2017-06-01

The Effects of Nucleosome Positioning and Chromatin Architecture on Transgene Expression

Colton E. Kempton
Brigham Young University

The Effects of Nucleosome Positioning and Chromatin Architecture on Transgene Expression

Colton E. Kempton

A dissertation submitted to the faculty of
Brigham Young University
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

Steven M. Johnson, Chair
Laura C. Bridgewater
Joel S. Griffitts
Brent Nielsen
Joshua A. Udall

Department of Microbiology and Molecular Biology

Brigham Young University

ABSTRACT

The Effects of Nucleosome Positioning and Chromatin Architecture on Transgene Expression

Colton E. Kempton
Department of Microbiology and Molecular Biology, BYU
Doctor of Philosophy

Eukaryotes use proteins to carefully package and compact their genomes to fit into the nuclei of their individual cells. Nucleosomes are the primary level of compaction. Nucleosomes are formed when DNA wraps around an octamer of histone proteins and a nucleosome's position can limit access to genetic regulatory elements. Therefore, nucleosomes represent a basic level of gene regulation. DNA and its associated proteins, called chromatin, is usually classified as euchromatin or heterochromatin. Euchromatin is transcriptionally active with loosely packed nucleosomes while heterochromatin is condensed with tightly packed nucleosomes and is transcriptionally silent. In order to become active, heterochromatin must first be remodeled. We have studied the effects of nucleosome positioning on transgene expression *in vivo* using *Caenorhabditis elegans* as a model. We show that both location and polarity of the DNA sequence can influence transgene expression. We also discuss some considerations for working with CRISPR/Cas9.

A major reason for doing *in vitro* nucleosome reconstitutions is to determine the effects of DNA sequence on nucleosome formation and position. It has previously been implied that nucleosome reconstitutions are stochastic and not very reproducible. We show that nucleosome reconstitutions are highly reproducible under our reaction conditions. Our results also indicate that a minimum depth of 35X sequencing coverage be maintained for maximal gains in Pearson's correlation coefficients.

Communicating science with others is an important skill for any researcher. The rising generation of scientists need mentors who can teach them how to be independent thinkers who can carry out scientific experiments and communicate their finding to others. With this goal in mind, we have devised a scaffolding pedagogical method to help transform undergraduates into confident independent thinkers and researchers.

Keywords: nucleosome, CRISPR/Cas9, *Caenorhabditis elegans*

ACKNOWLEDGEMENTS

I would like to thank my mentor Dr. Steven Johnson for all of his teaching and support during this experience. I would also like to thank the other members of my committee, Dr. Laura Bridgewater, Dr. Joel Griffitts, Dr. Brent Nielsen, and Dr. Joshua Udall for their support and guidance.

I would also like to thank Dr. Jonathan Alder for his help with my CRISPR/Cas9 experiments and for providing purified Cas9 to work with.

I would especially like to thank my wife Jenny for all of her support in this endeavor and acknowledge that without her encouragement, it would not have been possible.

Table of Contents

List of Figures

List of Tables

List of Appendices

Chapter 1. Introduction

1.1 CHROMATIN REGULATION

Human nuclei are 5 micrometers in diameter, but contain almost 2 meters of DNA. In order to fit inside, DNA must be highly compacted. However, compaction must be done in an organized manner for cells to maintain access to required genes. Eukaryotes use proteins to organize and condense their DNA into chromatin and nucleosomes are the fundamental level of compaction.

Nucleosomes are formed when DNA wraps 1.7 times (~147bp) around an octamer core of histone proteins [1]. Nucleosome formation on DNA results in a seven-fold compaction of DNA. The physical positions of nucleosomes on DNA are important because nucleosomes can limit access to regulatory elements in a genome's sequence. Therefore, nucleosomes represent a basic level of gene regulation and play a key role in chromatin formation.

