2020

Inserting dCas9 and single-guide RNAs into Drosophila using molecular cloning methods

BOSTON UNIVERSITY

SCHOOL OF MEDICINE

Thesis

INSERTING DCAS9 AND SINGLE-GUIDE RNAS INTO *DROSOPHILA* USING MOLECULAR CLONING METHODS

by

SARA NIETO

B.S., University of California Riverside, 2018

Submitted in partial fulfillment of the

requirements for the degree of

Master of Science

2020

Approved by

First Reader

Jean L. Spencer, Ph.D.
Instructor of Biochemistry

Second Reader

Martha L. Bulyk, Ph.D.
Professor of Medicine and Pathology, Harvard Medical School
Division of Genetics, Brigham & Women's Hospital
Co-Chair, Harvard Biophysics Graduate Program

DEDICATION

This is dedicated to my mom and dad; thank you for all that you have sacrificed for me to get to where I need to be. I would not have been able to get here without your endless love and encouragement. Also, to my family and friends who have provided unconditional support, even and especially when on the opposite side of the country.

ACKNOWLEDGMENTS

I would like to thank the entire Bulyk laboratory for welcoming me and teaching me more than I could have ever imagined. Thank you, Dr. Martha Bulyk, for giving me the opportunity to learn and grow under your guidance. Watching you be a leader in the research field has been inspiring and empowering. Thank you, Steve, for allowing me to work on the silencers project. You have taught me many lab techniques and skills, and more importantly, how to persevere when science fails. To Julian, thank you for help with the silencers project and for setting me up for success with your preemptive work. To Kian and Tim, thank you for allowing me to constantly pick your brain and for your patience and assistance when troubleshooting. To Sabrina, Raehoon, Luca, and Kaia, thank you for the support and help whenever I had questions, I enjoyed learning about all your projects and watching your dedication to your research. Although my time in the lab did not end as I hoped, I am thankful for our time together and will continue to keep up with your phenomenal contributions to genetics and bioinformatics.

INSERTING DCAS9 AND SINGLE-GUIDE RNAS INTO *DROSOPHILA* USING MOLECULAR CLONING METHODS

SARA NIETO

ABSTRACT

Non-coding DNA in the human genome is widely studied to investigate its effect on coding DNA and gene expression. Non-coding DNA contains cis-regulatory elements that influence transcription of genes upstream, downstream, or nearby. These regulatory elements have largely been studied as enhancers that promote the transcription of genes. To explore these regulatory elements as silencers, we chose validated bifunctional elements to study their silencing capability and their chromatin markers.

We used chromatin immunoprecipitation methods with dCas9 to target these elements using single-guide RNAs (sgRNAs). We experimented with various cloning methods to insert dCas9 into the pUAS vector. We initially planned to use the Gibson Assembly method, but after no success, we tried site-directed mutagenesis and traditional cloning with restriction enzymes. We were able to successfully insert dCas9 into the pUAS vector with traditional cloning, and we were then able to inject the construct into *Drosophila melanogaster*.

We designed sgRNAs to target desired elements of DNA that we chose to study as cis-regulatory elements. The sgRNA sequences were cloned into the pCFD5 vector and injected into another line of flies. The transgenic flies containing the pUAS/dCas9 plasmid will then be crossed with the flies containing the pCFD5/sgRNA to develop offspring that express the target elements and could undergo chromatin pulldown to

examine the bifunctional regulation of these DNA elements in cells. Results from a quantitative PCR (qPCR) assay on *Drosophila* expressing the cloned pUAS vector with dCas9 and a sgRNA for the *white* gene showed chromatin pulldown efficiency and successful transfection.

The *Drosophila* chromatin targeted by the sgRNAs will be pulled down, solubilized, and then analyzed on a western blot to screen for chromatin modifications, primarily histone modifications. We can then identify chromatin markers associated with elements when they act as silencers in the mesoderm versus when they act as non-mesodermal enhancers. We can also determine if the silencer acts by interacting with a promoter or with an enhancer to repress gene expression. If ENCODE can profile the data found in this project, the chromatin markers can act as a predictive tool for the identification of silencers.

TABLE OF CONTENTS

TITLE……...i

COPYRIGHT PAGE……...ii

READER APPROVAL PAGE……..iii

DEDICATION iv

ACKNOWLEDGMENTS v

ABSTRACT vi

TABLE OF CONTENTS viii

LIST OF FIGURES x

LIST OF ABBREVIATIONS xi

INTRODUCTION 1

Non-Coding DNA 1

History of Genetic Regulatory Elements 1

Gene Regulatory Elements That Can Enhance or Silence Gene Expression 3

Epigenetics 5

Regulatory Elements and Disease 6

Importance of Studying Regulatory Elements 7

Ways to Pull Down Genomic Elements 8

Drosophila Genome 9

Plasmid Cloning 10
The Silencers Project 13
SPECIFIC AIMS 16
METHODS 17
Getting dCas9 Into *Drosophila* DNA 17
Cloning sgRNAs 20
Sanger Sequencing 22
Chromatin Pulldown 23
RESULTS 25
Troubleshooting PCR Amplification 26
Increasing DNA Concentrations 28
Trials with Site-Directed Mutagenesis 29
More Gibson Assembly Attempts 30
Success with Traditional Cloning 31
Designing sgRNAs to Guide dCas9 to Targets 38
Ligating sgRNA Sequences Into pCFD5 39
DISCUSSION 46
Future Directions 46
REFERENCES 50
CURRICULUM VITAE 54

LIST OF FIGURES

Figure	Title	Page
1	The organization of the ENCODE Consortium.	2
2	Cis-regulatory elements and a regulatory domain in the human genome.	3
3	Basic structure of histones and chromatin.	4
4	Transcriptional regulatory elements involved in human diseases.	7
5	Plasmid cloning to target elements.	12
6	Oligonucleotide design for each sgRNA.	21
7	Schematic of the NEB standard Gibson Assembly workflow.	26
8	DNA concentrations of the dCas9 insert using different PCR polymerases.	27
9	Site-directed mutagenesis workflow.	30
10	Traditional plasmid cloning workflow.	32
11	Restriction digest gel results.	34
12	Whole plasmid sequenced pUAS vectors with dCas9 insert.	35
13	Element example from UCSC Genome Browser.	39
14	pCFD5 with tRNA-flanked guide RNAs.	42
15	sgRNA Sanger sequencing results.	43
16	qPCR results for chromatin pulldown efficiency.	44

LIST OF ABBREVIATIONS

AMP…….Ampicillin

BioTAP-XL……Chromatin Crosslinking Approach

Cas9…..……CRISPR-Associated Protein 9

ChIP-Seq……Chromatin Immunoprecipitation Sequencing

CRISPR……....Clustered Regulatory Interspaced Short Palindromic Repeats

dCas9……..…Deactivated CRISPR-Associated Protein 9

ddH_2O……...…….Distilled and Deionized Water

dH_2O……...…Deionized Water

DNase……..……Deoxyribonuclease

DNase-Seq……..……DNase I Hypersensitive Sites Sequencing

dNTP……Deoxynucleoside Triphosphate

dsDNA……..……Double-Stranded DNA

ENCODE……Encyclopedia of DNA Elements

ESCC……...Esophageal Squamous Cell Carcinoma

FAIRE-Seq…….Formaldehyde-Associated Isolation of Regulatory Elements Sequencing

FACS-Seq……..…..Fluorescence-Activated Cell Sorting Sequencing

GFP……..Green Fluorescent Protein

gRNA……..Guide RNA

IgG…….…….…….Immunoglobulin G

kb…….Kilobase

LB……...…….Lysogeny Broth

MCS……Multiple Cloning Site

NEB…….……...New England Biolabs

NHGRI……National Human Genome Research Institute

NIH……..…….National Institutes of Health

Oligo……..……...Oligonucleotide

PCR……Polymerase Chain Reaction

PEG…...……Polyethylene Glycol

PTM……..……..Post-Translational Modification

qPCR…….....Quantitative Polymerase Chain Reaction

qsFS……....Quantitative Silencer FACS-seq

RNA-Seq……RNA Sequencing

RT-PCR……...Real-Time Polymerase Chain Reaction

SAP…..……...Shrimp Alkaline Phosphatase

SDS…….Sodium Dodecyl Sulfate

sFS…….....Silencer-FACS-Seq

sgRNA…..……...…….Single-Guide RNA

SOC…….……Super Optimal Broth

SPRI……Solid-Phase Reversible Immobilization

TALE…….Transcription Activator-Like Effector

T_m…….Melting Temperature

tRNA…….……..Transfer RNA

INTRODUCTION

Non-Coding DNA

The majority of the DNA that makes up the entire human genome does not code for proteins and is known as non-coding DNA. In the 1970s, the term “junk DNA” became popular by geneticist Susumu Ohno who used the term to describe over 90 percent of the genome (Gregory, 2005). The belief that a majority of the human genome was useless was popular for many decades. However, further research in genetics showed that non-coding DNA plays a significant role in the transcription of genes. One of the essential functions of non-coding DNA includes the role of regulatory elements. Regulatory elements have a tremendous impact on coding DNA by influencing the transcription of nearby genes.

In the field of genetics research, these regulatory elements have commonly been studied as enhancers that activate the transcription of genes. However, there has not been adequate research on how these regulatory elements act as silencers to repress gene expression.

History of Genetic Regulatory Elements

Initially, many scientists thought that the DNA outside the protein-coding regions was “junk DNA” which had no known function. In the early 2000s, the National Human Genome Research Institute (NHGRI) disproved this belief when they proposed and funded the Encyclopedia of DNA Elements (ENCODE) project. The ENCODE project

analyzed a small percentage of the once termed “junk DNA” by teaming up with “35 research labs . . . (to analyze) 44 regions of the genome” (Pennisi, 2012). The project detected and analyzed functional elements with many assays, such as RNA-Seq, ChIP-Seq, DNA methylation, and DNase I hypersensitivity (Figure 1). The ENCODE project initially analyzed about one percent of the genome thought to be nonfunctional and discovered that these DNA regions play a crucial role in gene expression and regulation, which inspired further research.

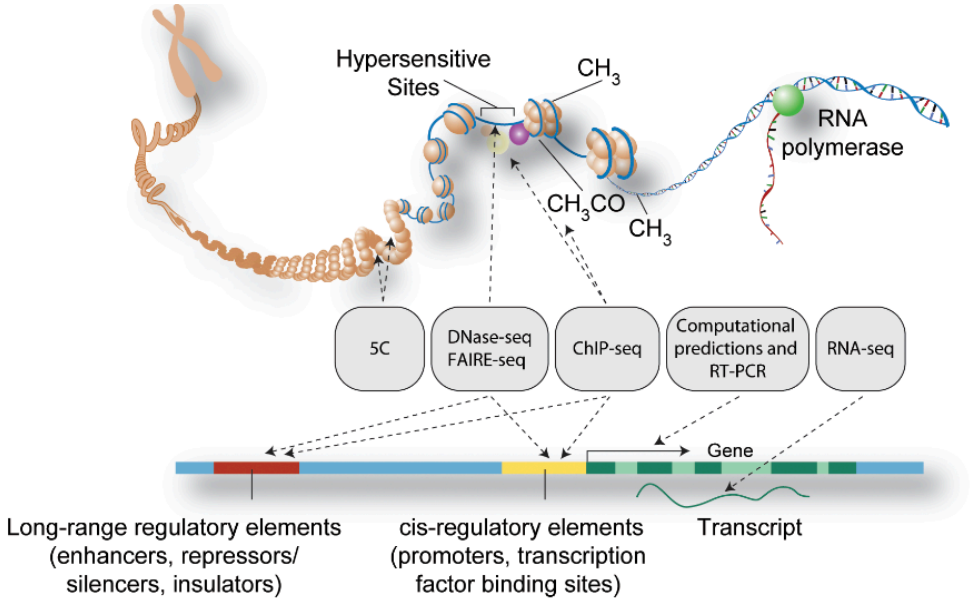

Figure 1. The organization of the ENCODE Consortium. This schematic shows the various assays and methods that were used in the ENCODE project to analyze non-coding regulatory elements. ENCODE = encyclopedia of DNA elements; ChIP-Seq = chromatin immunoprecipitation sequencing; FAIRE-Seq = formaldehyde-associated isolation of regulatory elements sequencing; DNase-Seq = DNase I hypersensitive site sequencing; RNA-Seq = RNA sequencing; RT-PCR = real-time PCR. Figure taken from (Consortium, 2011).

Gene Regulatory Elements That Can Enhance or Silence Gene Expression

Gene regulatory elements have been studied for quite some time and are becoming more popular in research as a diagnostic screening for many illnesses and diseases. It was a common belief that a majority of the human genome is not involved in gene expression. However, it has since been proved that gene regulatory elements play a role in enhancing, silencing, insulating, and promoting gene expression to various nearby regions. These regulatory elements act at specific genes by forming a loop that causes the different regions of DNA to interact (Li et al., 2015). As shown in Figure 2, the loop allows the enhancers or silencers to be near the promoter elements so that it can directly influence the transcription of genes (Li et al., 2015).

