA by RAP

To compare the *in vitro* dCas9 GAPDH mRNA enrichment to the RAP method, RAP was performed as previously described by McHugh et al. (2015). Input RNA and flow-through and elution fractions from the RAP assay were tested by RT-qPCR in order to detect GAPDH mRNA. Results were normalized to β -actin expression. A ~200-fold enrichment of GAPDH mRNA was detected in RAP elution fractions compared to input (**Figure 3.16**). In comparison to the *in vitro* dCas9 system, RAP appears to more efficient at enriching GAPDH mRNA. In addition, because of the strong wash conditions, there was more stringent removal of non-specifically bound RNA, in this case β -actin mRNA.

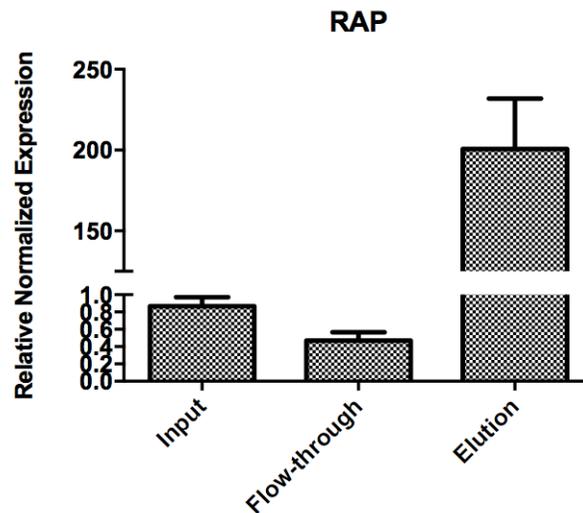


Figure 3.16 *GAPDH mRNA enrichment by RAP*

GAPDH mRNA was enriched by RAP in HEK 293FT cells, as previously described (McHugh et al., 2015). To detect GAPDH mRNA, input, flow-through and elution were probed for GAPDH and β -actin mRNA (reference) by RT-qPCR.

CHAPTER 4

RESULTS: Test *in vivo* CRISPR/dCas9 system for GAPDH mRNA enrichment

4.1 Create and verify constructs for dCas9 expression *in vivo*

To express dCas9 in HEK 293FT cells for GAPDH mRNA enrichment, we designed and created constructs. The construct's backbone was derived from pLKO.1 hygro, a construct that contains a variety of different promoters and restriction enzyme sites. Into the pLKO.1 hygro backbone and under the hPGK promoter, we added dCas9, cloned from a previously reported plasmid that expressed dCas9 *in vivo* and a halo-tag to the N-terminus of the dCas9 protein (Nelles et al., 2016). We decided on using a halo-tag to capture dCas9/gRNA/PAMmer/mRNA/RBP complexes because of the strong covalent interaction 'halo' forms with its corresponding beads. This means very stringent wash conditions can be used to remove non-specifically bound RNA and/or proteins. We also added the promoter "T2A" and puromycin to pLKO.1 hygro backbone so HEK 293FT cells expressing the construct could be selected for (**Figure 4.1 A**). The plasmid was confirmed by sanger sequencing (Mobix, McMaster).

To test the expression of the halo-tag dCas9 construct in HEK 293FT cells, we used PEI to transfect the created construct into HEK 293FT cells and selected for cells containing the construct with puromycin. After selection, we harvested and lysed the cells and tested for expression of dCas9 by a western blot probed with a Cas9-specific antibody. Compared to the pCDH control that does not contain or express dCas9, the created constructs expresses dCas9 (Lane 2 ,3 **Figure 4.1 B**). There was also a size shift to be noted

compared to flag-tagged dCas9 due to the presence of halo-tag. Collectively, these results show we were successfully able to produce a construct expressing halo-tagged dCas9 in HEK 293FT cells.

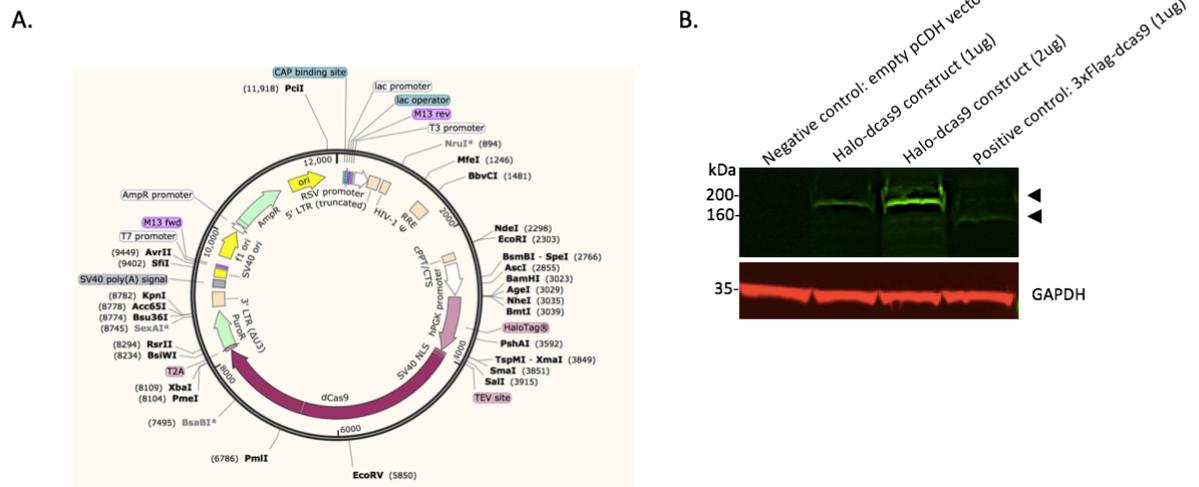


Figure 4.1 Verifying halo-tagged dCas9 expression in HEK 293FT cells

A construct was designed containing dCas9, N-terminal halo-tag, nuclear localization signal (NLS) and puromycin for dCas9 expression *in vivo* (A). pCDH control (lane 1), halo-tagged dCas9 construct (lane 2, 3) and flag-tagged dCas9 (lane 4) were transfected with PEI into HEK 293FT cells for 72-hours and then harvested. dCas9 expression was verified by western blot probed with a Cas9-specific antibody (B). Black triangles indicate halo-tagged dCas9 at 200 kDa and flag-tagged dCas9 at 160 kDa.

4.2 Testing binding efficiency and elution conditions for halo-dCas9

To test the binding efficiency of dCas9, we obtained lysate from our HEK 293FT cell-line that stably expresses halo-dCas9 and incubated it along with Halo-Link™ resin. Based on the recommendation from the Halo-ChIP protocol (Promega), we initially eluted the halo-dCas9 with 4x laemmli buffer supplemented with beta-mercaptoethanol. Input, flow-through, and elution fractions from the assay were run on western blot probed with a

Cas9-specific antibody. When probed with Cas9-specific antibody, the western blot shows that the majority of halo-dCas9 binds to the Halo-LinkTM resin, as it is only detected in the elution (**Figure 4.2A**), though the amount of halo-dCas9 recovered is considerably low.

We decided to optimize the elution conditions, this time using TEV protease to digest the linker (a TEV recognition site) between dCas9 and halo, effectively eluting it from the Halo-LinkTM resin. We found with TEV protease, we were able to recover more dCas9 (**Figure 4.2 B**).