Chromatin can be classified as euchromatin or heterochromatin. Euchromatin has been described as open, accessible, and transcriptionally active with loosely packed nucleosomes. It has also been observed in multiple organisms including humans that actively transcribed genes have nucleosome free regions just upstream of their transcriptional start sites [2]. This observation supports the idea that nucleosomes are involved in regulation of gene expression. Conversely, heterochromatin has been described as condensed, inaccessible, and transcriptionally silent with tightly packed nucleosomes. In order for silent chromatin to become transcriptionally active, it must first be remodeled. Nucleosomes are very stable *in vivo* unless they are actively remodeled by ATP-dependent chromatin remodelers (*e.g.* ISWI, CHD) [3]

If it were possible to manipulate chromatin architecture and nucleosome positions, it would be possible to alter gene transcription. This would be a very useful tool and could be very important for applications like gene therapy. The primary objective of gene therapy is to

permanently correct genetic defects. To achieve this, functional genes must be delivered to specific tissues and expressed at physiologically relevant levels. Currently gene therapy is limited in its usefulness because it has been difficult to safely deliver the therapeutic genes and maintain their expression *in vivo.* Many delivery and expression methods and systems are currently being developed to deliver these genes (*e.g.* plasmids and viral vectors) [4, 5]. However, despite delivery of functional copies of genes in some approaches, the expression of these genes is not maintained long term and they are eventually silenced [6-8]. Silencing of therapeutic genes leads to a relapse in the patient's condition. A better understanding of how nucleosomes are positioned could potentially alleviate this problem as well as increase our general understanding of how genes are regulated.

DNA sequence influences how and where nucleosomes sit. The persistence length (the length for which it behaves more like a rod than a string) of DNA is longer than the diameter of the histone octamer and DNA must be bent to wrap around it. DNA is not a homopolymer and each dinucleotide step has its own unique stereochemistry [9]. Consequently, DNA is anisotropic, and DNA sequences with intrinsic bending would help facilitate wrapping around the histone octamer. Indeed, it has been shown that having AA/TT dinucleotides spaced every 10 bp or every turn of the helix increases intrinsic bending and allows the histone octamer to bind with increased affinity [9, 10]. However, homopolymeric runs of A/T are recalcitrant to nucleosome formation [11]. The extent to which nucleosome positions are dictated by sequence or other factors is still debated.

We have studied the effects of nucleosome positioning on transgene expression *in vivo* using the nematode *Caenorhabditis elegans* as a model. The nucleosome positioning sequences we tested included the 601 sequence [10] and the Trifonov sequence [12]. The Trifonov

sequence was derived by analyzing a large dataset of *C. elegans* nucleosome cores [13]. Positional preferences of the different dinucleotide steps within the nucleosome cores from all six chromosomes in *C. elegans* were calculated and a "bendability pattern" was derived from these calculations [12]. The 601 sequence is a synthetically derived sequence that strongly positions nucleosomes *in vivo* [10] and is a standard in the nucleosome positioning field. Both of these sequences were included in our nucleosome positioning experiments.

Transgenic lines of *C. elegans* can be created fairly easily by microinjection into the syncytial arm of the gonad. Linear and/or circular DNA is taken up into developing oocytes where it is concatamerized and forms an extrachromosomal array. Worm progeny inherit the array(s) in a non-Mendelian manner in subsequent generations unless integration occurs [14, 15]. In transgenic worms, transgene expression from these extrachromosomal arrays can vary when transgenic lines are maintained for multiple generations and can even be ultimately silenced despite the continued presence of the array. Gradual silencing of transgenes in *C. elegans* is reminiscent of the current dilemma faced in gene therapy, thus providing a simple system in which this phenomenon can be studied.

1.2 SUMMARY OF RESEARCH CHAPTERS

Chapter 2 outlines our efforts to identify the effects that underlying DNA sequence has on nucleosome formation and its consequences for transgene expression. We also identify different patterns and aspects of affecting transgene expression through DNA manipulation. Furthermore, we discuss some of the practical aspects and merits of using CRISPR/Cas9 for extrachromosomal array integration in *C. elegans* as compared to more traditional integration methods like gamma irradiation. In Chapter 3 we discuss the purposes and reproducibility of *in vitro* nucleosome reconstitution experiments. We also make recommendations regarding the

depth of sequencing coverage that should be maintained when evaluating the effects of the underlying DNA sequence on nucleosome formation in the absence of other factors. Chapter 4 details a method for improving undergraduate science education. We discuss a scaffolding pedagogical method to improve students' ability and confidence to plan and carry out independent research and communicate their findings with their peers. In Chapter 5 we discuss the findings of this research as well as some of the questions raised during the course of this research. Future directions and strategies are also discussed.