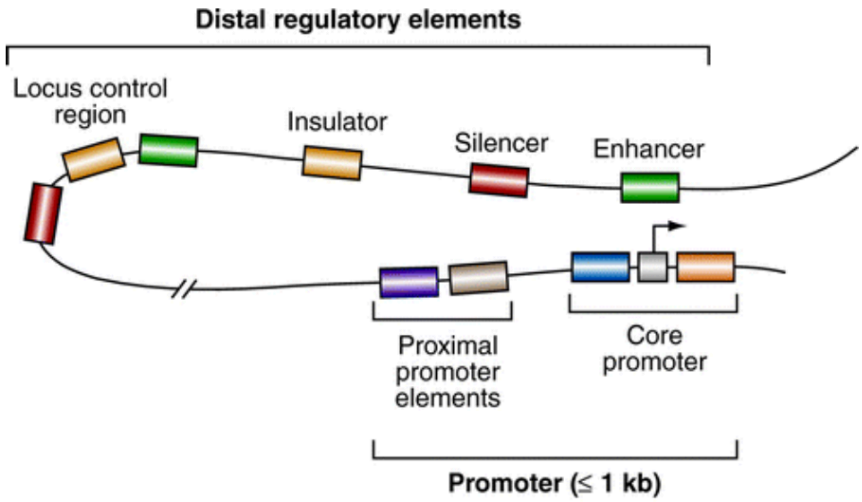

Figure 2. Cis-regulatory elements and a regulatory domain in the human genome. Enhancers and other regulatory elements acting at a promoter by forming a loop and enhancing transcription of specific genes. Figure taken from (Maston et al., 2006).

Other ways that these enhancers target genes are by localizing factors which act to recruit the RNA polymerase or by allowing the enhancer-promoter complexes to interact with the appropriate transcription compartment directly (Krivega & Dean, 2012). In comparison, regulatory elements that act as silencers have been shown to repress transcription by altering the chromatin to a repressive state, either by causing histone modifications or by preventing the formation of the preinitiation complex during transcription (Maston et al., 2006). Histones are proteins that double-stranded DNA (dsDNA) coils around to become more compact, leading to the formation of chromosomes. When the DNA is tightly coiled, it is known as heterochromatin, which is a repressive state with little or no transcription activity. When the chromatin is coiled lightly around the histones, it is in the active state, is "more accessible to transcription," and is known as euchromatin (Tamaru, 2010). Figure 3 shows the basis of histone formation.

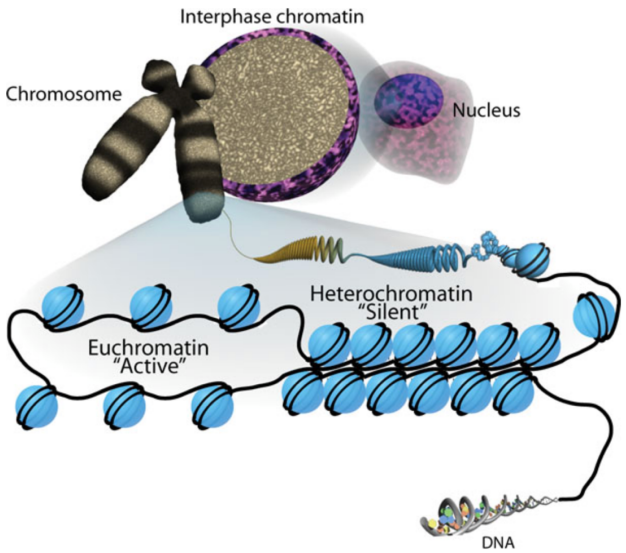

Figure 3. **Basic structure of histones and chromatin**. The blue spheres represent the histone proteins which are wrapped by the dsDNA. The tightly wrapped histones show the heterochromatin, whereas the loosely wrapped histones show euchromatin. dsDNA = double-stranded DNA. Figure taken from (Sha & Boyer, 2009).

The idea that most genes are naturally repressed and require additional modifications or factors to activate gene transcription is a common belief. These modifications are known to be influenced by the environment, and even though these modifications do not change the primary DNA sequence, they can be heritable or non-heritable (Gemenetzi & Lotery, 2020). The current research focus is to discover the factors or modifications associated with the regulatory elements and to determine whether it is possible to alter the transcription state of elements by switching from a repressed state to an active state by epigenetically modifying these regions.

Epigenetics

Epigenetics often controls the regulation or dysregulation of gene expression. The National Institutes of Health (NIH) defines epigenetics as "chemical compounds . . . added to the entirety of one's DNA (genome) as a way to regulate the activity (expression) of all the genes within the genome" (National Institutes of Health, 2020). These chemical compounds allow cells and tissues, which "contain the same genetic material," to behave differently throughout the body as a result of different gene expression (Williams, 2013). The most common types of epigenetic modifications that influence gene expression are methylation, acetylation, and histone protein modifications. The study of epigenetic modifications is vital because there are varying patterns of modifications in "different tissues within an individual" and finding commonalities in these patterns among different people with the same diseases or disorders can be

beneficial in genetic screening and the diagnosis of illnesses (National Institutes of Health, 2020).

Regulatory Elements and Disease

Several human diseases are known to be associated with regulatory elements. It is crucial to identify the genetic bases of diseases so that there can be screening methods for these illnesses. Although the diseases shown in Figure 4 are known to be caused by deletions or mutations rather than epigenetic changes, it is a notable example of how a change in regulatory elements can have a significant impact on the expression of genes.

In addition, recent studies have found that gene enhancers that lose healthy regulation, or are "hijacked," may cause the overexpression of oncogenes, which in turn can induce tumorigenesis or metastasis (Franco et al., 2018). This misregulation is another example of the crucial roles that the regulatory elements play in expression and overall cell development. By studying enhancer activity, enhancer location, transcription factors or chemical compounds involved, and target sites of these elements, researchers can determine how these enhancers cause cancer and how changes at the genetic level can activate other illnesses (Franco et al., 2018). The analysis of how these enhancers function is crucial to the understanding of diseases, such as some subtypes of breast cancer in which upstream enhancers were shown to regulate downstream genes involved in the cancer (Franco et al., 2018). Enhancers and their chromatin modifications have been well studied, but to downregulate the overexpression and misregulation of genes, the study of silencers is crucial.

Regulatory Element	Disease	Mutation (bound factor)	Affected Gene
Core promoter	β-thalassemia	TATA box, CACCC box, DCE	*β-globin*
Proximal promoter	Bernard-Soulier Syndrome	133 bp upstream of TSS (GATA-1)	*GpIbβ*
	Charcot-Marie-Tooth disease	215 bp upstream of TSS	*connexin-32*
	Congenital erythropoietic porphyria	70, 90 bp upstream of TSS (GATA-1, CP2)	uroporphyrinogen III synthase
	Familial hypercholesterolemia	43 bp upstream of TSS (Sp1)	*low density lipoprotein receptor*
	Familial combined hyperlipidemia	39 bp upstream of TSS (Oct-1)	*lipoprotein lipase*
	Hemophilia	CCAAT box (C/EBP)	*factor IX*
	Hereditary persistence of fetal hemoglobin	~175 bp upstream of TSS (Oct-1, GATA-1)	*Aγ-globin*
	Progressive myoclonus epilepsy	Expansion ~70 bp upstream of TSS	*cystatin B*
	Pyruvate kinase deficient anemia	72 bp upstream of TSS (GATA-1)	*PKLR*
	β-thalassemia	CACCC box (EKLF)	*β-globin*
	δ-thalassemia	77 bp upstream of TSS (GATA-1)	*δ-globin*
	Treacher Collins syndrome	346 bp upstream of TSS (YY1)	*TCOF1*
Enhancer	Preaxial polydactyly	1 Mb upstream of gene	*SHH*
	Van Buchem disease	Deletion ~35 kb downstream of gene	*sclerostin*
	X-linked deafness	Microdeletions 900 kb upstream	*POU3F4*
Silencer	Asthma and allergies	509 bp upstream of TSS (YY1)	*TFG-β*
	Fascioscapulohumeral muscular dystrophy	Deletion of D4Z4 repeats	4q35 genes
Insulator	Beckwith-Wiedemann syndrome	CTCF binding site (CTCF)	*H19/Igf*
LCR	α-thalassemia	62 kb deletion upstream of gene cluster	*α-globin* genes
	β-thalassemia	~30 kb deletion removing 5′HS2–5	*β-globin* genes

Figure 4. Transcriptional regulatory elements involved in human diseases. This figure shows that many diseases involving transcriptional regulatory elements have mutations in the promoter region, as well as in enhancers and silencers. Mutations in these regions can have major impact on the transcription of nearby genes. Figure taken from (Maston et al., 2006).

Importance of Studying Regulatory Elements

Both the regulation and the expression of genes play a crucial role in cell development. Misregulation, such as the overexpression or silencing of genes, can lead to various illnesses, including cancer. It is important to be able to enhance expression in specific genes to promote transcription, but it is just as essential to be able to silence

genes in order to downregulate gene expression during cell development. In this project, we focus on bifunctional regulatory elements and the associated chromatin modifications present when they act as enhancers versus silencers. Finding signature chromatin markers that are specific to tissue-specific genes or the activity state of the genes can act as a screening method for certain diseases. The chromatin modifications H3k27me3 and H3k9me2 are examples of known chromatin modifications that are used for prognostic screenings. An article published in the *International Journal of Clinical and Experimental Pathology* discusses how these chromatin modifications can "act as potential biomarkers" to screen for esophageal squamous cell carcinoma (ESCC) (Lin et al., 2019). These prognostic markers are just one example of how identifying chromatin markers for genes can have a substantial impact on medicine.

Ways to Pull Down Genomic Elements

The study of these regulatory elements and their markers requires that we perform a chromatin pulldown from the genome. One of the various methodologies used to pull down elements was first developed in 1957 and is known as the triple helix (triplex) "DNA pull-down procedure" (Isogawa et al., 2018). This method works by adding a "third strand" of DNA that is rich with pyrimidine or purine bases complementary to the "purine-rich strand in the major groove of double-stranded DNA," which forms a triplex that "allows sequence-specific recognition of DNA" that can be pulled down and studied (Isogawa et al., 2018). Another method is the chromatin crosslinking approach (BioTAP-XL), which crosslinks chromatin with formaldehyde, sonicates and solubilizes the

chromatin, and then purifies it using a BioTAP tag containing "a Protein A moiety and a (biotinylated) targeting signal" (Alekseyenko et al., 2015). The BioTAP-XL method can use transcription activator-like effector (TALE) proteins to target desired elements and analyze their tissue-specific chromatin markers (Rogers et al., 2015). In the preliminary studies, we used TALE proteins and a dCas9-sgRNA complex to pull down the silencer element. The method we are using in this study is the CRISPR (clustered regularly interspaced short palindromic repeats)-Cas9 (CRISPR-associated protein 9)-guided chromatin immunoprecipitation. The dCas9 endonuclease utilized in this method is an inactive mutant of Cas9 that is still able "to bind to the gRNA-guided gene target" but is unable to cut into the DNA sequence (Zhang et al., 2016). When the target of interest is successfully bound to dCas9, the complex can then be immunoprecipitated (Zhang et al., 2016). We perform Cas9-guided chromatin immunoprecipitation to target elements in the mesoderm of *Drosophila melanogaster* and to pull down their chromatin.

Drosophila **Genome**

The model organism that we use to study silencers and their associated markers is *Drosophila melanogaster*, also known as the fruit fly, and specifically its embryonic mesoderm. There are several key reasons why scientists for over a century have chosen *Drosophila* as a model to better understand genetics. The main reason is that about "75% of disease-related genes in humans" have an ortholog in *Drosophila* with approximately 80% to 90% conservation of the DNA sequence (Pandey & Nichols, 2011). Conservation

means that if chromatin markers for silencers are discovered in *Drosophila,* the same markers and elements should be observed in humans.

Another reason why *Drosophila melanogaster* is an excellent model organism is that it replicates very rapidly. *Drosophila* can hastily produce "hundreds of identical offspring" in a matter of 10 to 12 days, which is more efficient than the period of reproduction seen in most mammals (Pandey & Nichols, 2011). In addition, the *Drosophila* genome has a "P element transposon" which has been altered. This transposon allows "overexpression and deletion" in various genes and is very beneficial in the UAS/Gal4 system, which we take advantage of in our project (Lloyd & Taylor, 2010). In this project, we primarily study embryonic mesoderm during development because it contains transcription regulatory elements that are evolutionarily conserved. The ability to cross a fly with an altered gene "downstream of a UAS site" with a fly expressing Gal4 results in the binding of Gal4 to UAS and "allows the gene of interest to be expressed selectively in desired tissue or tissues" (Lloyd & Taylor, 2010). This expression allows the testing of potential bifunctional characteristics of regulatory elements, primarily silencers, in different tissues such as the mesoderm versus the ectoderm.

Plasmid Cloning

The goal of this project is not only to identify the chromatin marker patterns present in silencers for creating a catalog but also to identify how these markers differ in mesoderm tissue versus non-mesoderm tissue. These markers could allow the prediction

of the presence of these silencers in the genome. The findings from the overall project should demonstrate that regulatory elements are not strictly enhancers or silencers but rather bifunctional regulatory elements that can act as either depending on the presence of co-factors or transcription factors in the specific cell types or tissues. The first goal of the project is to clone various vectors with either the dCas9 insert or sgRNAs to target elements that can be pulled down and studied, as illustrated in Figure 5. The inactive endonuclease, dCas9, targets elements through the direction of site-specific sgRNAs. In order to inject the dCas9 endonuclease and the sgRNAs into *Drosophila*, they must first be cloned into their respective plasmids.