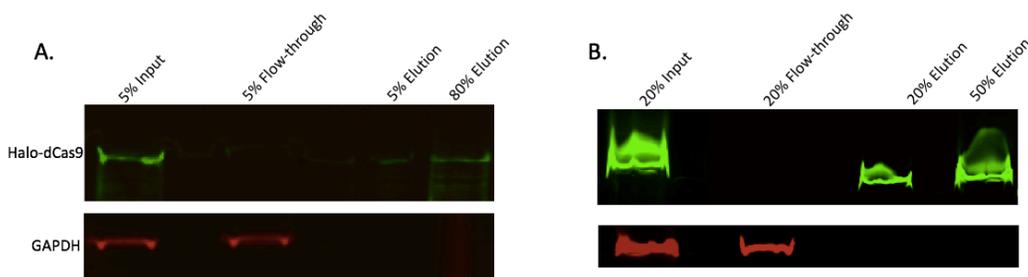


Figure 4.2 *Optimizing elution conditions for halo-dCas9 and the Halo-LinkTM resin*
Lysate from HEK 293FT cells stably expressing halo-dCas9 was incubated with Halo-LinkTM resin and eluted with either 4x laemmli buffer supplemented with beta-mercaptoethanol (**A**) or TEV protease (**B**). Input, flow-through and elution fractions from both conditions were analyzed on a western blot probed with a Cas9-specific antibody.

4.3 Create and verify constructs for gRNA expression *in vivo*

To express gRNA in HEK 293FT cells for GAPDH mRNA enrichment, we created constructs expressing gRNA under a U6 promoter. The constructs were created following a previously reported method, where a unique forward primer for the protospacer, a common reverse primer with ‘tracrRNA’ sequence and psLQ1651 vector were used to amplify the desired gRNA that was then cloned into the psLQ1651 vector (Liu et al., 2017).

The protospacers used in the forward primer were the same as those used in *in vitro* gRNAs. The plasmids were confirmed by sanger sequencing (Mobix, McMaster) and expression was confirmed by lenti-viral infection of gRNA into HEK 293FT cells (**Figure 4.3**).

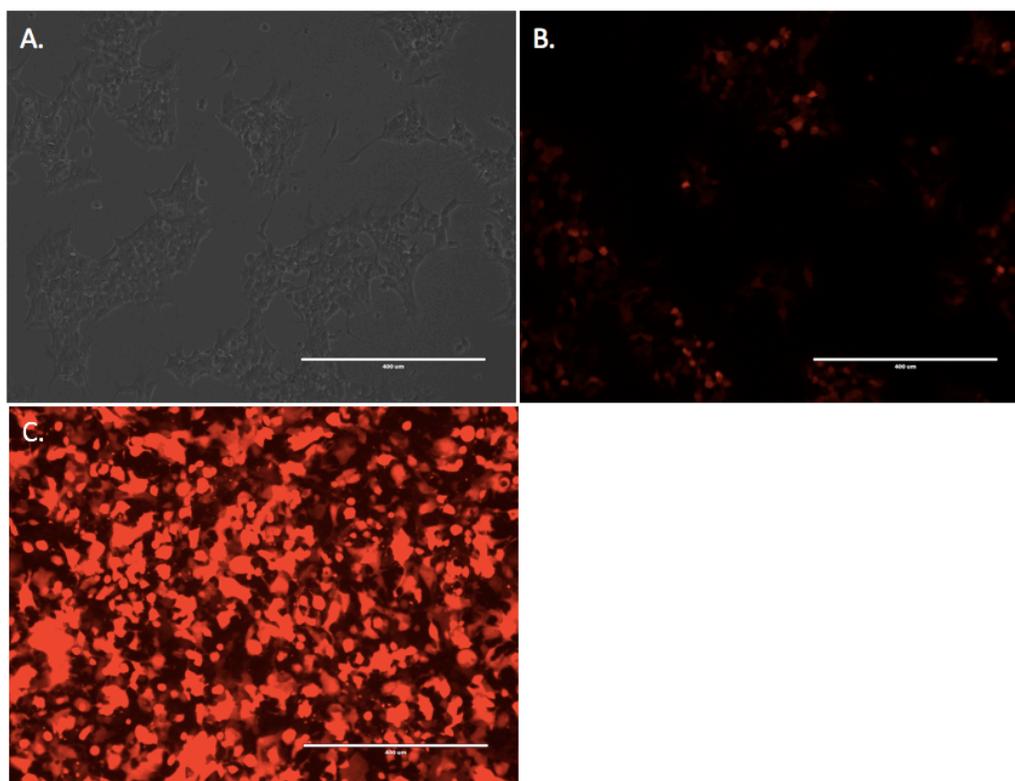


Figure 4.3 *Verifying gRNA expression in HEK 293FT cells*

Constructs expressing GAPDH gRNAs were designed as previously described (Liu et al., 2017) and transiently transfected in HEK 293FT cells with PEI. Images were taken at 24 hours (A/B) and 48 hours (C) post-transfection under 4X using a EVOS™ FL Imaging System (ThermoScientific).

4.4 Determining PAMmer transfection efficiency in HEK 293FT cells

As previously mentioned, gRNA sequences used were the same for the *in vitro* and *in vivo* dCas9 systems. Due to this, PAMmers sequences for the *in vivo* system would also be same

as the *in vitro* system. Due to the 2' O methyl modification, PAMmers would have to be delivered to HEK 293FT cells by transient transfection with Lipofectamine RNAiMax. By using a transfection reagent such as Lipofectamine RNAiMax, it is difficult to predict the efficiency of the PAMmer delivery to HEK 293FT cells. Therefore, to determine the optimal amount required and transfection efficiency of Lipofectamine RNAiMax, we transfected ~25,000 HEK 293FT cells with GAPDH PAMmer v1.1, containing a 5' 6-FAM tag, using varying amounts of Lipofectamine RNAiMax (0, 3, 6, 9 μ L) (**Figure 4.4**). We found that both 6 and 9 μ L of Lipofectamine RNAiMax resulted in similar transfection efficiency, with 62.1% and 58.6% of live cells being fluorescein isothiocyanate (FITC) positive (FITC signal is directly proportional with 5' 6-FAM signal) respectively by flow cytometry. The 3 μ L group had a slightly lower transfection efficiency of PAMmer v1.1, with only 44.2% of live cells being FITC positive. The control group, with no Lipofectamine RNAiMax, did not result in FITC signal. Thus, we decided that 6 μ L would be sufficient to achieve a moderate-high transfection efficiency and we could scale up accordingly for pull-down and MS experiments.

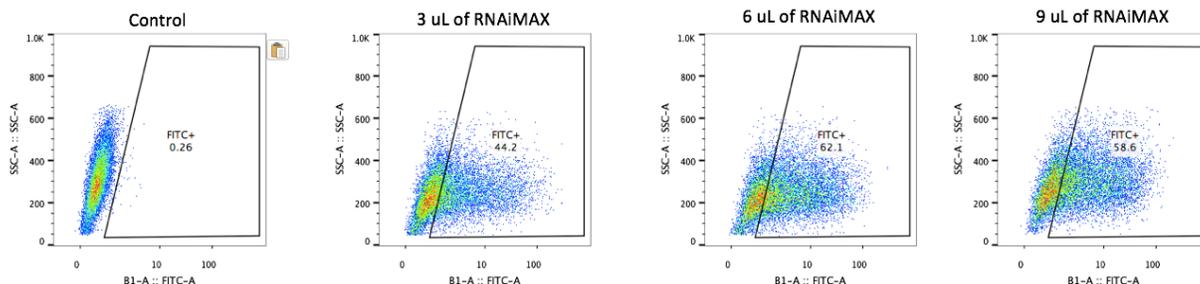


Figure 4.4 Testing PAMmer transfection efficiency in HEK 293FT cells

GAPDH PAMmer v1.1, containing a 5' 6-FAM tag, was transiently transfected in HEK 293FT cells using either 0, 3, 6 or 9 μ L of Lipofectamine $\text{\textcircled{R}}$ RNAiMAX. At 48 hours' post-transfection, percentage of live cells expressing GAPDH PAMmer 1 was determined using the MACSQuant $\text{\textcircled{R}}$ analyzer.