Chapter 2. Nucleosome Positioning Experiments and CRISPR Experiments

2.1 NUCLEOSOME POSITIONING EXPERIMENTS

Developing a system to evaluate positioning sequences

We used the model organism *Caenorhabditis elegans* to study the effects of chromatin architecture and nucleosome positioning on transgene expression and silencing in somatic cells *in vivo*. In *C. elegans* there are four genes that encode myosin heavy chains. They are *myo-1*, *myo-2, unc-54,* and *myo-3.* The genes *myo-1* and *myo-2* are expressed in pharyngeal muscle cells while *unc-54* and *myo-3* are expressed in body-wall muscle cells [16]. Each gene has an enhancer element as well as a tissue specific promoter. These enhancer elements were defined by their ability to activate a heterologous promoter [17].

One such heterologous enhancer and promoter construct is present in the plasmid pPD151.79 (obtained from Andrew Fire). See Figure 2-1 and Appendix C. Observations of transgenic lines show robust tissue-specific transgene expression that gradually diminishes over multiple generations. In this construct, GFP is expressed as a lacZ fusion protein and is under the control of the *myo-2* promoter and the *unc-54* enhancer element. The *myo-2* promoter constitutively induces GFP::*lacZ* expression specifically in pharyngeal muscle cells (Figure 2-2). The *unc-54* enhancer expands this expression specifically to the body-wall muscle cells [18]. Transgene expression in the pharynx of the worm remains active while transgene expression in body-wall muscle cells undergoes silencing when lines are maintained for multiple generations. See Figure 2-3 for an example. Few details are known about transgene expression from extrachromosomal arrays and the phenomenon of transgene silencing on these arrays is also ambiguous [19]. We hypothesize that transgene silencing is due to the enhancer being