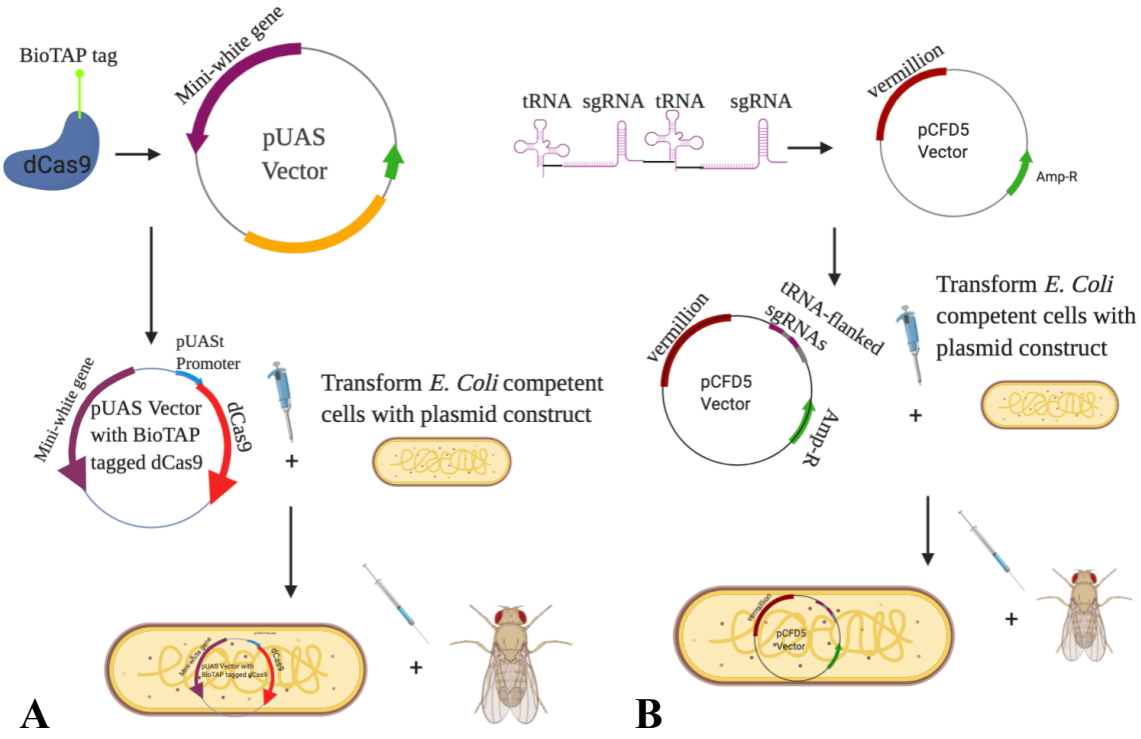

Figure 5. Plasmid cloning to target elements. **(A)** This diagram shows the workflow of inserting dCas9 into the pUAS vector before injecting it into *Drosophila*. **(B)** This diagram shows the workflow of injecting tRNA-flanked sgRNAs into the pCFD5 vector prior to injecting it into *Drosophila. Drosophila* individually containing the newly constructed plasmids will eventually be crossed so the sgRNAs can guide dCas9 to the target elements under Gal4 expression. dCas9 = deactivated CRISPR-associated protein 9; sgRNA = single-guide RNA; tRNA = transfer RNA.

There are various plasmid cloning methods such as Gibson Assembly, site-directed mutagenesis, and traditional cloning. Gibson Assembly is a more recently discovered cloning method that allows the ligation of a gene of interest into a vector. We initially chose this method because it "filters out any potential off-target DNA cleavage caused by CRISPR/Cas9," requires fewer steps, and is said to clone constructs seamlessly (Wei et al., 2020). A second cloning method is site-directed mutagenesis. This procedure works by using a dsDNA "vector with an insert of interest and two synthetic

oligonucleotide primers that both contain the desired mutation" (Agilent Technologies, 2015). The primers are "designed with mutually complementary 5' segments while the 3' segments are complementary . . . to the template" (Wei et al., 2020). This method was attempted because it has a mutagenesis efficiency "greater than 80%" and "relies on homologous recombination . . . to re-circulize plasmids containing homologous ends" (Wei et al., 2020). A third method of plasmid cloning, the most common method, is traditional cloning which "makes use of restriction enzymes and ligation of DNA *in vitro*" (Stevenson et al., 2013). Although this method is "often inefficient and laborious," traditional cloning proved to be successful when cloning the constructs in the silencers project (Iizasa & Nagano, 2006).

When cloning plasmids, most methods are followed with a DpnI restriction enzyme digest. DpnI works by targeting the methylated sites in the sequence GATC commonly found in the parent strands and *Escherichia coli* (*E. coli*) bacteria strains (New England Biolabs [NEB], n.d.). The enzyme digest allows the parent plasmid that does not contain the desired mutation to be destroyed, leaving the newly constructed plasmid.

The Silencers Project

Before joining Dr. Martha Bulyk's laboratory at Brigham and Women's Hospital (Boston, MA), research scientist Stephen Gisselbrecht, Julian Segert, and others completed much preliminary work with enhancer elements in *Drosophila*. The idea of studying chromatin markers on regulatory elements began in Dr. Bing Ren's laboratory at the University of California, San Diego (San Diego, CA). This group was able to identify

a “relationship between chromatin modifications at transcriptional enhancers” (Stevenson et al., 2013). This profound information has since been profiled in ENCODE and has led to the creation of computational models such as ChromHMM, which examines chromatin states and “observe(s) their biological functions” (Ernst & Kellis, 2012). This work led to our research regarding silencers and their possible associated chromatin markers, which have not been as heavily researched as enhancers.

The Bulyk laboratory began the silencer project when creating an assay that allows the screening of tissue-specific silencers, known as silencer FACS-Seq (sFS). This new screening technology confirmed the phenomenon of bifunctional regulatory elements. The silencing activity of these suspected elements was measured by observing green fluorescent protein (GFP) activity of the elements inserted into constructs and comparing them to the GFP activity of the ubiquitous enhancers. The silencers discovered in sFS will be used as a positive control in quantitative silencer FACS-Seq (qsFS) assays to study the activity of non-mesodermal enhancers in mesodermal tissue. The results of this experiment will be published by the Bulyk laboratory. On 10 of the validated bifunctional elements, the Bulyk laboratory will perform chromatin pulldown experiments and analyze the chromatin markers associated with these tissue-specific elements.

During preliminary experiments, the Bulyk laboratory performed BioTAP-XL-TALE approaches as a way to test chromatin pulldown on the confirmed bifunctional element brk_NEE. They were able to successfully transfect *Drosophila* mesodermal cells with a plasmid construct containing a TALE expressed under the Gal4 promotor to target

and pull down the chromatin of the element. This element can then act as a control when targeting potential silencers and pulling down the chromatin from the embryos. The goals of this project are to correctly insert dCas9 and sgRNAs into vectors so that elements acting as silencers can be targeted, pulled down, and analyzed for characteristic chromatin markers that could become prognostic markers.

SPECIFIC AIMS

Following the discovery of chromatin markers associated with specific enhancers, we have chosen regulatory elements to target and pull down so that we can examine chromatin markers associated with silencers. The main goals of this project are (1) to create plasmids for the dCas9 endonuclease and sgRNAs and (2) to target and examine regulatory elements that could potentially act in a bifunctional manner as both silencers and enhancers, focusing on the silencers and their chromatin markers. Before the injection of the dCas9 endonuclease and the sgRNAs into *Drosophila melanogaster*, we experimented with various molecular cloning techniques to efficiently and correctly clone the desired plasmid constructs.

By targeting the elements that act as silencers and comparing them with their orthologous enhancers in *Drosophila ananassae*, we can compare the chromatin modifications between the two. By identifying modifications that are specific to the silencers and to a particular cell or tissue type, we can characterize histone modification patterns to the chromatin states of a region in the genome. Because there is high conservation of the chromatin markers associated with the genome of *Drosophila melanogaster* to humans, we could potentially use these chromatin markers to screen for silencers in the human genome (Ho et al., 2014).

METHODS

Getting dCas9 Into *Drosophila* DNA

To target silencers in the *Drosophila melanogaster* genome, we first ligated the dCas9 fragment into the pUAS vector. The dCas9 insert fragment was 4,107 bases, and the pUAS vector was 7,890 bases. The initial plan was to use Gibson Assembly (NEB, Ipswich, MA) for inserting dCas9, but after no success, we attempted other plasmid cloning methods. In all of these cloning techniques, the first step was to amplify the dCas9 fragment and the pUAS vector.

PCR Amplifying DNA Fragments

The first step was to isolate the plasmid DNA, pUAS, from the bacteria by performing a miniprep using a Qiagen kit and protocol (Qiagen Inc., Germantown, MD). After isolating the plasmid and measuring the DNA concentration, we then performed a polymerase chain reaction (PCR). J. Segert, a current Ph.D. student who previously worked on the project, designed the primers for the amplification portion of Gibson Assembly. Developed using the program SNAPgene (GSL Biotech, San Diego, CA), the primers were 30-40 nucleotides long with complementary overhangs that created a 16 bp overlap to allow annealing of the pUAS vector to the dCas9 insert. The designed primers consisted of forward and reverse primers for the dCas9 insert and the pUAS vector, as well as the N-terminus and C-terminus tagged versions of each.

We followed the 50 µL reaction protocol for the Q5 polymerase PCR reaction provided by New England Biolabs (NEB, Ipswich, MA). The protocol was 25 µL of Q5 2x master mix, 2.5 µL of the forward primer, 2.5 µL of the reverse primer, 1 µL of the template DNA, and 19 µL of deionized water. We repeated the protocol for the N-terminus tagged pUAS vector, C-terminus tagged pUAS vector, N-terminus tagged dCas9 insert, and C-terminus tagged dCas9 insert. Each reaction cycle in the thermocycler consisted of the following: initial denaturation at 98 °C for 30 seconds; 30 cycles of 98 °C for 7 seconds, 70 °C for 20 seconds, and 72 °C for 25 seconds per kilobase (kb); final extension at 72 °C for 2 minutes; and finally holding at 4 °C. After the PCR was complete, we ran each PCR reaction product on a 0.7% agarose gel. When we observed the correct band sizes, the DNA from the gel bands was extracted and purified. After a few trials, we added a step of performing a restriction enzyme digest using DpnI in an attempt to increase the DNA concentrations.

Gibson Assembly

For the Gibson Assembly reaction, the amplified, extracted, and purified DNA fragments for the pUAS vector and the dCAs9 insert were calculated into picomoles and converted into microliters (µL) using their measured DNA concentrations. The reaction required the two DNA fragments, the Gibson Assembly Master Mix (which includes the 5’ exonuclease, Taq polymerase, DNA ligase, and deoxynucleoside triphosphates [dNTPs]), and deionized water. After mixing the DNA fragments, Gibson Assembly Master Mix, and deionized water together in a 50 µL microcentrifuge tube, we placed the

tube in the thermocycler at 50 °C for 1 hour. After the reaction was complete, we transformed competent *E. coli* cells with the reaction product, plated on agar plates of lysogeny broth (LB) and ampicillin (AMP) (LB+AMP plates), and incubated for 12-18 hours. We screened the colonies on these plates with various primers, including dCas9 primers and primers explicitly designed for screening which we labeled "N-screen" and "C-screen." These primers screened for the region where the pUAS vector and the dCas9 insert annealed.

Site-Directed Mutagenesis

Another method that was performed in an attempt to anneal the dCas9 insert to the pUAS vector was QuikChange site-directed mutagenesis (Agilent Technologies, Santa Clara, CA). The mutation we chose was the insertion of restriction enzyme sites, AgeI and SpeI, into the pUAS vector by first creating primers that bound to the denatured plasmid and were complementary to the opposing strands. The protocol in the thermocycler was as follows: 30 seconds at 95 °C, cycles 2-18 at 95 °C for 30 seconds, 55 °C for 1 minute, and 68 °C for 1 min/kb. After the primers annealed, a DNA polymerase extended the strands and allowed the restriction enzyme site to be inserted into the plasmid.

After the mutation was thought to have been inserted into the plasmid, we then PCR amplified the new plasmid construct using a Phusion polymerase (NEB, Ipswich, MA). This step was followed by a DpnI restriction enzyme digest. We then transformed the top 10 competent *E. coli* cells and screened them to confirm that the restriction

enzyme site AgeI or SpeI was inserted into the pUAS vector. This screening was completed by performing an AgeI restriction enzyme digest and a SpeI restriction enzyme digest, respectively, and running the product on an agarose gel. We also ran a PCR product of the uncut vector on the gel as a negative control.

Restriction Enzyme Cloning

The restriction enzyme site chosen to insert into the plasmid was also engineered to flank the dCas9 insert in the phage donor. After both the phage donor with the dCas9 insert and the backbone with the inserted restriction enzyme site were restriction enzyme-digested with AgeI (NEB), they were treated with a shrimp alkaline phosphatase (SAP) (NEB) to remove the 5' end phosphate to prevent self-ligation of DNA fragments. After the ligation reaction was complete, we then transformed *E. coli* cells with the newly ligated fragments and screened them with various primers and restriction enzyme digests to confirm the dCas9 insert was present and had the correct orientation in the pUAS vector.