4.5 Optimizing formaldehyde cross-linking conditions for *in vivo* dCas9 system

To link any dCas9-gRNA-PAMmer-GAPDH mRNA complexes and to capture RBPs associated with GAPDH mRNA and other RBPs, we cross-linked HEK 293FT cells with formaldehyde. Total mRNA-protein complexes, captured with oligo-dT beads from formaldehyde cross-linked HEK 293FT cells were analyzed on a silver-stained SDS-PAGE to determine cross-linking conditions (**Figure 4.5**). As the percentage of formaldehyde was increased, more protein was recovered from the oligo-dT beads as displayed by the appearance of more intense protein bands (35, 50 kDa) from 0% to 3% formaldehyde. There was an increase in protein band intensity from 0.5% formaldehyde to 1.0% formaldehyde and from 1.0% formaldehyde to 3.0%. 3% formaldehyde appears to be the most efficient. From these results, we concluded the cross-linking conditions for the GAPDH mRNA/protein enrichment would be 3% formaldehyde in HEK 293FT cells.

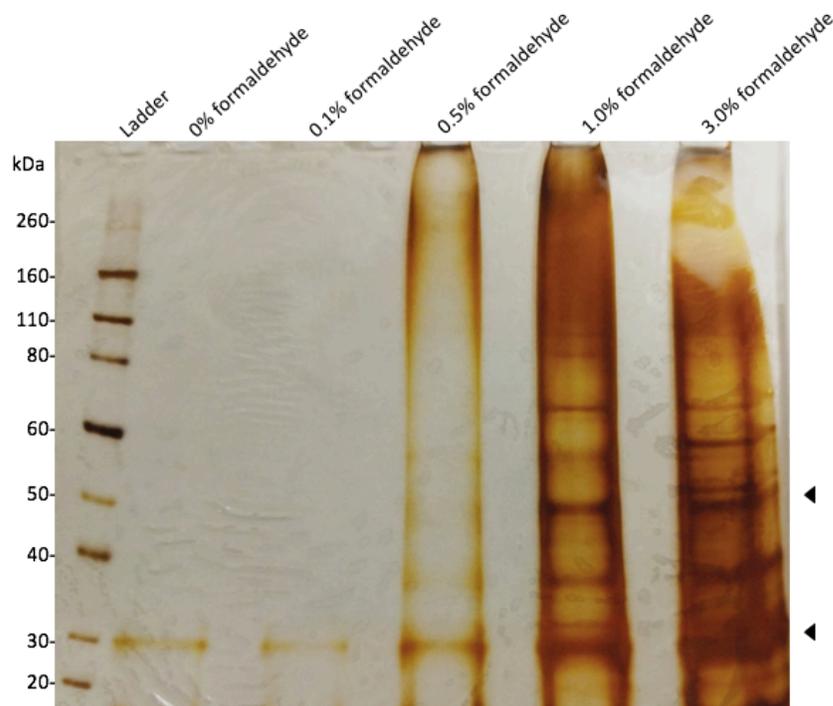


Figure 4.5 *Optimizing formaldehyde cross-linking conditions in HEK 293FT cells*
HEK 293FT cells were treated with 0, 0.1, 0.5, 1, or 3% (lane 2-6 respectively) formaldehyde to cross-link RNA to protein. Cells were harvested, lysed and mRNA-protein complexes were captured with oligo-dT beads. mRNA-protein complexes from cross-linking were analyzed on a silver-stained SDS-PAGE. Black triangles indicate protein bands (35, 50 kDa) that display different intensities at different cross-linking conditions.

4.6 Enriching GAPDH mRNA with the *in vivo* dCas9 system

To test GAPDH mRNA enrichment using the *in vivo* CRISPR/dCas9 system, our stable HEK 293FT cell-line expressing halo-dCas9 was infected with either gRNA 1 and PAMmer v1.1 with no tag, gRNA 1 and PAMmer v1.1 with 5' 6-FAM tag or with nothing. Cells were lysed and the resulting complexes formed with halo-dCas9 were captured with the Halo-LinkTM resin. Input, flow-through and elution fractions from the enrichment assay were tested by RT-qPCR in order to detect and quantify GAPDH mRNA. Results were

normalized to β -actin mRNA expression. For the dCas9 alone group, more GAPDH mRNA was detected in the flow-through fraction (~6-fold). There is a small enrichment of GAPDH mRNA detected in elution fraction of ~2-fold (**Figure 4.6 A**). The low CT values of the elution fraction for both GAPDH and β -actin mRNA suggest dCas9 may still retain both transcripts despite lacking a gRNA and PAMmer (24.995, 24.721 respectively).

For the gRNA 1 and 5' 6-FAM tagged PAMmer v1.1 group, ~1.4-fold fold more GAPDH mRNA than β -actin mRNA was detected in the flow-through fraction (**Figure 4.6 B**). This was accompanied by a lack of -fold enrichment of GAPDH mRNA in the elution fraction. It is worth noting that the CT value of β -actin mRNA and GAPDH mRNA were close in value (26.59, 26.432), which indicates potential capturing of β -actin mRNA as with the *in vitro* dCas9 system. For gRNA 1 and PAMmer v1.1 without a tag, we found there nearly ~1.3 fold more GAPDH mRNA in the flow-through fraction (**Figure 4.6 C**). There was also a lack of GAPDH mRNA enrichment in the elution fraction. The CT value of β -actin mRNA is relatively low (25.881), indicating some retention of β -actin mRNA between the flow-through and elution fraction. These findings suggest that further optimization is required to maximize the binding between GAPDH mRNA and halo-dCas9. Also, as with the *in vitro* dCas9 system, there was retention of non-specific RNA, in this case β -actin mRNA. Therefore, as mentioned there may be competitive binding between GAPDH and β -actin mRNA to halo-dCas9, contributing to the of lack enrichment of the target RNA.