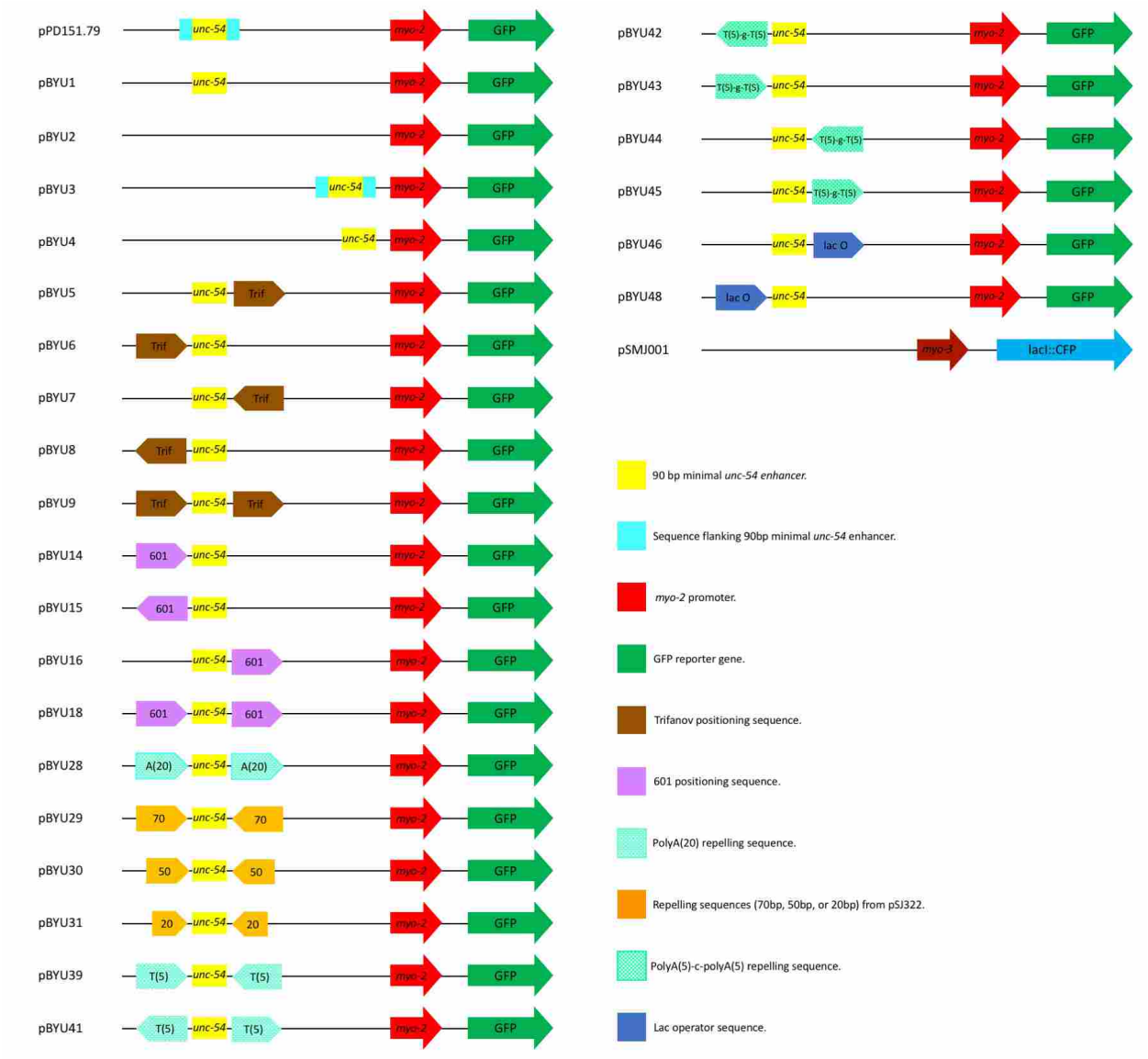


Figure 2-1 Nucleosome positioning/repelling constructs.

Nucleosome positioning/repelling sequences and their orientations relative to the 90bp minimal *unc-54* enhancer and the *myo-2* promoter.