Cloning sgRNAs

We followed a Cas9-guided chromatin immunoprecipitation protocol to target elements using sgRNAs. To do so, S. Gisselbrecht (Bulyk laboratory) selected 10 bifunctional elements. We then designed 3-4 sgRNAs to target each specific element (Figure 6). We composed the sgRNAs using various online programs such as CHOPCHOP (http://chopchop.cbu.uib.no/), CRISPRscan

(https://www.crisprscan.org/?page=gene), and UCSC genome (https://genome.ucsc.edu/cgi-bin/hgGateway). These tools helped discover target sequences. We focused on the conserved elements between the *Drosophila melanogaster* and *Drosophila ananassae*. These regions were avoided so that dCas9 could bind to the target without causing off-target effects. Based on these sequences, we were able to design sgRNA primers that allowed the cloning of the sgRNAs into the vector. To design and clone the primers, we followed the protocol described in the methods of Port and Bullock (2016).

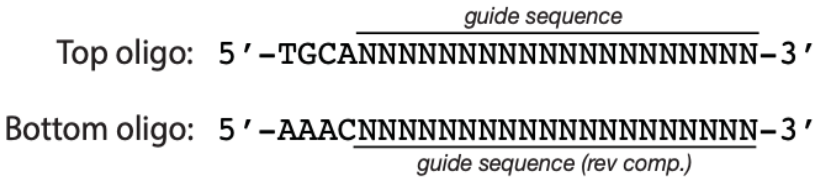

Figure 6. Oligonucleotide design for each sgRNA. Each designed guide sequence has a top oligonucleotide (oligo) and a bottom oligo that are reverse complimentary to each other. The oligos also have four additional bases added to the 5' end and are desalted for purification. sgRNA = single-guide RNA. Figure taken from (Port & Bullock, 2016).

The first step was to digest the pCFD5 vector with the restriction enzyme, BbsI. The protocol recommended the digestion of 8 µg of the pCFD5 DNA with 1 µL of BbsI, 3 µL of a 10x buffer, and the remaining volume consisting of deionized water. The total reaction volume was 30 µL, which was incubated at 37 °C for 3 hours. The BbsI-digested pCFD5 was run on a 1% agarose gel, band size verified, and then extracted and purified. We eluted the product in 25 µL of deionized water and measured the DNA concentration with a NanoDrop spectrophotometer (Thermo Fisher Scientific, Waltham, MA).

The next step after ordering the designed oligonucleotides (oligos) was to set up the phosphorylation and annealing reaction. This step required 1 µL of 100 µM top oligonucleotide, 1 µL of 100 µM bottom oligonucleotide, 1 µL of T4 ligation buffer, 6.5 µL of deionized water, and 0.5 µL of T4 polynucleotide kinase. The reaction mixture was incubated at 37 °C for 30 minutes, followed by 95 °C for 5 minutes, and then ramped down to 20 °C at a rate of 5 °C per minute.

After the annealing step was the ligation step. This reaction required 60 ng of BbsI-digested pCFD5 (calculated depending on the measured DNA concentration), 1 µL of the annealed oligonucleotides diluted 1:20 in distilled and deionized water (ddH2O), 1.5 µL of 10x T4 ligation buffer, 1 µL of T4 DNA ligase, and remaining X µL of ddH2O (up to 15 µL). It was then incubated at room temperature for 30 minutes. Following ligation, we then transformed *E. coli* competent cells by adding 2 µL of the reaction into 50 µL of the cells. The competent cells were recovered in 700 µL of super optimal broth (SOC) media, plated 100 µL on LB+AMP plates, and incubated overnight in a 37 °C incubator. The colonies were then Sanger sequenced with a specific primer to confirm the insertion and correct orientation of sgRNAs into the pCFD5 vector.

Sanger Sequencing

For Sanger sequencing of the colonies, we sent our samples to Dana-Farber/Harvard Cancer Center (Boston, MA) after following the sample preparation protocol posted on their website (https://dnaseq.med.harvard.edu/). According to the protocol, we measured the DNA concentrations of the ligation products using a

NanoDrop spectrophotometer and placed 10 μL of each diluted ligation reaction product on a template plate and 12 μL of the primer on a separate primer plate. The primer used to sequence the pCFD5 plasmid for the sgRNA was the U6:3 forward primer sequence 5'-ACGTTTTATAACTTATGCCCCTAAG-3' from the methods of Port and Bullock (2016).

Chromatin Pulldown

After the sgRNAs were sequence-verified, we injected the plasmids into the flies. To test the dCas9 pulldown efficiency, we crossed the flies with the pCFD5 plasmid containing *white* gene sgRNA and Gal4 with the flies that contained the pUAS plasmid with BioTAP-tagged dCas9; flies crossed with Gal4 alone acted as a negative control. The resulting progeny then underwent a chromatin pulldown to examine their chromatin modifications. We were able to detect if the sgRNA had successfully guided dCas9 to the target elements after performing the chromatin pulldown and extraction. Meanwhile, we used a validated sgRNA to target the *white* gene to test the efficiency of the chromatin pulldown; the sgRNA was provided by the Ueda Laboratory from the National Institute of Genetics (Mishima, Japan) (Shu & Ueda, 2013).

To extract the chromatin, we added 5 mL of buffer A1 with 1.6% formaldehyde in up to 200 mg of dechorionated embryos. Using a loose pestle, we applied five strokes before switching to a tight pestle and applying another five strokes. We then exposed the embryos to formaldehyde at room temperature for up to 15 minutes, added 0.55 mL of

2.5 M glycine, and incubated them at room temperature for 5 minutes before placing on ice.

In the next step, the nuclei were spun down for 5 minutes at 4,000 g, washed three times with 3 mL of buffer A1, one time with 3 mL of lysis buffer without sodium dodecyl sulfate (SDS), and then resuspended as much as possible in 0.5 mL of lysis buffer. After the addition of 0.1% SDS and 0.5% N-Lauroylsarcosine, the embryos were incubated at 4 °C for 10 minutes while rotating. The remaining nuclei were then sonicated for two rounds in a Diagenode Bioruptor (Diagenode Inc., Denville, NJ) to shear the chromatin. Each round consisted of 15 cycles alternating the power on and off for 30-second intervals, and at the end, the samples were spun at maximum speed for 5 minutes before recovering and reserving the supernatant. The pellet was resuspended in 0.5 mL of lysis buffer and 0.05% SDS and incubated at 4 °C for 10 minutes while rotating and spinning for 5 minutes at maximum speed. We then recovered and pooled the supernatants before spinning them at maximum speed for 10 minutes and recovering again.

After we prepared the chromatin, J. Segert (Bulyk laboratory) performed the pulldown by means of immunoprecipitation with immunoglobulin G (IgG) agarose beads. These beads bind to the protein A moiety in the BioTAP tag bound to dCas9. This binding allowed the pulldown of dCas9 protein and the now bound target elements.

RESULTS

The main goal for this project was to insert the dCas9 and sgRNAs into their respective vectors to target regulatory elements. As we worked toward this overarching goal, we ran into some issues with Gibson Assembly and the crucial process of inserting the dCas9 endonuclease into the pUAS vector.

We followed the New England Biolabs (NEB) Gibson Assembly protocol when performing the reactions, as shown in Figure 7. The first few times that we attempted the Gibson Assembly method, it did not work, so we decided to vary different factors of the reaction in an attempt to make it more efficient. In the beginning, we were adding 5 µL of the reaction product to the competent cells for cell transformations; however, other lab members suggested that we dilute the reaction product before mixing it with the competent cells. We then performed a 1:4 dilution on the reaction product and added 2 µL of that dilution to the competent cells. As an experiment, we tested 5 µL of the undiluted reaction in the competent cells, as we had previously been doing, compared with 2 µL of the diluted product. The experiment showed that there was an increase in the number of colonies on the NEB-provided positive control plate, proving that it increased the transformation efficiency. As a result, we continued to use a dilution when performing cell transformations.

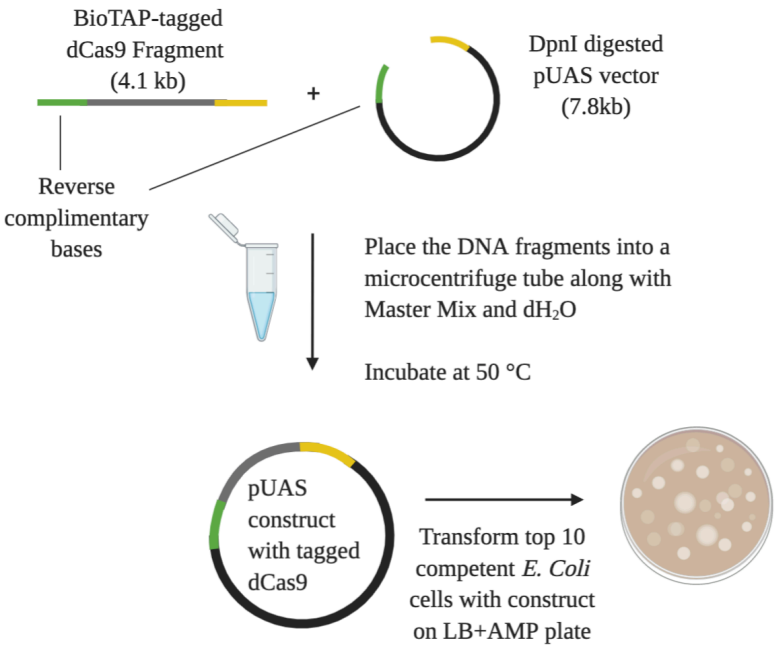

Figure 7. **Schematic of the NEB standard Gibson Assembly workflow.** This schematic demonstrates the Gibson Assembly protocol. Following the cell transformations, the colonies were screened with various primers. NEB= New England Biolabs; dH_2O = deionized water; LB+AMP = lysogeny broth + ampicillin.

Troubleshooting PCR Amplification

We decided to troubleshoot the PCR protocol before the Gibson Assembly reaction in an attempt to increase the DNA concentrations of the dCas9 insert and the pUAS vector for optimization. We first began to experiment with the annealing temperatures. The annealing temperatures were chosen by using the NEB primer T_m (melting temperature) calculator and ranged from 64 °C to 70 °C. Each designated temperature was set for each fragment when running the PCR with Q5 polymerase. When issues began to rise, we decided to attempt a gradient PCR. The annealing temperature set for the gradient PCR ranged from 55 °C to 70 °C. Using this PCR gradient still did not

produce the correct band size for the desired DNA fragments, so we examined other factors.

We also explored other polymerases, such as KAPA (Roche Diagnostics, Indianapolis, IN) and Phusion instead of the original Q5 polymerase, to get sharper bands on the gel and higher DNA concentrations. For each polymerase, we followed the respective protocol for the N-tagged and C-tagged dCas9 inserts and ran them on a gel. We were able to obtain more pronounced bands on the gel, as well as better DNA concentrations when amplifying with the KAPA polymerase, as shown in Figure 8. It is important to note that the Phusion polymerase did not give the correct band size of 4.1 kb for any of the inserts, and the Q5 polymerase did not give the correct band size for the C-tagged version of the dCas9 insert.

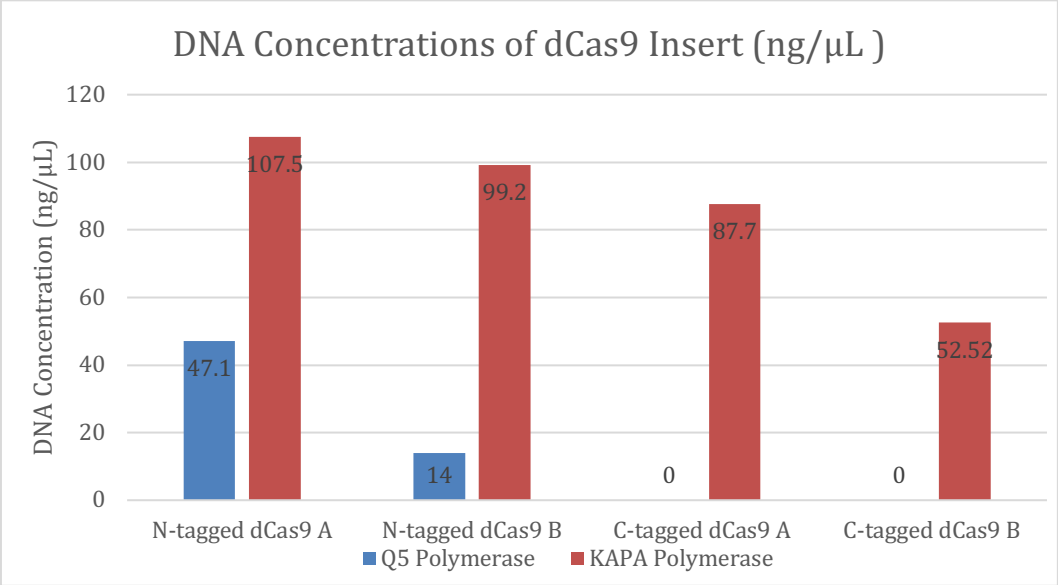

Figure 8. DNA concentrations of the dCas9 insert using different PCR polymerases. This figure shows the N-tagged version and the C-tagged version of dCas9 A and dCas9 B. The two templates of dCas9 were used in an attempt to find the optimal concentrations. The Q5 polymerase did not show the correct band size for the C-tagged

version of the dCas9 insert. KAPA polymerase showed the highest DNA concentrations. dCas9 = deactivated CRISPR-associated protein; PCR = polymerase chain reaction.