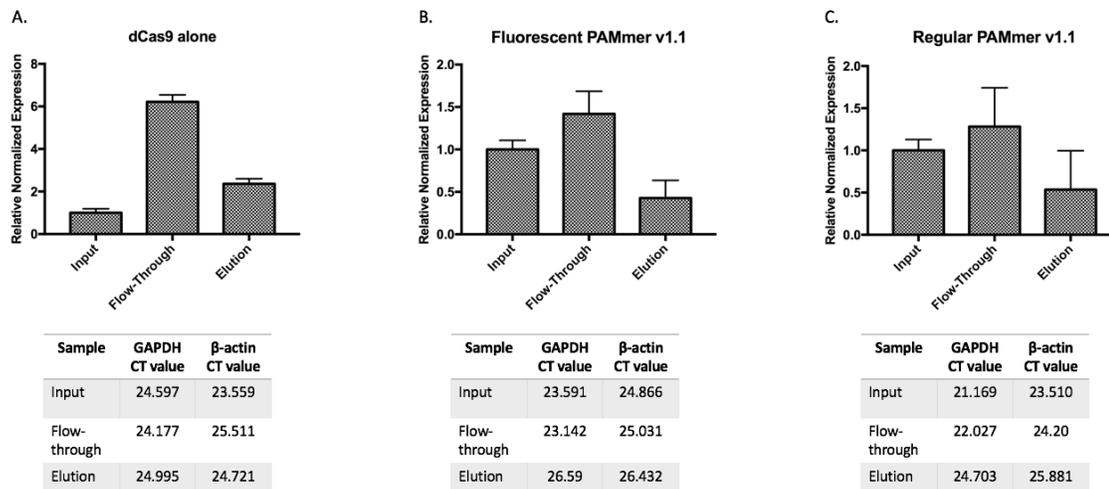


Figure 4.6 *GAPDH mRNA enrichment by in vivo dCas9 system*

A stable HEK 293FT cell-line expressing halo-dCas9 was infected with either gRNA 1 and PAMmer v1.1 without a tag or gRNA 1 and 5' 6-FAM tagged PAMmer v1.1 or nothing. Cells were cross-linked with formaldehyde, lysed and incubated with Halo-LinkTM resin. Halo-dCas9 was eluted by treating with TEV protease overnight. Input, flow-through and elution fractions from the enrichment assay were probed for GAPDH mRNA by RT-qPCR. Top half is the relative expression plots of GAPDH mRNA compared to the reference, β -actin mRNA. Bottom half is the raw CT values from RT-qPCR. Error bars are SEM.

CHAPTER 5

CONCLUDING REMARKS AND FUTURE DIRECTIONS

5.1 The use of specific mild and potent detergents is incompatible with the *in vitro* dCas9 system

Our initial results from the GAPDH mRNA enrichment by the *in vitro* dCas9 system showed ~6-fold enrichment of GAPDH mRNA with gRNA 1 and PAMmer v1.1., which was much smaller than GAPDH mRNA enriched by RAP. CT values showed dCas9 also captured and retained non-target RNA species, specifically β -actin mRNA. In an attempt to remove non-specifically bound RNA, we tested different, well-known IP wash conditions. This included the use of both very mild or potent detergents in varying concentrations as well as different salt concentrations, which have been shown to help reduce non-specific binding (Yang, Zhang, & Bruce, 2009). Even with the use of higher amounts of very weak detergents, the *in vitro* dCas9 system resulted in a loss of both the target GAPDH mRNA and β -actin mRNA.

Despite dCas9 and gRNA binding to the target RNA through hydrogen bonding, the purified dCas9 protein may not favor certain types and/or amounts of detergents. Specifically, it is possible even with mild amounts of SDS and NP-40, the purified dCas9 can become partially or completely unfolded (Bhuyan, 2010). Thereby resulting in a non-functional protein that becomes dissociated from the gRNA and can no longer survey and enrich RNA. If the *in vitro* dCas9 system is to be used as an RNA-RBP enrichment tool, the capturing of non-specific RNA and lack of flexibility to different buffer conditions are

problematic. For example, though GAPDH and β -actin mRNA have information available about RBPs bound to the target mRNA, making the MS analysis slightly more specific, if we were to focus on a lesser known target RNA, it would be difficult to decipher if the RBPs identified are in fact associated with the target or are just an anomaly. With RAP, this is not a concern since it is a probe-based method. In this case, therefore, RAP-MS would be a more specific and efficient RNA-RBP enrichment tool.

5.2 UV Stabilization of dCas9 complexes may enhance capturing of non-specific RNA

Due to the ease by which GAPDH mRNA was lost with a change in detergents and buffer composition, we attempted to stabilize the fragile complex formed between the dCas9-gRNA-PAMmer and the target RNA by introducing covalent bonds through UV cross-linking. With UV cross-linking, regardless of if the protein is denatured, it will still be linked to the target RNA. Thus, in principle, the target RNA could be captured and stringent wash conditions can be used. When using UV cross-linking and the NP-40 buffer conditions, we found higher UV intensities did help enrich slightly more GAPDH mRNA compared to non cross-linked samples. However, this did not result in a considerable fold change in the elution fraction. This is because there was still retention of β -actin mRNA, though we did find that higher UV intensities did not lead to higher retention of β -actin mRNA. In addition, there was loss of GAPDH mRNA to the flow-through fraction. This may be as a result of the dCas9 complex disassociating when samples are on ice waiting to reach the target UV intensity.

With UV cross-linked samples, more stringent wash conditions (for example, the

ChIP buffers with SDS) would not help minimize and/or eliminate the retention of non-specific RNA. This is because a concern with UV cross-linking is when dCas9 is in very close proximity or even binds to non-specific RNA. In this case, dCas9 would be linked to that RNA (Poria & Ray 2017). In a sample with total RNA or even cell lysate, the likelihood of dCas9 being ‘close’ to non-specific RNA is high. Thus, even with more stringent wash conditions, we could still capture that non-specific RNA and non-specific RBPs. Moreover, our results so far indicate that dCas9 is binding to β -actin mRNA, leading to its retention. Presumably, UV cross-linking would only act to strengthen that interaction. Therefore, the use of stringent wash conditions would not effectively eliminate β -actin mRNA if samples are UV cross-linked.

5.3 Assessing the RNA-specificity of the *in vitro* dCas9 system

Through our work with the the *in vitro* dCas9 system, we have consistently showed that though our dCas9-gRNA-PAMmer system may preferentially bind to GAPDH mRNA, there was a portion of the dCas9 also binding to and retaining β -actin mRNA. This indicates promiscuous binding by dCas9 to RNA. However, it was unclear whether this capturing effect was specific to β -actin mRNA or true for all RNA. During RT-qPCR, ‘relative’ expression or fold enrichment is determined by the choice of a reference gene. The reference gene should not be enriched nor detected in the elution fraction of the enrichment assay, with CT values typically ranging from 38-40. For example, in RAP, the CT values of the reference β -actin mRNA during GAPDH mRNA enrichment were approximately around 30-32 in the elution, while GAPDH mRNA is ~20-22, suggesting β -actin mRNA

was not significantly captured. Since our current reference gene was detected at high amounts in the elution fraction for the *in vitro* dCas9 system, we wondered whether a change in reference gene would result in a similar detection and enrichment of that RNA.

Upon changing the reference gene to U2 snRNA, we found there was a large loss of GAPDH mRNA to the flow-through and wash fraction, though there was slightly ‘more’ GAPDH mRNA detected in the elution fraction when compared to U2 snRNA. Interestingly, when comparing β -actin mRNA to U2 snRNA in these fractions, there was a very similar pattern of capturing followed. It also appeared as if β -actin mRNA was preferentially captured, as less β -actin was detected in the flow-through and more was detected in the elution fraction.

Changing the reference gene for our RT-qPCR analysis revealed two major concerns with using the dCas9 system to enrich RNA and its RBPs. For one, there may be competition for substrate binding between GAPDH and β -actin mRNA for dCas9. Specifically, β -actin mRNA may act as a ‘competitive’ inhibitor, thereby occupying the dCas9 binding site and making it unavailable for the actual target, GAPDH mRNA (Berg, Tymoczko, & Stryer, 2002). This could help explain the large loss of GAPDH mRNA to the flow-through fraction and the little fold change detected in our initial results for the *in vitro* dCas9 system. It is also possible that β -actin mRNA is not the only competitive inhibitor. We have tested only two reference genes, therefore, it is possible there are other RNAs exhibiting similar binding and capturing as β -actin mRNA. This would further limit the capturing and enrichment of GAPDH mRNA. Adding more dCas9 would not alleviate this issue, as it may act to provide more enzyme binding sites to non-specific RNA.

The other issue is specificity. Changing the reference gene suggested there was a high non-specific signal in our elution fraction with β -actin mRNA and potentially, other RNAs to a higher degree as well. As mentioned, in terms of RBP enrichment, this would be unfavorable as it would be difficult to determine whether the identified protein candidates are true or false positives.