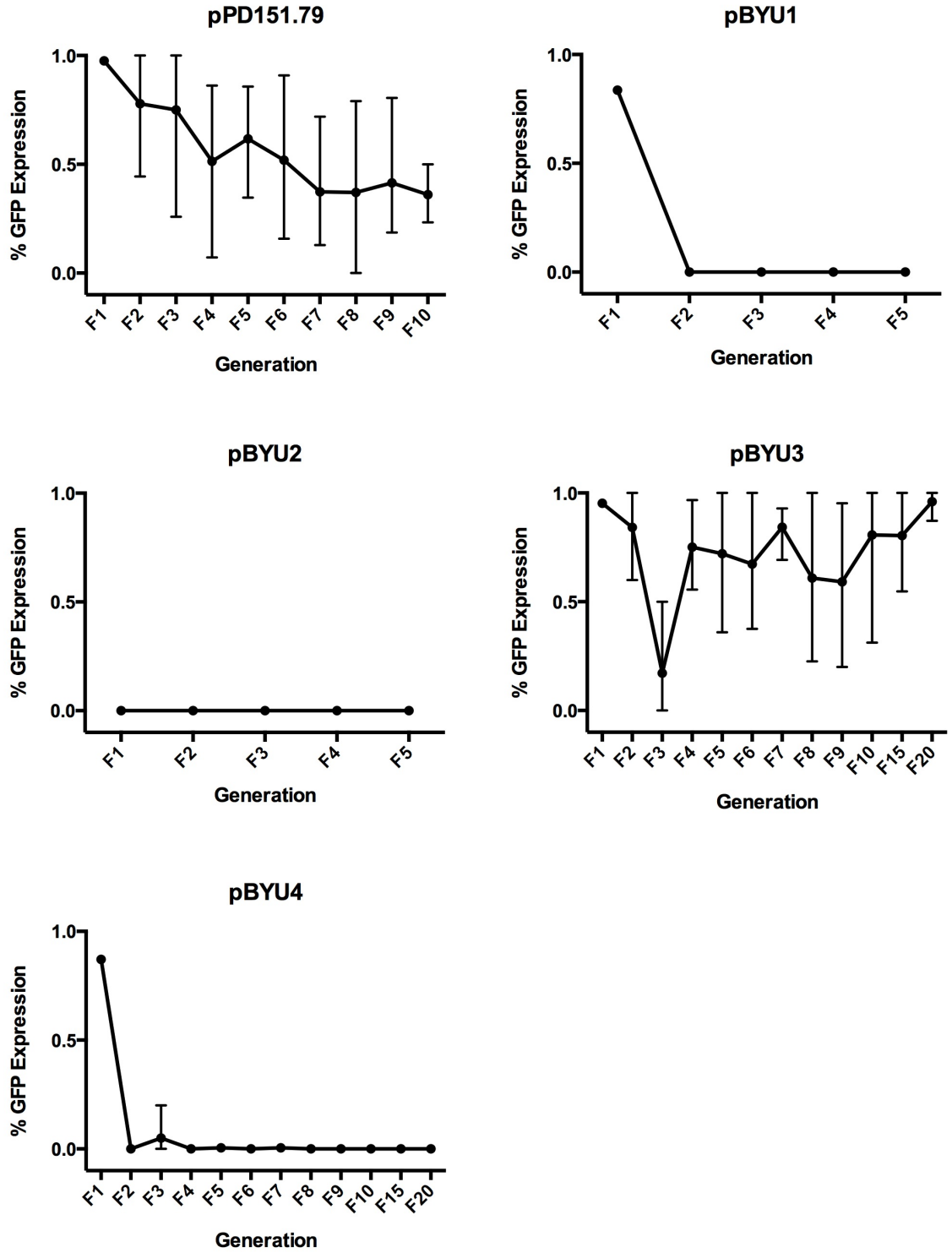


Figure 2-2 GFP expression in body-wall muscle cells over multiple generations.

GFP expression in body-wall muscle cells over multiple generations for select plasmids. All plasmids except pPD151.79 are derived from pBYU1. See Figure 2-1.

sequestered in nucleosomes and the subsequent formation of heterochromatin. In our experiments, we have attempted to prevent the silencing of GFP expression in body-wall muscle cells by manipulating the DNA sequence surrounding the *unc-54* enhancer.

In these experiments worms were said to be GFP positive for body-wall muscle cells or pharyngeal muscle cells, if GFP could be visually detected in any of the nuclei of either muscle cell type. For each construct, transgenic F_1 animals were first identified by a dominant roller phenotype conferred by the co-injection marker. Each animal was then scored for GFP expression in body-wall and pharyngeal muscle cells. F_1 N-values and GFP percentages were calculated based on the entire population of transgenic F_1 animals. GFP expression levels for later generations were calculated by scoring 25-40 animals from each generation that was reported. Three or more biological replicates were done for each construct tested.

The minimal unc-54 enhancer

The *unc-54* enhancer has been defined as a 90bp sequence with modular properties [18]. However, pPD151.79 not only had this 90bp sequence, but it also had extra flanking sequence from *unc-54* intron-3 (80bp upstream and 145bp downstream) where the enhancer is located *in vivo*. In order to sterically hinder nucleosome formation on the enhancer, we wanted to position nucleosomes flanking the enhancer and leave insufficient room for a nucleosome to form on the enhancer itself. Thus, we removed the enhancer with flanking sequence and replaced it with the 90bp minimal enhancer and renamed it pBYU1. See pPD151.79 and pBYU1 Appendix C. Since 90bp is insufficient for nucleosome formation, we hypothesized that the 90bp enhancer would remain nucleosome free.

The 90bp *unc-54* enhancer turned out to be more useful than the enhancer with the concomitant flanking sequence. Expression data for pPD151.79 showed a gradual decrease for

GFP expression over multiple generations. This gradual silencing was highly variable between isolated worm lines but the trend showed a decrease in the occurrence of GFP body wall expression. However, pBYU1 expression data showed that GFP body wall expression had been completely silenced by the second generation. See Figure 2-2. This worked to our advantage as it allowed us to test different positioning sequences without needing to maintain and score the worms for multiple generations.

In order to verify that the 90bp enhancer element behaved like an enhancer and was responsible for the GFP expression in body-wall muscle cells, we created several deletion vectors. When the enhancer element was completely removed (pBYU2), GFP expression was not observed in the body-wall muscle cells, although pharyngeal GFP expression was unaffected.