Increasing DNA Concentrations

In addition to altering the PCR amplification, we evaluated some alternative purification methods. The initial gel purification of the DNA began with performing gel band extractions of the correct band sizes for the insert and vector. We quickly realized that running a gel and purifying the bands were compromising the DNA concentrations, which were relatively lower than the recommended amount needed for the Gibson Assembly reaction. In an attempt to increase the DNA concentrations, we followed a suggestion to purify the PCR product with solid-phase reversible immobilization (SPRI) beads instead of the gel. Considering that we had been getting the correct band size for the individual fragments after performing the DpnI enzyme digest, the chance of the vector self-ligating with itself was unlikely. The SPRI beads function by cleaning up the DNA when the beads combine with "polyethylene glycol (PEG) and salt," which allows the carboxyl groups on the surface of the beads to attract and bind to the "negatively charged DNA" (Hadfield, 2012). With DNA binding to the beads, excess compounds are washed away, leaving a higher DNA concentration of the PCR product.

After achieving higher DNA concentrations for the dCas9 insert and the pUAS vector with the SPRI beads, we attempted Gibson Assembly several more times. We had already increased the incubation time to the optimal range of an hour; however, we had not attempted different incubation temperatures. While researching alternative incubation temperatures, most protocols and scientists mentioned that the recommended temperature

of 50 °C was the optimal temperature for Gibson Assembly and that there was not much success with other temperatures. Nevertheless, we discovered one paper with a protocol called the "Modified Gibson Assembly Protocol" that suggested incubating the reaction at "60 °C for 4-8 minutes" before the usual one-hour incubation at 50 °C (Birla & Chou, 2015). We attempted this "Modified Gibson Assembly Protocol" by altering the incubation step and retaining the other steps of the previous protocol. However, we did not have success with this modified protocol, as evidenced by the lack of colonies on any of the plates.

Trials with Site-Directed Mutagenesis

Following multiple unsuccessful Gibson Assembly trials, we briefly attempted site-directed mutagenesis. S. Gisselbrecht (Bulyk laboratory) recommended this method because he was able to quickly design the required oligonucleotide primers and because he had success with the method from previous projects. We followed the QuikChange site-directed mutagenesis protocol (Agilent, 2015) illustrated in Figure 9. After we believed that the restriction enzyme site had been mutated into the pUAS vector, we performed a DpnI digest, transformed *E. coli* cells, performed enzyme digests to confirm the insertion, and then ran them on an agarose gel. We observed that the band sizes on the gel were not correct, indicating that the site-directed mutagenesis method was unsuccessful.

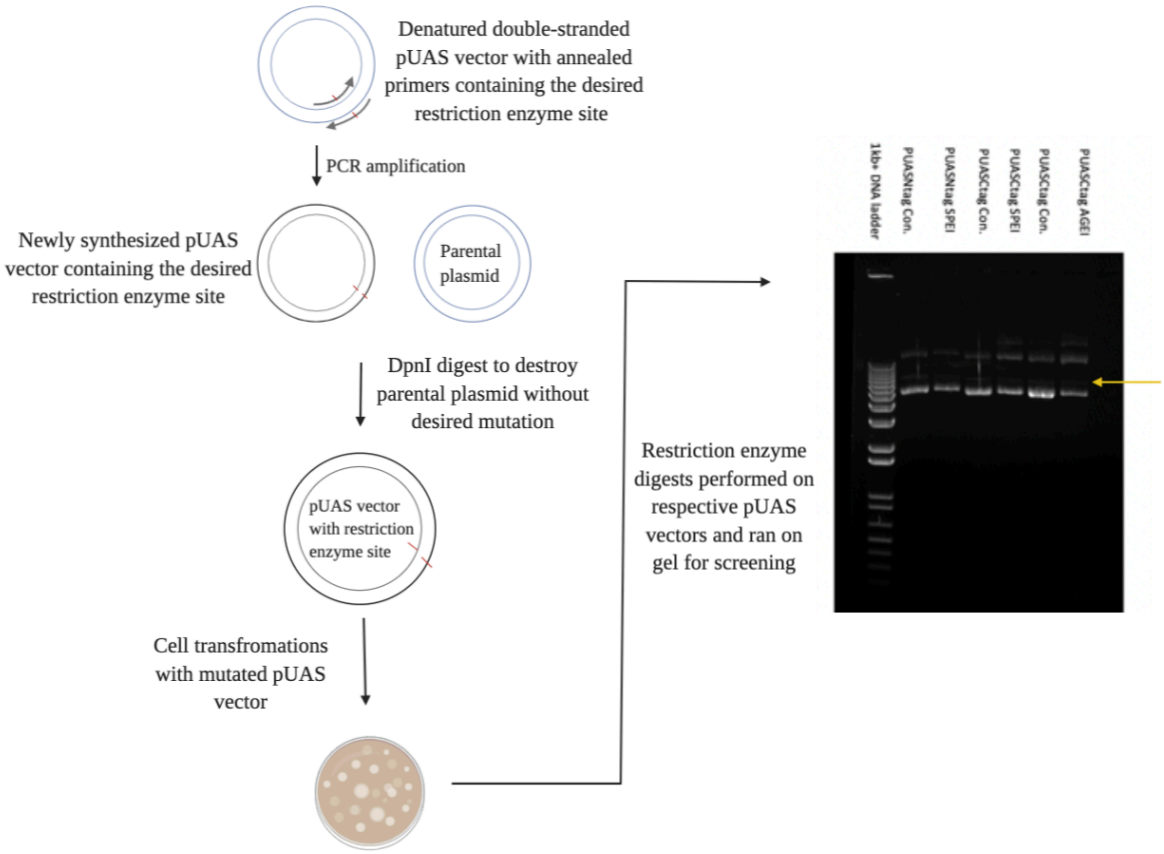

Figure 9. Site-directed mutagenesis workflow. This schematic shows the method of site-directed mutagenesis that was performed on both the N-tagged and C-tagged pUAS vectors. The restriction enzyme sites that were attempted to be inserted in the vectors were the AgeI and SpeI sites. The gel with the results of the enzyme digests showed incorrect band sizes for both vectors. The yellow arrow indicates where the bands should be, around 7.9 kb. PCR = polymerase chain reaction.

More Gibson Assembly Attempts

After the failed attempt with site-directed mutagenesis, Brian Rabe from Dr. Constance Cepko's laboratory (Harvard Medical School, Boston, MA) gave us some of his Gibson Assembly Master Mix, which he termed "Enhanced Gibson Assembly." B. Rabe did not disclose the contents of his master mix, except that it contained the necessary components of an endonuclease, a ligase, a polymerase, and nucleotides. When

using his master mix with his recommended protocol, we noticed that there was an increase in the number of colonies on the plates of the competent cells transformed with the reaction product. Because we had a large number of colonies, we had to screen them to confirm that the cloning was successful. To do so, we PCR amplified the Gibson Assembly products with various primers, including the dCas9 primers and other primers designed explicitly for screening (N-screen and C-screen primers). S. Gisselbrecht (Bulyk laboratory) designed the N-screen and C-screen primers to screen the plasmid for the expected ligation site of both fragments instead of just the dCas9 insert. This screening primer allowed us to not only detect the dCas9 insertion but also confirm its orientation as well. Another screening method that we performed involved restriction enzyme digests of the Gibson Assembly product with EcoRI and PvuII; these screenings included an uncut version of the vector as a negative control. Some of the screened colonies on the agarose gel showed nearly correct band sizes, so we additionally performed enzyme digests on these colonies as a second screening and ultimately discovered that none of them were correct. Although the "Enhanced Gibson Assembly" method did not work in this project, it increased the success of the cell transformations and proved to be more efficient than the standard Gibson Assembly protocol.

Success with Traditional Cloning

Finally, S. Gisselbrecht (Bulyk laboratory) suggested that we try traditional cloning. To ensure that neither the insert nor the vector is cut at incorrect sites, the choice of the restriction enzyme site to insert is crucial. We chose the AgeI restriction enzyme

site to insert into both the N-tagged and C-tagged versions of the pUAS vector (Figure 10). This restriction enzyme site was selected because it can flank the dCas9 insert without cutting it. We then engineered the AgeI restriction enzyme site into the multiple cloning site (MCS) of the pUAS vector (Addgene, n.d.). The majority of this procedure was performed by S. Gisselbrecht because he had successfully cloned many fragments with traditional cloning multiple times during his research career.

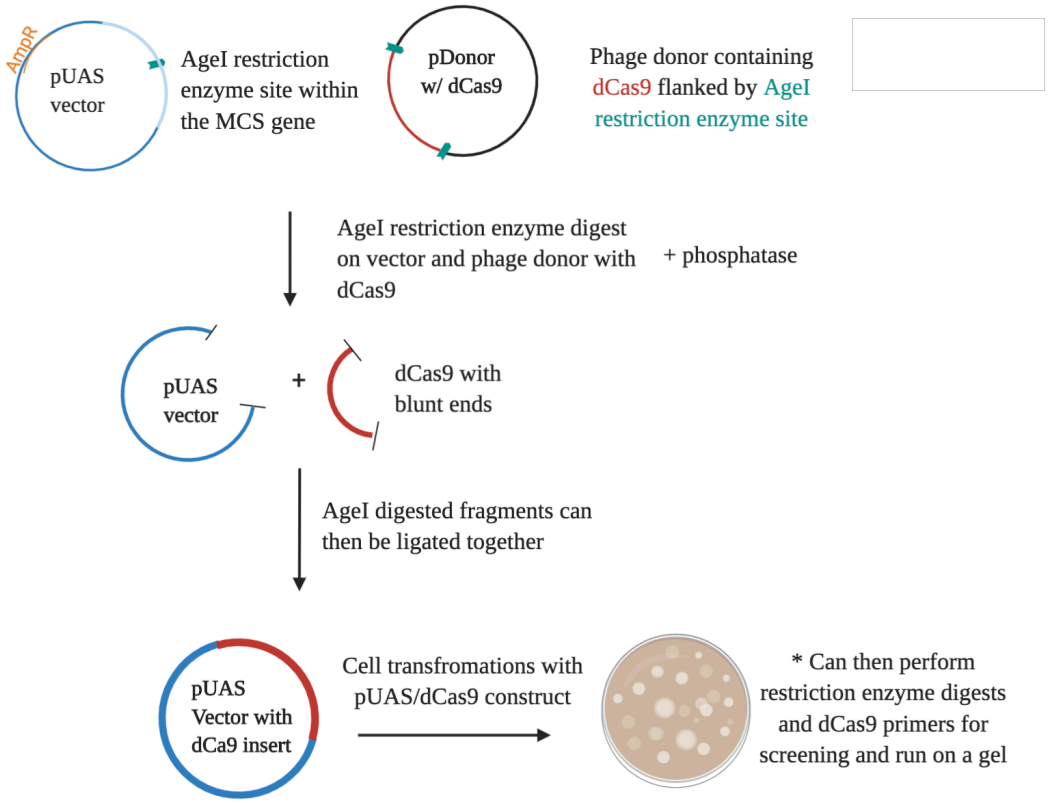

Figure 10. Traditional plasmid cloning workflow. This schematic depicts the method of traditional cloning with restriction enzymes. This method includes a restriction enzyme digest and ligation of the two fragments. dCas9 = deactivated CRISPR-associated protein; MCS = multiple cloning site.

The pUAS vector and the dCas9 fragment in the phage donor were PCR amplified and restriction enzyme-digested with AgeI to cut the fragments at the desired site and with EcoRI and PvuII to screen the fragments. After confirmation that the pUAS vector and the phage donor with dCas9 were correctly digested, the two fragments were ready to undergo ligation. We were finally able to construct the N-tagged and C-tagged versions of the pUAS/dCas9 vector. Figure 11 shows the results of the enzyme digests performed on the successfully ligated N-tagged construct. After performing the enzyme digests, both whole plasmids were Sanger-sequenced and confirmed to have dCas9 successfully inserted into the vector in the right position and the correct orientation (Figure 12).