5.4 Assessing the gRNA and PAMmer specificity to GAPDH mRNA

There are three points of contention when assessing the RNA-specificity of the dCas9 system, the dCas9, gRNA and PAMmer. The gRNAs and PAMmers (**Table 3.1** and **Table 3.2**) were designed specific to the GAPDH mRNA sequence. Furthermore, it has been suggested by O'Connell et al. (2014) and in studies with DNA, there must be specific placement of the PAM upstream of the target sequence, in order for the target sequence to be unwound, interrogated by sgRNA and targeted for cleavage. Each gRNA and PAMmer pair were designed as such, with PAMmer being slightly upstream and overlapping the target GAPDH mRNA site. A quick BLAST search between β -actin mRNA and each gRNA and PAMmer sequence reveals there are no similarities detected.

As previously mentioned, the gRNA can be divided into two regions, the seed and non-seed sites. It has been shown that the 'mismatches' in the gRNA are tolerated on a positional basis. For example, the non-seed (PAM-distal) regions can tolerate mismatches, which leads to off-target cleavage events. Even within non-seed regions, there are positional effects, with 1-5 nucleotide mismatches tolerated at the 5' region, while only 1-2 single-nucleotide mismatches may be tolerated closer to the seed regions, with even two

mutations in this location resulting in a less 5% cleavage efficiency (Anderson et al., 2015). The seed region is very stringent and does not tolerate mismatches (Wu et al., 2015), though there has been debate about the ‘true’ classification of the seed region. Traditionally, the seed region was classified as 12 nucleotides in PAM-proximal region (Wu et al., 2015). However, recent studies with CHIP-seq of DNA bound to dCas9 in murine embryonic stem cells suggest there is only a 5 bp PAM-proximal region matched with the immunoprecipitated DNA, indicating that this is the ‘true’ seed region (Wu et al., 2015). Taking the 5 bp PAM-proximal region in gRNA 1, would we find there is a ‘match’ along both β -actin and GAPDH mRNA (**Figure 5.1 A/B**). Though there is also a match for U2 snRNA, it does have a single nucleotide mismatch (**Figure 5.1 C**). Alignment of the 5 bp seed region for gRNA 2 and gRNA 3, show a similar match to both both β -actin and GAPDH mRNA. Interestingly the 5 bp seed region of gRNA 3 also matches U2 snRNA, whereas there is the 5 bp seed region of gRNA 2 does not (only 2 bp is matched). If the seed region for dCas9 to RNA is also in fact 5 bp, this would explain the potential capturing of β -actin mRNA but the lack of U2 snRNA targeting despite the non-seed region being mismatched for both. In addition, if true, this would further solidify the inability to use the dCas9 system as an enrichment tool, since there is a high likelihood of selecting a 5 bp seed sequence that would match with other RNA. This is confirmed via a BLAST search that shows several hundred targets that contain the seed region of the GAPDH gRNAs.

A. β -actin mRNA GCAAGAGAGGCATCCTCACCCCTGAAGTACCCCATCGAGC
5 bp seed -----ACCCT-----

B. GAPDH mRNA CAAGAGCACAAAGAGGAAGAGAGAGACCCTCACTGCTGGGG
5 bp seed -----ACCCT-----

C. U2 snRNA CATCGACCTGGTATTGCAGTACCTCCAGGAACGGTGCACCC-
5 bp seed -----ACCCT

Figure 5.1 Alignment of 5 bp seed region of gRNA 1 and β -actin mRNA, GAPDH mRNA and U2 snRNA to determine sequence similarity

A. β -actin mRNA CCCCAAGGCCAACCGCGAGAAGATGACCCAGATCATGTTTGAGACCTTCAA
5 bp seed -----ATGAC-----

B. GAPDH mRNA TGGTAAAGTGGATATTGTTGCCATCAATGACCCCTTCATTGACCTCAACT
5 bp seed -----ATGAC-----

C. U2 snRNA ATCGCTTCTCGGCCTTTTGGCTAAGATCAAGTGTAGTATCTGTTCTTATC
5 bp seed AT-----

Figure 5.2 Alignment of 5 bp seed region of gRNA 2 and β -actin mRNA, GAPDH mRNA and U2 snRNA to determine sequence similarity



Figure 5.3 Alignment of 5 bp seed region of gRNA 3 and β -actin mRNA, GAPDH mRNA and U2 snRNA to determine sequence similarity

In context of Cas9 and dsDNA, PAM recognition proceeds sgRNA interrogation. The initial tests with the *S. Pyogenes* Cas9 system in context of RNA cleavage and/or binding showed that this system is largely dependent on the addition of a PAMmer. Specifically, cleavage would not occur without the addition of a PAMmer and, in terms of dCas9, PAMmer activated dCas9 bound nearly >500-fold tighter than dCas9 alone (O’Connell et al., 2014). However, it is interesting to note that the PAMmer sequences, may not be required to initiate Cas9 binding to ssRNA (Strutt, Torrez, Kaya, Negrete & Doudna, 2018). For enrichment purposes, however, this would entail that dCas9 could bind to other RNAs that match the 5 bp seed region of a gRNA, even if the PAMmer was intended to target a specific RNA. In addition, it is interesting to note that directly upstream of seed region binding site for β -actin mRNA there is an endogenous ‘NCC’ sequence for all the gRNAs. Though it has been suggested that this is insufficient to induce cleavage for ssRNA targets, it has not been tested whether or not this could still elicit binding by dCas9 to RNA. It’s possible that dCas9 may temporarily bind to this site and unbind once it

recognizes it is not dsDNA.

5.5 Cleavage versus binding specificity of Cas9

Cas9 has traditionally been referred to as a highly specific enzyme. However, it is important to distinguish between Cas9's cleavage specificity and its binding specificity. In terms of dsDNA, Cas9 acts first as a surveying enzyme. It will survey the genome in search of sequences that contain the correct PAM, initiate duplex melting and if sgRNA confirms the target is correct, then it will initiate cleavage (Jiang & Doudna, 2017). In this traditional model, Cas9 requires extensive base-pairing between dsDNA and sgRNA in order for binding and cleavage to occur (Wu et al., 2015). Results from ChIP-seq with dCas9, however, suggest two state model of Cas9 (Wu et al., 2015). In this model there two bound states for Cas9, one in which it is bound to the PAM and the 5 bp seed region and one in which it is bound to the PAM and extensively to the DNA target. The later bound state is required for cleavage initiation. This is supported by cleavage assays where mutations in both the seed and non-seed regions of the sgRNA led to less extensive base-pairing with the dsDNA target and dramatically reduced Cas9 cleavage efficiency (Wu et al., 2015).

The ChIP-seq studies with dCas9 reveal that though Cas9 exhibits high cleavage specificity, its binding to dsDNA is less specific. If the PAM and 5 bp seed region is correct, then Cas9 will remain bound to that dsDNA region (Wu et al., 2015). In terms of our *in vitro* dCas9 system, this could help explain the retention of β -actin mRNA. As previously indicated, the 5 bp seed region does match β -actin mRNA. In addition, it has recently been shown PAMmers may not be required to initiate binding between Cas9 and ssRNA targets

(Strutt et al, 2018). Therefore, the combination between the 5 bp seed region and lack of need for the PAMmer may initiate the first bound state of dCas9 and lead to non-specific retention of β -actin mRNA.

5.6 The effects of mutations introduced into the *in vitro* dCas9 protein

The *S. Pyogenes* dCas9 protein purified for the *in vitro* enrichment assays contain several different mutations. Notably, there is the D10A and H840A mutations, which inactivate the RuvC and HNH cleavage domains respectively (O’Connell et., 2014). It is unlikely that these mutations would affect Cas9 specificity and binding, as they have been well-studied and tested, for example, in ChIP-seq, enrichment of dsDNA targets and live-imaging (Raha et al., 2010; Liu et al., 2017; Nelles et al., 2016). In addition to the inactivation of the cleavage domains, O’Connell et al. (2014) have mutated naturally occurring cysteine (C) residues to serine (S) (C80S and C574S) and introduced a cysteine point mutant at methionine 1 (C1M) to permit specific biotin labelling. The C to M mutation occurs in the cleavage domain (RuvC), while the C to S mutations both occur in the REC lobe. However, neither C to S mutation participates or acts to support the binding of sgRNA and/or the dsDNA (Nishimasu et al., 2014). Moreover, both C and S are polar and have nearly identical structures, the only difference is that sulfur atom in C is replaced with an oxygen atom in S. This further suggests that the C to S mutations likely have little effect on changing the folding dynamics and functionality of dCas9.

5.7 *In vivo* dCas9 system fails to capture the majority of GAPDH mRNA

With our *in vivo* dCas9 system, we successfully established a stable HEK 293FT cell-line expressing halo-dCas9 and conditions to express gRNA and PAMmers in live cells. When we attempted to target GAPDH mRNA, however, we saw no enrichment of GAPDH. CT values also suggest, there is higher amount of β -actin mRNA retained in the elution fraction than preferred. The control group, with only dCas9, did have some retention of GAPDH mRNA. The retention of β -actin mRNA is not shocking, as the gRNA sequences used were consistent between the *in vitro* and *in vivo* approaches. If the 5 bp seed region effect is true, it is likely that β -actin mRNA binds to and is retained by dCas9. The system also uses formaldehyde cross-linking to link any complexes formed with halo-dCas9, further supporting the retention of both GAPDH and β -actin mRNA.

Though there was not a substantial enrichment of β -actin mRNA, it would be interesting to probe for other RNAs in the elution fraction to further elucidate the specificity of this approach. In addition, the *in vivo* method could also be further optimized, for example, there may be a lack of binding between halo-dCas9 and Halo-LinkTM resin. However, since the halo-dCas9 pull-down efficiency was optimized by a western blot, it is unlikely that this is the case. In addition, the current *in vivo* dCas9 protocol collects all cells after transfection with the gRNA and PAMmer. As our results from PAMmer transfection indicate, transfection efficiency is not 100%. Therefore, the cells collected are a mixture of those with only halo-dCas9, with halo-dCas9 and gRNA, halo-dCas9 and PAMmer and halo-dCas9, gRNA and PAMmer. Thus, it may be worth sorting and selecting for the cells that contain all three CRISPR components prior to pull-down. This would help reduce the

background noise from other cells which do not contain all the necessary components.

For optimization, it is possible to re-perform the assay once again, with either longer binding time or the use of more beads and the addition of cell-sorting. But, this is extremely costly. For example, the Lipofectamine RNAiMAX reagent is a necessary component for the pull-down, as the PAMmers cannot be delivered to the cells otherwise. Its retail price, however, is quite high for even small amounts (\$750.00 per 750 μ L). In comparison to RAP, this would not be practical for large-scale MS experiments, which would require potentially 3-4 vials of Lipofectamine RNAiMAX per gRNA group and control. In addition, though 'halo' binds stronger to its ligand than biotin, large amounts of the Halo-Link™ resin are required to perform each pull-down. For a MS experiment, 2 mL of Halo-Link™ resin would be required per gRNA and control group. Each 10 mL vial of Halo-Link™ resin retails for \$470.00. Although it is true that cost can be drawback for most experiments, in this case, cost and practicality become vital concerns when the specificity of the approach is largely questionable.

5.8 Future directions

In order to use either the *in vitro* and *in vivo* dCas9 approach to target specific RNA and its RBPs, it is first vital to perform a CLIP-seq experiment that targets dCas9 and its bound RNA to determine 1) the exact seed region which dCas9 uses to bind to target RNA and 2) the surveying nature of dCas9 to RNA. Our results suggest, even without a completely matched gRNA or PAMmer, dCas9 can bind to and retain RNA. Therefore, it would be interesting to perform CLIP-seq on four groups: dCas9 alone, dCas9 and gRNA

only, dCas9 and PAMmer only and dCas9 with both the gRNA and PAMmer. Sequencing analysis would reveal whether there are individual, non-specific effects of the PAMmer and/or gRNA on dCas9 targeting. . Our stable HEK 293FT cell-line expressing halo-dCas9 can be used to perform those experiments, as the dCas9 can be targeted with an antibody to the halo-tag or the dCas9 itself and can be easily manipulated to express different gRNAs and PAMmers. In the meantime, we could also optimize for PAMmer binding to aid with specificity in our *in vitro* assays, either by using a PAMmer with higher affinity (locked nucleic acids probes) or using a heating then cooling approach to promote annealing between the ssRNA and PAMmer.

Though we have been able to identify suitable conditions for *in vitro* system that achieve a ~6-fold GAPDH mRNA enrichment, if the seed region is indeed 5 bp or less, as with the ChIP-seq experiments, then the dCas9 system would not be a suitable method for RNA-RBP enrichment. As mentioned, there is a high likelihood that other RNA sequences will carry a similar 5 bp sequence, making this approach highly non-specific. Even if the seed region was 10-12 bp, as in the traditional model, this sequence is still too small to elicit high affinity to the target RNA. It would be interesting to see the results of the CLIP-seq experiments with the 'gRNA only group, since the lack of a PAMmer appears to still result in targeting of ssRNA. If, as predicted, this promotes non-specific capturing of RNA, then it would further indicate this is an unsuitable approach. It would also be interesting to alter the seed region of the gRNA, to determine if this is the reason for off-target enrichment events by dCas9.

Though Cas9 appears to be limited in its binding specificity to RNA, there are

potentially other CRISPR/Cas systems that could be used to target RNA and its RBPs more precisely. For example, in 2018, Type II-A and II-C variants of Cas9 were discovered to target and cleave ssRNA independent of a PAMmer sequence (Strutt, et al., 2018). These variants were programmable and sequence specific, though it would be interesting to assess the exact binding dynamics and seed regions of such variants. In addition, we are currently working on targeting RNA with Cas13a, a newly discovered RNA-guided RNA cleaving enzyme. Cas13a does not share homology with previously classified Cas proteins; it has two HEPN domains, which critical for cleaving RNA substrates, consistent with the role of HEPN domains in other proteins (Abudayyeh et al., 2017). Cas13a is known to target specific RNA species through a short crRNA sequence of 24 nucleotides, that interacts with Cas13a through a uracil rich stem-loop structure (Cox et al., 2017). Cas13a also tolerates only single mismatches to the target RNA, whereas 2 or more mismatches dramatically reduces cleavage efficiency (Cox et al., 2017). The Cas13a has its own PAM-like sequence, referred to as a protospacer flanking sequence (PFS) that is found on the 3' end of the spacer sequence and contains either a single adenosine, uracil or cytosine (Cox et al., 2017). De-activated variants of Cas13a have been established, capable of targeting RNA when supplemented with the appropriate crRNA (Abudayyeh et al., 2017). One major point of contention with Cas13a is cleavage versus binding specificity. It would be important to assess the seed region here as well and probe for other genes during RT-qPCR analysis.