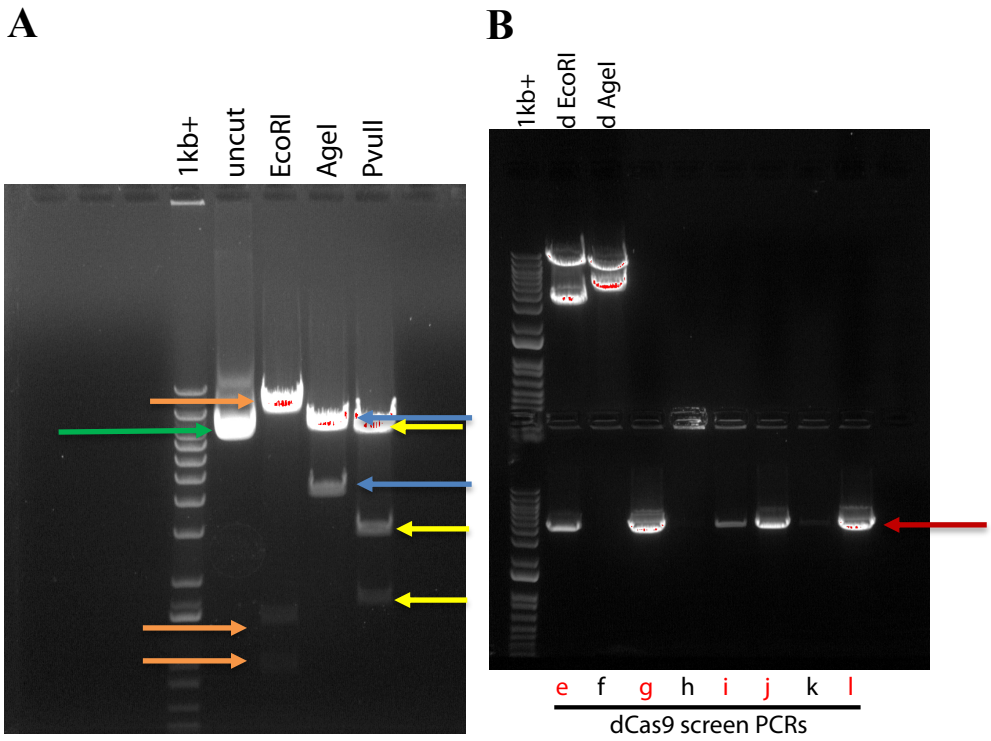

Figure 11. Restriction digest gel results. (A) This gel shows restriction enzyme digests of the cloned pUAS N-tagged dCas9 vector. Lane 2 is the uncut, empty vector which would be around the size of 8 kb, as seen with the green arrow. Lane 3 is the EcoRI-digested vector which should have band sizes of 940, 1300, and 9600 bases, as seen with the orange arrows. Lane 4 is the AgeI-digested vector which shows the individual fragment band sizes of 7.9 kb and 4.1 kb, as seen with the blue arrows. Lane 5 is the vector digested with PvuII to show the fragment sizes of 1600, 2800, and 7600, as seen with the yellow arrows. **(B)** This gel shows restriction enzyme digests of the cloned pUAS C-tagged dCas9 vector. Lanes 2 and 3 on the top portion of the gel show the incorrect band sizes of the enzyme-digested vector of one of the clone copies. The bottom portion of the gel shows the screening of 8 colonies with dCas9 primers. Colonies E, G, I, J, and L all give the correct dCas9 insert amplification at the band size of 4.1 kb, as seen with the red arrow. The colonies that showed the correct band sizes for the enzyme digests or the dCas9 PCR were sent for whole plasmid sequencing. dCas9 = deactivated CRISPR-associated protein; kb = kilobase; PCR = polymerase chain reaction.

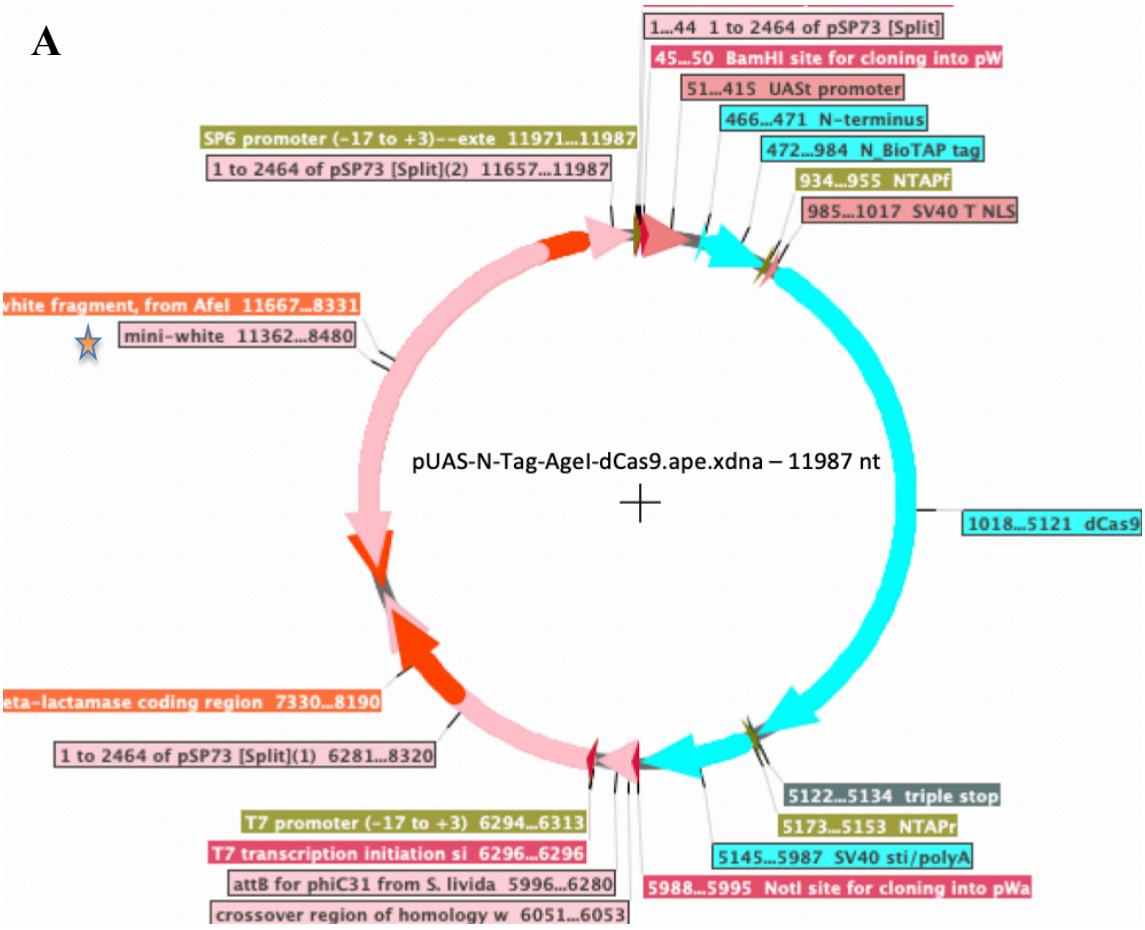

Figure 12. Whole plasmid sequenced pUAS vectors with dCas9 insert. This figure shows the **(A)** pUAS vector with N-terminus tagged version of dCas9 and **(B)** the pUAS vector with C-terminus tagged dCas9. These plasmids were created using the DNA sequences returned from the whole plasmid sequencing in the software Serial Cloner (Version 2.6.1). Both vectors show important characteristics such as the *white* gene, the UAS promoter, the bioTAP tags, and sites where the screening primers anneal. dCas9 = deactivated CRISPR-associated protein.

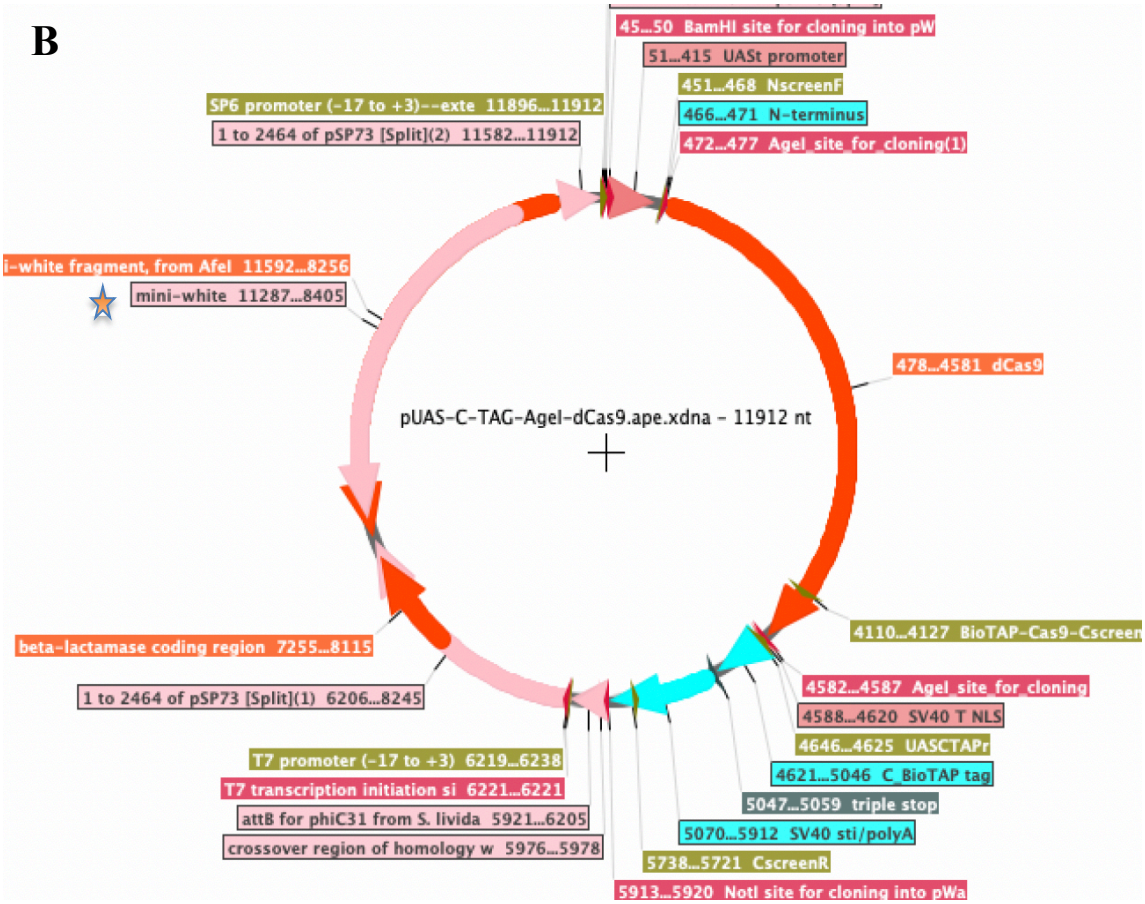

Figure 12 (continued).

After the plasmids were Sanger sequenced, they were injected into *yellow*, *white* flies with PhiC31 integrase at the attP40 site on chromosome II. The integrase is under the control of the *nanos* promoter which allows deposition of the mRNA in eggs before fertilization. The plasmid was injected into a fly stock containing PhiC31 integrase and the wild-type eye gene at the attB site and then crossed with virgin females. We injected the flies with the sequence-verified plasmid constructs and determined transformants by

observing the color of the eyes. The insertion of the *white* gene into the attP40 site is confirmed in transformant flies that have pale-yellow eyes rather than the mutant white color. The pale-yellow eye color was observed in the flies that we injected with the pUAS/dCas9 vector.

The wild-type males with the pUAS plasmid were pulled down and crossed with virgin females that contained a balancing chromosome. The balancing chromosome allows us to maintain a lethal recessive mutation, as well as a pure stock of flies with suppressed wild-type genes. We removed the female virgins and males that expressed the plasmid, along with the wild-type red eyes and curly wings. Curly wings are a dominant mutation on chromosome II that, when homozygous, is lethal. We then isolated females with the plasmid and non-curly wings (wild-type gene), which were then self-crossed.

Flies with the pUAS vector with the dCas9 insert have recently been collected. We will cross them with females containing the Gal4 promoter. The Gal4 promoter is present in *Drosophila* and is used to drive the expression of the pUAS vector in the mesoderm. The female virgins with the gal4-expressed plasmids will be crossed with a male containing another balancing chromosome. By pulling down males with the gal4-expressed plasmid and crossing it with a female with a balancing chromosome, we will be able to create a pure line of virgin females that have the gal4-expressed pUAS vector with the tagged dCas9.

Designing sgRNAs to Guide dCas9 to Targets

The crossbreeding of the pUAS construct was performed by S. Gisselbrecht (Bulyk laboratory). During this period, we continued to design sgRNAs and insert them into a pCFD5 vector. As mentioned earlier, we attempted to design at least 3 sgRNAs for the 10 chosen target elements and composed oligo primers for each of these sequences. We chose these elements to study because they are enhancers that can also act as silencers, proving to be established bifunctional elements. This information was found using the previously mentioned sFS assay and is in the process of being published by the Bulyk laboratory. It was necessary to avoid areas of conservation in the genome because these regions had been conserved throughout evolution, implying that they are important and most likely involved in mechanisms such as transcription factor binding or protein interactions (Figure 13). These types of interactions in the genome were avoided to minimize off-target effects. After finding potential target sequences that did not have dense areas of conservation between *Drosophila melanogaster and Drosophila Ananassae*, we were able to successfully design sgRNAs for 8 of the 10 potential elements.