5.9 Concluding remarks

This body of work has allowed us to comprehensively examine and assess the RNA-specificity of the dCas9 system as a tool for RNA-RBP enrichment, which has not been previously determined before. We developed and/or optimized two approaches to target RNA-RBPs using dCas9, the *in vitro* and the novel *in vivo* approach. Through our optimization, we found ~6-fold enrichment of GAPDH mRNA with the *in vitro* approach, though there was some background signal of non-specific RNA. Probing for other target RNA suggested that dCas9 could potentially capture RNA with similar 5 bp seed sequences. Therefore, we have critically established that there can be promiscuous binding by dCas9 to RNA targets. We also have conclusively determined, in comparison to RAP, this tool cannot tolerate denaturing conditions, thereby limiting the potential to alleviate background signals. Overall, our findings establish critical steps that all RNA-RBP enrichment tools should use when determining the specificity to their approach, namely probing for multiple reference genes in the elution and in terms of the CRISPR/Cas system, determining the seed region. This can potentially aid in saving time and cost with MS and in determining true and specific RBP candidates for downstream analyses.

REFERENCES

- Abudayyeh, O.O., Gootenberg, J.S., Esselzsbichler, P., Han, S., Joung, J., Belanto, J.J., ... Zhang, F. (2017). RNA-targeting with CRISPR-Cas13a. *Nature*, *550*(7675), 280-284. doi: 10.1038/nature24049.
- Anderson, E.M., Haupt, A., Scheil, J.A., Chou, E., Machado, H.B., Strezoska, Z., ... Brabant-Smith, A.V. (2015). Systemic analysis of CRISPR-Cas9 mismatch tolerance reveals low levels of off-target activity. *Journal of Biotechnology*, *211*, 56-65.
- Bann, D.V., & Parent, L.J. (2012). Application of live-cell RNA imaging techniques to study the retroviral RNA trafficking. *Viruses*, *4*(6), 963-979.
- Bertrand, E., Chartrand, P., Schaefer, M., Shenoy, S.M., Singer, R.H., & Loop, R.M. (1998). Localization of ASH1 mRNA particles in living yeast. *Molecular Cell*, *2*(4), 437-445.
- Berg, J.M., Tymoczko, J.L., & Stryer, L. (2002). Section 8.5 Enzymes Can Be Inhibited by Specific Molecules. *Biochemistry: 5th edition*. Houndsmills, New York: W.H. Freeman & Company.
- Bhuyan, A.K. (2010). On the mechanism of SDS-induced protein denaturation. *Biopolymers*, *93*(2), 186-199.
- Catherman, A.D., Skinner, O.S., & Kelleher, N.L. (2015). Top down proteomics: Facts and perspectives. *Biochemical and Biophysical Research Communications*, *445*(4), 683-693.

- Chu, C., Zhang, Q.C., da Rocha, S.T., Flynn, R.A., Bharadwaj, M., Calabrese, J.M., ... Chang, H.Y. (2015). Systemic discovery of *Xist* RNA binding proteins. *Cell*, *161*(2), 404-416.
- Chu, C., Spitale, R.C., & Chang, H.Y. (2015). Technologies to probe functions and mechanisms of long non-coding RNAs. *Nature Structural and Molecular Biology*, *22*, 29-35. doi:10.1038/nsmb.2921.
- Cox, D.B., Gootenberg, J.S., Abudayyeh, O.O., Franklin, B., Kellner, M.J., Jong, J., & Zhang, F. (2017). RNA-editing with Cas13a. *Science*, *358*(6366), 1019-1027.
- Daigle, N., & Ellenberg, J. (2007). $[[\lambda]]N$ -GFP: an RNA reporter system for live-cell imaging. *Nature Methods*, *4*(8), 633-636.
- Demidov V.V., & Frank-Kamenetskii M.D. (2004). Two sides of the coin: affinity and specificity of nucleic acid interactions. *Trends in Biochemical Science*, *29*(2), 62–71.
- Deepak, S.A., Kottapalli, K.R., Rakwal, R., Oros, G., Rangappa, K.S., Iwahashi, H., ... Agrawal, G.K. (2007). Real-time PCR: Revolutionizing detection and expression analysis of genes. *Current Genomics*, *8*(4), 234-251.
- Diamandis, E.P., & Christopoulos, T.K. (1991). The biotin-(strept)avidin system: Principles and applications in biotechnology. *Clinical Chemistry*, *37*(5), 635-636.
- Dominguez, A.A., Lim, W.A., & Qi, L.S. (2016). Beyond editing: Re-purposing the CRISPR-Cas9 system for genome regulation and interrogation. *Nature Reviews Molecular Cell Biology*, *(1)5*, 5-15.

- Gallgher, S., & Sass, J. (2001). Protein analysis by SDS-PAGE and detection by coomassie blue or silver staining. *Current Protocols in Pharmacology*, 3(3B). doi:10.1002/0471141755.pha03bs02.
- Gebauer, F., Preiss, T., & Hentze, M.W. (2012). From cis-regulatory elements to complex RNPs and back. *Cold Spring Harbor Perspectives in Biology*, 4(7), 1-14. doi: 10.1101/cshperspect.a012245.
- Hafner, M., Landthaler, M., Burger, L., Khorshid, M., Hausser, J., Berninger, P., ... Tuschl, T. (2010). PAR-CLIP- A method to identify transcriptome-wide the binding sites of RNA binding proteins. *Journal of Visualized Experiments*, 2034(41). doi: 10.3791/2034.
- Harvey, R.F., Smith, T.S., Mulrone, T., Querioz, R.M.L., Pizzinga, M., Dezi, V., ... Willis, A.E. (2018). *Trans-acting translational regulatory RNA binding proteins. Wiley Interdisciplinary Reviews: RNA*, 9(3), 1-19. doi: 10.1002/wrna.1465.
- Jiang, F., & Doudna, J. (2017). CRISPR-Cas9 structures and mechanisms. *Annual Review of Biophysics*, 46, 505-529. doi: 10.1146/annurev-biophys-062215-010822.
- Jinek, M., Chylinski, K., Fonfara, I., Hauer, M., Doudna, J., & Charpentier, E. (2012). A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science*, 337(6096), 816–821.
- King, H.A., Cobbold, L.C., Pichon, X., Poyry, T., Wilson, L.A., Booden, H., ... Willis, A.E. (2014). Remodeling of a polypyrimidine tract-binding protein complex during apoptosis activates cellular IRES. *Cell Death & Differentiation*, 21(1), 161-171. doi: 10.1038/cdd.2013.135.

- König, J., Zarnack, K., Rot, G., Curk, T., Kayikci, M., ... Ule, J. (2010). iCLIP reveals the function of hnRNP particles in splicing at individual nucleotide resolution. *Nature Structure & Biology*, *17*(7), 909-915. doi: 10.1038/nsmb.1838.
- Lango, P.A., Kavran, J.M., Min-Sung, K., & Leahy, D.J. (2013). Transient mammalian cell transfection with polyethylenimine (PEI). *Methods in Enzymology*, *529*, 227-240. doi:10.1016/B978-0-12-418687-3.00018-5.
- Liu, T., Pan, S., Li, Y., Peng, N., & She, Q. (2018). Type III CRISPR-Cas systems: Introduction and its application for genetic manipulations. *Current Issues in Molecular Biology*, *26*, 1-14. oi: 10.21775/cimb.026.001.
- Liu, X., Zhang, Y., Chen, Y., Li, M., Zhou, F., Li, K., ... Xu, J. (2017). In situ capture of chromatin interactions by biotinylated dCas9. *Cell*, *170*(5), 1028-1043.
- Lodish, H., Berk, A., Ziprusky, S.L., Matsudaria, P., Baltimore, D., & Darnell, J.E. (2000). From gene to protein: Basic molecular genetic mechanisms. In S. Tenney, K. Ahr, R. Steyn, K. Ueno (Eds.), *Molecular Cell Biology: 4th edition* (pp. 119-125). Houndsmills, New York: W.H. Freeman & Company.
- Mahmood, T., & Yang, P.C. (2012). Western blotting: Technique, theory and trouble shooting. *Northern American Journal of Medicine and Science*, *4*(9), 429-434.
- Machyna, M., & Simon, M.D. (2018). Capturing RNAs using hybridization capture methods. *Briefings in Functional Genomics*, *17*(2), 96-103.
- McHugh, C. A., Chen, C.K., Chow, A., Surka, C.F., Tran, C., McDonel, P., ...Guttman, M. (2015). The Xist lncRNA directly interacts with SHARP to silence transcription through HDAC3. *Nature*, *521*(7551), 232–236.