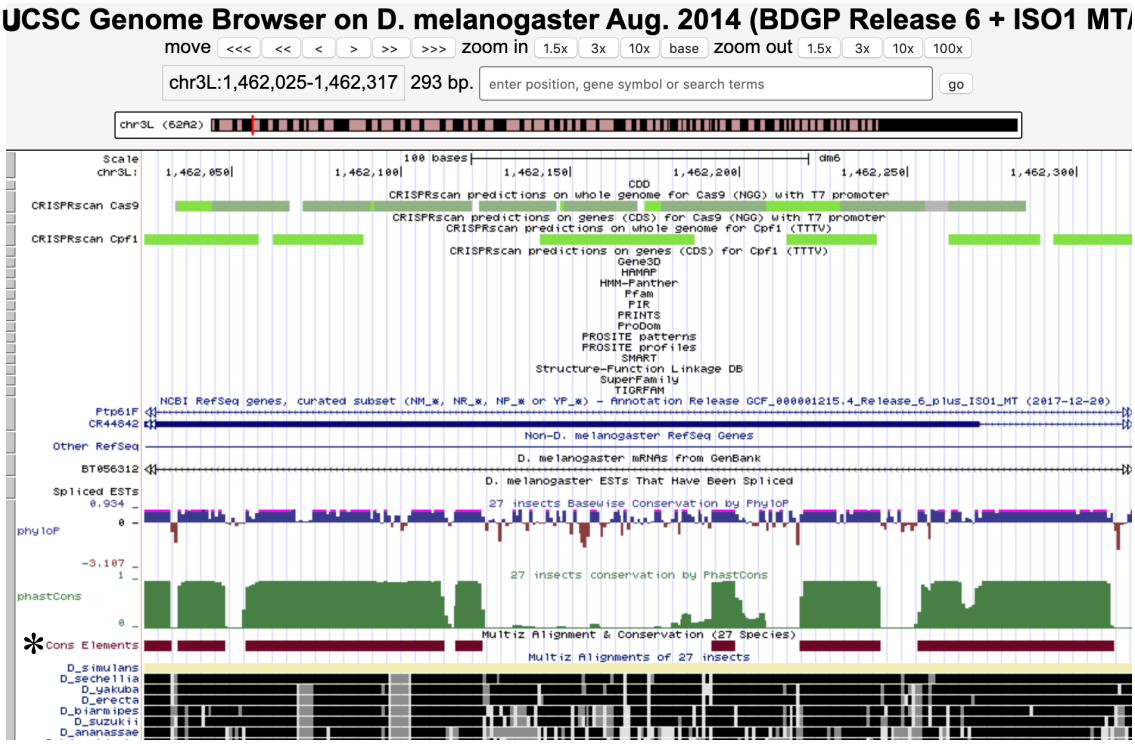

Figure 13. Element example from UCSC Genome Browser. This is an example of the image for the target element "Rho_NEE_long" from the UCSC (University of California, Santa Cruz) Genome Browser that was used to design target sequences. The segment named "Cons Elements" (marked by *) helped to select DNA segments that lack magenta bars, which represents regions of conservation. This selection method aimed at lowering the amount of off-target effects. Image taken from (UCSC Genome Browser, https://genome.ucsc.edu/cgi-bin/hgGateway).

Ligating sgRNA Sequences Into pCFD5

After we designed and ordered the primers for the top and bottom oligonucleotides, we performed the phosphorylation and annealing reaction. The designed primers gave the sgRNA sequence necessary to ligate into the pCFD5 vector. Following the miniprep of the pCFD5 vector, we digested it with the BbsI restriction enzyme, ran the digested vector on an agarose gel, gel-purified the product, and measured the DNA concentrations with NanoDrop. The initial DNA concentrations for the vector

were relatively low at about 80 ng/μL. Because we had enough product for the ligation reactions and the cell transformations, we continued with the process; however, we did not get colonies on any of the plates (which included ligation reactions with almost 30 different sgRNA sequences).

We began to troubleshoot by trying to increase the DNA concentration of the pCFD5 vector. We attempted to midiprep the vector to increase the volume yield available for the many ligations we had to perform; however, the DNA concentrations were deficient at around 40 ng/μL. We returned to the Qiagen miniprep kit but used a newer kit because the previous one might have been outdated. We also let the BbsI enzyme digest run longer in the 37 °C incubator. Initially, the enzyme digest ran for 2 hours, but a lab member suggested to let it run overnight for 16 hours. It was important not to go over 16 hours because it would increase the chances of the enzymes possibly cleaving at sites other than the restriction enzyme sites and ultimately destroying the plasmid. In addition, we decreased the agar percentage of the gel from 1% to 0.7% to make the bands sharper so we could be more precise in the gel band DNA extraction and purification. As a result of these changes, we were able to get adequate DNA concentrations over 100 ng/μL and felt confident about moving forward to perform the ligation reactions.

Ultimately, we carried out the ligation reaction and cell transformations two times because the first attempt was unsuccessful. Previous cell transformations had been performed with NEB top 10 competent cells with the strain 10-beta competent *E. coli*. However, the recommended cells to be transformed in this protocol were the 5-alpha

competent *E. coli* cells. Both strains of the competent cells contain the recA- and endA- genes, which are essential for preventing recombination and endonuclease degradation, respectively. Following the protocol recommendation, we used the 5-alpha competent cells. During the first trial of the cell transformation, we diluted the transformed cells by 10-fold before spreading them onto the LB+AMP plates. This first cell transformation was unsuccessful in that no colonies appeared on the plates. One notable factor, however, was that on removing the plates from what was supposed to be a 37 °C incubator, the temperature was 28 °C. Fellow lab members realized that the temperature was changed overnight and might have been a factor in the lack of colonies.

In the second trial, we used undiluted ligation product to transform the cells, spread about 200 µL of the transformed cells onto its labeled LB+AMP plate, and incubated the plates at 37 °C for 16 hours. We were able to successfully get colonies on all the plates. To confirm the correct insertion of each sgRNA in the pCFD5 vector, whole plasmid sequencing was applied to the cells transformed by the plasmid constructs. We chose two colonies for each sgRNA sequence and miniprepped them before sending them for Sanger sequencing. Because of the possibility that we would inject flies with the sgRNA vector constructs, we had to miniprep the ligation reaction products with the ZymoPURE miniprep kit (Zymo Research, Irvine, CA) rather than the usual Qiagen kit. In previous experiments we had observed that the vector constructs miniprepped with the Qiagen kit killed the embryos, whereas those miniprepped with the ZymoPURE kit did not. As we were unaware of this issue when we began miniprepping the samples, about half the correct plasmid sequences had to be processed again using the ZymoPURE kit

before injecting the embryos. The results of the Sanger sequencing should follow the example in Figure 14.

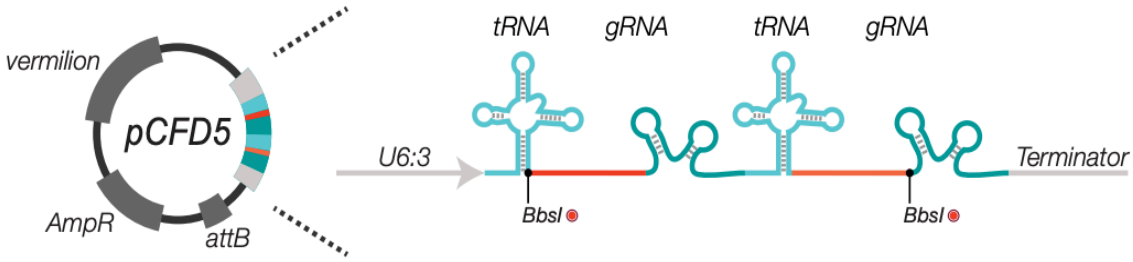

Figure 14. pCFD5 with tRNA-flanked guide RNAs. This schematic shows the inserted sgRNAs (also referred to as gRNAs) flanked by tRNAs and under the expression of the U6:3 promoter in the pCFD5 vector. It also shows the gRNA inserted into the pCFD5 at the BbsI-digested sites. sgRNA= single-guide RNA; gRNA = guide RNA; tRNA = transfer RNA. Figure taken from (Port & Bullock, 2016).

After getting the sequence results for the whole plasmids, we were able to determine that about half the pCFD5 plasmids had the sgRNAs inserted correctly into the BbsI-digested sites. The sequencing results showed that the sgRNA sequence for the elements Rho_NEE_long, hkb_0.6kbRIRv, PNR_p4, ths_Neu4_early_embryonic_enhancer, ind_moduleA, ind_moduleBC, and lz_CrystalCellEnhancer1236-737 was successfully inserted into the pCFD5 (Figure 15).

Element Name	Template A	Template B
RNL #1	Correct	Incorrect
RNL #2	Inccorrect	Incorrect
RNL #3	Incorrect	Incorrect
Hkb #1	Correct	Correct
Hkb #2	Incorrect	Correct
Hkb #3	Correct	Correct
PP3 #1	Incorrect	Incorrect
PP3 #2	Incorrect	Incorrect
PP4 #1	Incorrect	Incorrect
PP4 #2	Correct	Incorrect
PP4 #3	Correct	Correct
TNE #1	Correct	N/A
TNE #2	Correct	Incorrect
TNE #3	Correct	Correct
IMA #1	Correct	Correct
IMA #2	Incorrect	Incorrect
IMA #3	Correct	N/A
IMBC #1	Incorrect	Correct
IMBC #2	Correct	Incorrect
IZ #1	Correct	Correct

Figure 15. sgRNA Sanger sequencing results. The results from whole plasmids after Sanger sequencing were entered into the Serial Cloner (Version 2.6.1) gene editing program. The "Element Name" column includes shorthand labels for the target elements. When the sgRNAs were successfully ligated into pCFD5 vector, we screened 2 colonies of the transformed cells for each designed guide sequence. Columns "Template A" and "Template B" are the Sanger sequence results for the 2 colonies of each guide sequence. "Correct" means the guide sequence was successfully inserted in the correct location and correct orientation; "Incorrect" means the guide sequence was not identified in the vector; and "N/A" is for the samples that were lost during the inoculation step of miniprepping. sgRNA = single-guide RNA.

Recently, S. Gisselbrecht and J. Segert (Bulyk laboratory) conducted experiments to confirm that the dCas9 protein was being expressed and that the desired target elements were being efficiently pulled down. To test the efficiency of the chromatin pulldown, they used validated sgRNAs to target the *white* gene. We then collected offspring that should genetically contain the constructs. The chromatin was extracted and purified as described in Methods and then underwent a quantitative polymerase chain

reaction (qPCR) assay. The results of the qPCR assay showed that the *white* gene was successfully enriched in the chromatin extracted from the *Drosophila* embryos which expressed the *white* sgRNA in the N-tagged pUAS vector (Figure 16). This assay shows the specificity of the targeting and indicates a significant increase in chromatin enrichment of the desired target when the sgRNAs are present in vector versus when the sgRNAs are absent.

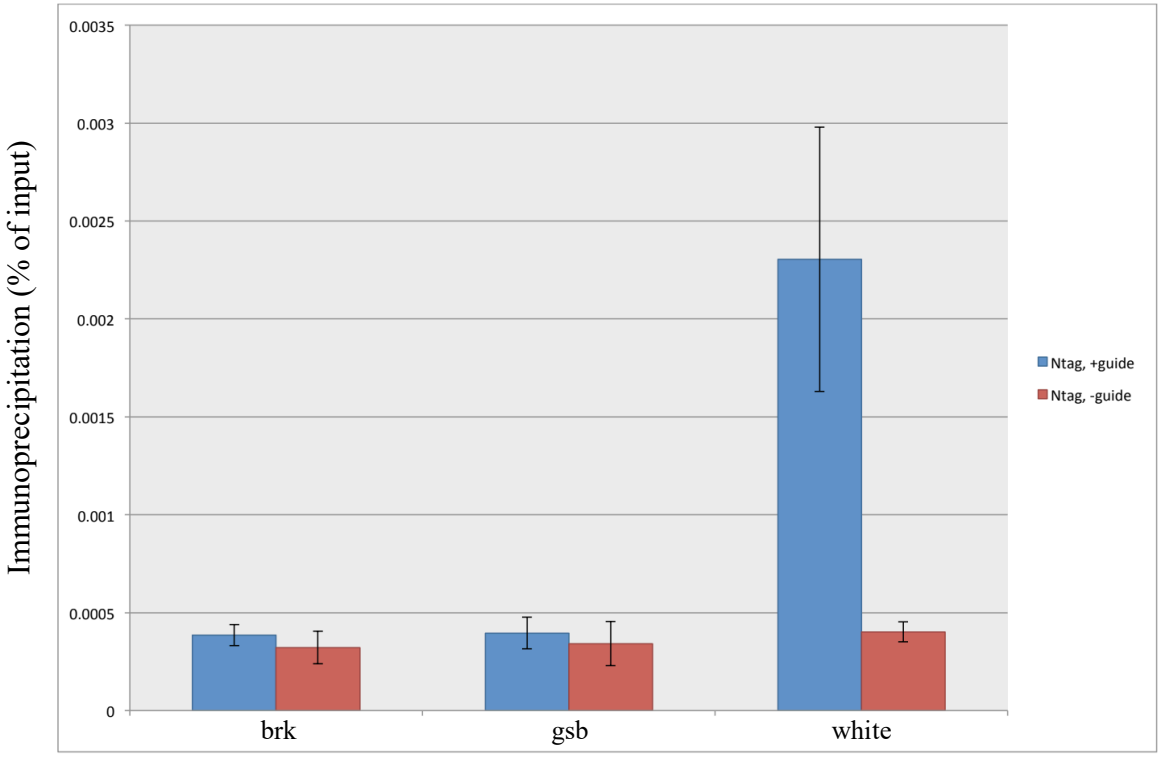

Figure 16. qPCR results for chromatin pulldown efficiency. This figure shows the qPCR results for 3 genes Brk, Gsb, and White. The Brk and Gsb genes are used as controls in this assay, and they represent non-targeted regions of the genome. The graph shows the amount of immunoprecipitation for the desired gene targets after performing a chromatin pulldown when a sgRNA is absent versus when a sgRNA is present. The blue bar is for a Gal4-driven N-tagged pUAS vector with the sgRNA for the desired gene. The red bar is for a Gal4-driven pUAS vector without a sgRNA. qPCR = quantitative PCR; sgRNA = single-guide RNA. Figure taken from (S. Gisselbrecht, personal communication, 02/23/2020).