- Mendelsohn, J., Howley, P., Isarel, M., Gray, J., & Thompson, C. (2004). Chapter 22-Mass spectrometry and Cancer. *The Molecular Basis of Cancer: 4th edition* (pp. 293-207). Amsterdam, Netherlands: Elsevier Inc.
- Morris, K.V., & Mattick, J.S. (2014). The rise of regulatory RNA. *Nature Reviews Genetics*, *15*(6), 423-437. doi: 10.1038/nrg3722.
- Nishimasu, H., Ran, F.A., Hsu, P.D., Konermann, S., Shehata, S.I., Dohmae, N., ... Nureki, O. (2014). Crystal structure of Cas9 in complex with guide RNA and the target DNA. *Cell*, *156*(5), 935-949.
- Nelles, D.A., Fang, M.Y., O'Connell, M.R., Xu, J.L., Markmiller, S.J., Doudna, J.A., & Yeo, G.W. (2016). Programmable RNA tracking in live cells with CRISPR/Cas9. *Cell*, *165*(2), 488-496.
- O'Connell, M.R., Oakes, B.L., Sternberg, S.H., Seletsky A.E., Kaplan, M., & Doudna, J.A. (2014). Programmable RNA recognition and cleavage by CRISPR/Cas9. *Nature*, *516*(7530), 263-266.
- Park, S.A., Ahn, S.I., & Gallo, J.M. (2016). Tau mis-splicing in the pathogenesis of neurodegenerative disorders, *BMB Reports*, *49*(8), 405-413.
- Pommer, A.J., Cal, S., Keeble, A.H., Walker, D., Evans, S.J., Kuhlmann, U.C., ... Kleanthous, C. (2001). Mechanism and cleavage specificity of the H-N-H endonuclease colicin E9. *Journal of Molecular Biology*, *314*(4), 735–749.
- Poria, D.K., & Ray, P.S. (2017). RNA-protein UV cross-linking assay. *Bio-Protocol*, *7*(6). doi: 10.21769/BioProtoc.2193.

- Raha, D., Hong, M., & Snyder, M. (2010). ChIP-seq: A method for global protein identification of regulatory elements in the genome. *Current Protocols in Molecular Biology*, 21(19), 1-14.
- Rath, D., Amlinger, L., Rath, A., & Lundgren, M. (2015). The CRISPR-Cas immune system: Biology, mechanisms and applications. *Biochimie*, 117, 119-128. doi: 10.1016/j.biochi.2015.03.025.
- Rinn, J.L., & Chang, H.Y. (2012). Genome regulation by long non-coding RNAs. *Annual Review of Biochemistry*, 81, 144-166. doi: 10.1146/annurev-biochem-051410-092902.
- Sexton, A.N., Maychna, M., & Simon, M.D. (2016). Capture hybridization analysis of RNA targets (CHART). *Current Protocols in Molecular Biology*, 21(25). doi: 10.1002/0471142727.mb2125s101.
- Shepard, P.J., & Hertel, K.J. (2009). The SR protein family. *Genome Biology*, 10(10), 242. doi: 10.1186/gb-2009-10-10-242.
- Shyu, A.B., Willikinson, M.F., & van Hoof, A. (2008). Messenger RNA regulation: To translate or to degrade. *EMBO Journal*, 27(3), 471-481. doi: 10.1038/sj.emboj.7601977.
- Stork, C., & Zheng, S. (2016). Genome wide profiling of RNA-protein interactions using CLIP-seq. *Methods in Molecular Biology*, 1421, 137-151. doi: 10.1007/978-1-4939-3591-8_12.
- Strutt, S.C., Torrez, R.M., Kaya, E., Negrete, O.A., & Doudna, J.A. (2018). RNA-dependent RNA targeting by CRISPR-Cas9. *eLife*, 5(7), doi: 10.7554/eLife.32724.

- Summer, H., Gramer, R., & Droge, P. (2009). Denaturing urea polyacrylamide gel electrophoresis (Urea PAGE). *Journal of Visualized Experiments*, 29(32), 1485. doi: 10.3791/1485.
- Sundberg, E.J. (2009). Structural basis of antibody-antigen interactions. *Methods in Molecular Biology*, 524, 23-36. doi: 10.1007/978-1-59745-450-6_2.
- Stepanov, G.A., Filippova, J.A., Komissarov, A.B., Kuligina, E.V., Richter, V.A., & Semenov, D.V. (2014). Regulatory role of small nucleolar RNA in human diseases. *Biomed Research International*, 2015, 1-10. doi: 10.1155/2015/206849.
- Sternberg, S.H., Redding, S., Jinek, M., Greene, E.C., & Doudna, J.A. (2014). DNA interrogation by the CRISPR RNA-guided endonuclease Cas9. *Nature*, 507(7490),62–67.
- Wang, Y., Liu, J., Huang, B., Xu, Y.M., Li, J., Huang, L.F., ... Wang, X.Z. (2015). Mechanisms of alternative splicing and its regulation. *Biomedical Reports*, 3(2), 152-158.
- Wu, X., Kriz, A.J., & Sharp, P.A. (2015). Target specificity of the CRISPR-Cas9 system. *Quantitative Biology*, 2(2), 59-70.
- Wu, X., Scott, D.A., Kriz, A.J., Chi, A.C., Hsu, P.D., Dadon, D.B., ... Sharp, P.A. (2015). Genome-wide binding of the CRISPR endonuclease Cas9 in mammalian cells. *Nature Biotechnology*, 32(7), 670-676.
- Yan, M., Kanbe, E., Peterson, L.F., Boyapati, A., Miao, Y., Wang, Y., ... Zhang, D.E. (2006). A previously unidentified alternatively spliced isoform of t(8;21) transcript promotes leukemogenesis. *Nature Medicine*, 12(4) 945–949.

- Yang, L., Zhang, H., & Bruce, J.E. (2009). Optimizing the detergent concentration conditions for immunoprecipitation (IP) coupled with LC-MS/MS identification of interacting proteins. *Analyst*, *134*(4), 755-762.
- Yang, Y.C., Di, C., Hu, B., Zhou, M., Liu, Y., Song, N., Lu, Z.J. (2015). CLIPdb: a CLIP-seq database for protein-RNA interactions. *BMC Genomics*, *16*(51), 1-8. .doi: 10.1186/s12864-015-1273-2.
- Zhang, Y., Fonslow, B.R., Shan, B., Baek, M.C., & Yates, J.R. (2013). Protein analysis by shotgun/bottom-up proteomics. *Chemical Reviews*, *114*(4), 2343-2394.