The results from this qPCR (Figure 16) also demonstrate that we were able to correctly insert dCas9 into the pUAS vector using traditional plasmid cloning with restriction enzymes and that we were able to efficiently target the desired elements to pull down their chromatin. We can now move forward with injecting the correctly sequenced sgRNAs into *Drosophila* and continue to crossbreed the lineage with *Drosophila* containing the twi:gal4 pUAS vector expressing dCas9 in the mesoderm.

DISCUSSION

Although the initial plan to insert dCas9 into the pUAS vector using Gibson Assembly ultimately did not work, we were able to successfully insert the dCas9 endonuclease into the pUAS vector using traditional cloning. We believe that Gibson Assembly did not work in our project because of the fragment sizes. Although the NEB protocol mentions using the method on fragment sizes as large as 15 kb, many scientists with whom I spoke to had success with Gibson Assembly using smaller fragment sizes from 1 to 2 kb. As previously mentioned, our fragment sizes were 4.1 kb and 7.9 kb, which may have been too large for optimal efficiency. If in future experiments we return to this cloning method, we would design different primers and increase the number of bases that create the overlap for annealing. The overlap that our primers created was about 16 bases. However, it was mentioned by NEB that overlaps can be up to 40 bases and that larger overlaps would decrease the amount of DNA necessary for the reaction. This increased overlap would be helpful because one of our problems with this cloning method was low DNA concentrations. In this case, we discovered that traditional cloning was the most optimal plasmid cloning method for our vectors and DNA fragments.

Future Directions

The correctly miniprepped sgRNAs are ready to inject into a yellow, vermillion fly with phiC31 integrase at attP on chromosome III. We will inject the sgRNA along with attB and the wild-type eye gene. We can then screen for non-vermillion eyes to

isolate males with the sgRNA and cross them with virgin females with a balancing chromosome. We will screen the offspring from this cross for the stubble gene (present in the balancing chromosome) and the wild-type eye gene. The resulting offspring can then be self-crossed until the wild-type female with the sgRNA is isolated and kept as a pure stock.

Eventually, crossing flies that contain the Gal4UAS/dCas9 plasmid with flies that contain the sgRNA will produce offspring that contains both the Gal4-expressed pUAS vector with dCas9, along with the sgRNA. After a pure stock is developed, we can then target the desired silencers and pull down the chromatin to study the associated chromatin markers with the previously mentioned screening methods.

As we wait for the transformants with the sgRNAs, we are currently looking into a positive control that we could use for the chromatin pulldown. A positive control would be beneficial to ensure that the chromatin pulldown is efficient. An element that could act as a positive control should be a known repetitive region in the genome. The potential elements we are considering when choosing a positive control include telomeric repeats, histone clusters, and transposable elements. We will continue to consider these options for a positive control when performing the chromatin pulldowns of the targeted elements.

By performing the chromatin pulldown for the silencers, we will collect the embryos at the stage of development when the expression of these elements is the most active. We will then perform the mentioned workflow of purifying the chromatin with IgG agarose beads, followed by the screening assays. The screening methods include western blot, mass spectrometry, ChIP-Seq, and qPCR.

These various screening methods will help us identify chromatin markers and modifications associated with individual silencers and compare them with their enhancer orthologs. The western blot assay allows us to analyze proteins and protein modifications. The proteins of focus are histones in the genome and their modifications, as well as associated co-repressors and repressors. Mass spectrometry is another assay that will permit us to identify co-repressors, repressors, and other post-translational modifications (PTMs). If there are specific patterns seen in histone marks across several silencers, we can further analyze them with ChIP-Seq. ChIP-Seq is an assay that uses antibodies against PTMs and histone characteristics seen across many silencers. The study of the enrichment of ChIP-Seq peaks for the chromatin marks in certain groups of silencers could allow us to identify patterns that might eventually act as a prediction tool for these silencers. Many of these assays will be performed when we detect transformants with the Gal4-driven pUAS vector and sgRNAs and pull down the chromatin for our desired target elements.

Meanwhile, we have been attempting in situ staining to confirm the expression of the pUAS/dCas9 vector in the mesoderm of transformants. To achieve this, we have been discussing different possible antibodies that would be most efficient. This approach would allow us to confirm the location of the tagged dCas9 by analyzing the GFP fluorescence activity. We will observe GFP activity when performing a qPCR assay on the targeted elements, which allows fluorescence to be quantitatively measured. J. Segert (Bulyk laboratory) has previously attempted this approach with various antibodies, including a secondary IgG antibody that could bind to the protein A moiety in the

bioTAP tag bound to dCas9. He tried this method with both an anti-rat antibody and a goat antibody, both of which did not bind well to the tag. We will continue to experiment with different antibodies that are more efficient in imaging the location of dCas9.

To discover more chromatin markers associated with silencers, we could analyze more enhancers and their ability to act as a silencer in different tissues. By observing either non-mesodermal enhancer activity as silencers in mesoderm or mesodermal enhancer activity as silencers in non-mesoderm, we can select more elements to study their bifunctional potential. We can also continue to use qsFS to screen hundreds of regulatory elements at a time, to identify these silencers, and to further study their characterizations with assays.

If chromatin patterns are associated with silencers, ENCODE can profile them, and they can then act as a predictive screening tool for the silencers in *Drosophila*. Although we predict the conservation of these target elements in humans, we would have to test the conservation of the silencer-associated chromatin marks as well. We could test this conservation from *Drosophila* to humans by performing ChIP-Seq and RNA knockout or knockdown assays in human cell cultures. We could also test the silencers for their specificity to tissues or cells by studying the elements in different tissues of the *Drosophila* and human cells. The results of this project can be profiled by ENCODE, along with the already discovered enhancer data, and together this information can act as a predictive tool for silencers in the human genome.

REFERENCES

Addgene. (n.d.). Plasmid Cloning by Restriction Enzyme Digest (with Protocols). Retrieved February 20, 2020, from https://www.addgene.org/protocols/subcloning/

Agilent Technologies. (2015). QuikChange Site-Directed Mutagenesis Kit Instruction Manual. Retrieved February 14, 2020, from https://www.agilent.com/cs/library/usermanuals/public/200518.pdf

Alekseyenko, A. A., McElroy, K. A., Kang, H., Zee, B. M., Kharchenko, P. V., & Kuroda, M. I. (2015). BioTAP-XL - crosslinking/tandem affinity purification to study DNA targets, RNA and protein components of chromatin associated complexes. *Current Protocols in Molecular Biology 109*(1), 1–32. https://doi.org/10.1002/0471142727.mb2130s109

Birla, B., & Chou, H. (2015). Rational Design of High-Number DsDNA Fragments Based on Thermodynamics for the Construction of Full-Length Genes in a Single Reaction. *PLoS ONE 10*(12): e0145682. https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0145682

Consortium, T. E. P. (2011). A User's Guide to the Encyclopedia of DNA Elements (ENCODE). *PLOS Biology*, *9*(4), e1001046. https://doi.org/10.1371/journal.pbio.1001046

Ernst, J., & Kellis, M. (2012). ChromHMM: Automating chromatin-state discovery and characterization. *Nature Methods*, *9*(3), 215–216. https://doi.org/10.1038/nmeth.1906

Franco, H. L., Nagari, A., Malladi, V. S., Li, W., Xi, Y., Richardson, D., Allton, K. L., Tanaka, K., Li, J., Murakami, S., Keyomarsi, K., Bedford, M. T., Shi, X., Li, W., Barton, M. C., Dent, S. Y. R., & Kraus, W. L. (2018). Enhancer transcription reveals subtype-specific gene expression programs controlling breast cancer pathogenesis. *Genome Research*, *28*(2), 159–170. https://doi.org/10.1101/gr.226019.117

Gemenetzi, M., & Lotery, A. J. (2020). Epigenetics in age-related macular degeneration: New discoveries and future perspectives. *Cellular and Molecular Life Sciences*. https://doi.org/10.1007/s00018-019-03421-w

Gregory, T. R. (2005). *The Evolution of the Genome*. Elsevier Science & Technology. http://ebookcentral.proquest.com/lib/bu/detail.action?docID=286753

Hadfield, J. (2012). CoreGenomics: How do SPRI beads work? *CoreGenomics*. http://core-genomics.blogspot.com/2012/04/how-do-spri-beads-work.html

Ho, J., Jung, Y., Liu, T., Alver, B., Lee, S., Ikegami, K., Sohn, K., et al. (2014). Comparative Analysis of Metazoan Chromatin Organization. *Nature,* 512(7515), 449–452. https://doi.org/10.1038/nature13415.

Iizasa, E., & Nagano, Y. (2006). Highly efficient yeast-based in vivo DNA cloning of multiple DNA fragments and the simultaneous construction of yeast/Escherichia coli shuttle vectors. *BioTechniques*, *40*(1), 79–83. https://doi.org/10.2144/000112041

Isogawa, A., Fuchs, R. P., & Fujii, S. (2018). Versatile and efficient chromatin pull-down methodology based on DNA triple helix formation. *Scientific Reports*, *8*(1), 1–12. https://doi.org/10.1038/s41598-018-24417-9

Krivega, I., & Dean, A. (2012). Enhancer and promoter interactions—Long distance calls. *Current Opinion in Genetics & Development*, *22*(2), 79–85. https://doi.org/10.1016/j.gde.2011.11.001

Li, Y., Chen, C., Kaye, A. M., & Wasserman, W. W. (2015). The identification of cis-regulatory elements: A review from a machine learning perspective. *Biosystems*, *138*, 6–17. https://doi.org/10.1016/j.biosystems.2015.10.002

Lin, S., Zhou, M., Li, Y., Chen, Y., Xu, W., Xia, W., Wang, S., Yin, R., Zhang, Q., Xu, L., & Fan, H. (2019). H3K27 trimethylation and H3K9 dimethylation as poor prognostic markers for patients with esophageal squamous cell carcinoma. *International Journal of Clinic & Experimental Medicine*, 12(7), 2657–2664. https://www-ncbi-nlm-nih-gov.ezproxy.bu.edu/pmc/articles/PMC6949571/

Lloyd, T. E., & Taylor, J. P. (2010). Flightless Flies: Drosophila models of neuromuscular disease. *Annals of the New York Academy of Sciences*, *1184*, e1-20. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3062507/

Maston, G. A., Evans, S. K., & Green, M. R. (2006). Transcriptional Regulatory Elements in the Human Genome. *Annual Review of Genomics and Human Genetics*, *7*(1), 29–59. https://doi.org/10.1146/annurev.genom.7.080505.115623

National Institutes of Health. (2020). *What is epigenetics?* Genetics Home Reference. https://ghr.nlm.nih.gov/primer/howgeneswork/epigenome

New England Biolabs. (n.d.). DpnI. https://www.neb.com/products/r0176-dpni#Product%20Information

Pandey, U. B., & Nichols, C. D. (2011). Human Disease Models in Drosophila melanogaster and the Role of the Fly in Therapeutic Drug Discovery. *Pharmacological Reviews*, *63*(2), 411–436. https://doi.org/10.1124/pr.110.003293

Pennisi, E. (2012). ENCODE Project Writes Eulogy for Junk DNA. *Science*. https://science-sciencemag-org.ezproxy.bu.edu/content/337/6099/1159

Port, F., & Bullock, S. L. (2016). Augmenting CRISPR applications in Drosophila with tRNA-flanked sgRNAs. *Nature Methods*, *13*(10), 852–854. https://doi.org/10.1038/nmeth.3972

Rogers, J. M., Barrera, L. A., Reyon, D., Sander, J. D., Kellis, M., Joung, J. K., & Bulyk, M. L. (2015). Context influences on TALE–DNA binding revealed by quantitative profiling. *Nature Communications*, *6*. https://doi.org/10.1038/ncomms8440

Sha, K., & Boyer, L. A. (2009). The chromatin signature of pluripotent cells. *StemBook*. https://doi.org/10.3824/stembook.1.45.1

Shu,K. and Ueda, R. (2013). Highly Improved Gene Targeting by Germline-Specific Cas9 Expression in Drosophila. *Genetics, 195*(3), 715–721. https://doi.org/10.1534/genetics.113.156737.

Stevenson, J., Krycer, J. R., Phan, L., & Brown, A. J. (2013). A Practical Comparison of Ligation-Independent Cloning Techniques. *PLoS ONE*, *8*(12). https://doi.org/10.1371/journal.pone.0083888

Tamaru, H. (2010). Confining euchromatin/heterochromatin territory: Jumonji crosses the line. *Genes & Development*, *24*(14), 1465–1478. https://doi.org/10.1101/gad.1941010

Wei, P., Xue, W., Zhao, Y., Ning, G., & Wang, J. (2020). CRISPR-based modular assembly of a UAS-cDNA/ORF plasmid library for more than 5500 Drosophila genes conserved in humans. *Genome Research*, *30*(1), 95–106. https://doi.org/10.1101/gr.250811.119

Williams, S. C. P. (2013). Epigenetics. *Proceedings of the National Academy of Sciences of the United States of America*, *110*(9), 3209. https://doi.org/10.1073/pnas.1302488110

Zhang, Y., Hu, J.-F., Wang, H., Cui, J., Gao, S., Hoffman, A. R., & Li, W. (2016). CRISPR Cas9-guided chromatin immunoprecipitation identifies miR483 as an

epigenetic modulator of IGF2 imprinting in tumors. *Oncotarget*, *8*(21), 34177–34190. https://doi.org/10.18632/oncotarget.10918

CURRICULUM VITAE