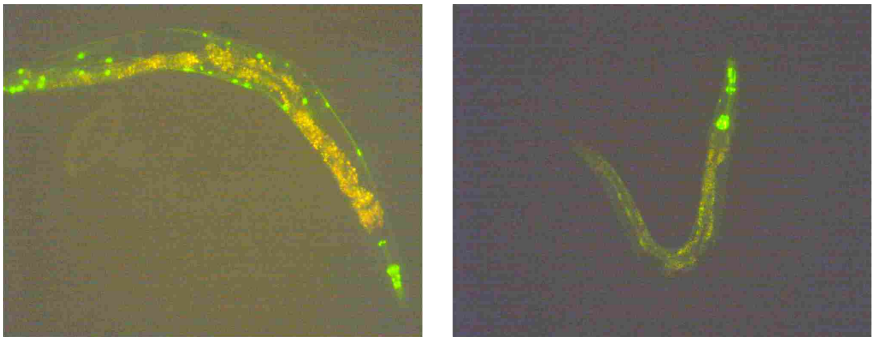


Figure 2-3 GFP expression in body-wall muscle cells and pharyngeal muscle cells.

GFP expression in body-wall muscle cells and pharynx (left). GFP expression in pharyngeal muscle cells only (right).

To see if the enhancer was position independent, we deleted the intervening DNA sequence between the enhancer and the *myo-2* promoter (pBYU3 & pBYU4). Body wall GFP expression for pBYU3 was erratic. For pBYU4 body wall GFP expression behaved nearly identically to pBYU1 except that a few animals showed body wall expression in the F_3 generation. We are unsure why expression was so erratic in pBYU3 or why the trend for

silencing over time as compared to pPD151.79 was lost. See Figure 2-2. These results confirmed that the 90bp is position independent and it is required for GFP expression in body-wall muscle cells. Despite the binary on-and-off of GFP expression in the body-wall cells for pBYU1, GFP expression in the pharynx remained unchanged and served as a good internal control to show that the transgene was not lost. In an effort to keep the *unc-54* enhancer in an active state and extend GFP expression beyond the F_1 generation, we tested whether it was possible to accomplish this by using DNA sequences inserted into pBYU1 upstream and or downstream of the enhancer to influence local nucleosome positions.

Effects of various positioning and repelling sequences on gene expression

We primarily used two different strategies to exclude nucleosomes from the enhancer in our GFP reporter constructs. First we used nucleosome positioning sequences. These sequences were placed upstream and or downstream of the enhancer so as to create well positioned nucleosomes that would occlude nucleosomes from forming in the intervening space where the enhancer is located. We tested the 601 sequence [20] and the Trifonov sequence [9, 12]. The 601 sequence is recognized as an important standard against which all other positioning sequences are measured. *In vitro* it can reliably position nucleosomes. It has also been reported that it can also position nucleosomes *in vivo,* although this positioning was transient. Loss of positioning also corresponded with loss of transgene expression. [21]. The Trifonov sequence was derived based on observed nucleosome positioning patterns in *C. elegans* and intrinsic DNA bendability.

Our second strategy was to use nucleosome repelling sequences to "push" nucleosomes away from the enhancer [*e.g.* polyA(20) sequence]. We also tried to use statistical positioning via a barrier to extend reporter expression. When a nucleosome encounters a barrier (*e.g.* bound transcription factor) it is positioned next to that barrier. This causes neighboring nucleosomes to

form a regularly spaced array of neighboring nucleosomes that radiates out from the barrier. These positioned nucleosomes are said to be positioned due to statistical positioning [22]. We also tried to establish a phased nucleosome array by using a CFP::lacI fusion protein bound to a *lacO* sequence to create a barrier that would position neighboring nucleosomes. See Table 2-1 for descriptions of theses positioning and repelling sequences.

While we were able to influence the frequency of GFP expression in the worms with each of these methods, the results were not what we expected. See Figures 2-4 and 2-5. With the exception of pBYU45, all positioning and repelling sequences tested caused equal or less GFP expression in body-wall muscle cells as compared to pBYU1. No body wall GFP expression could be detected beyond the F_1 generation for any of the constructs tested. However, these experiments did reveal some important principles about nucleosome positioning that will prove helpful in future experiments.

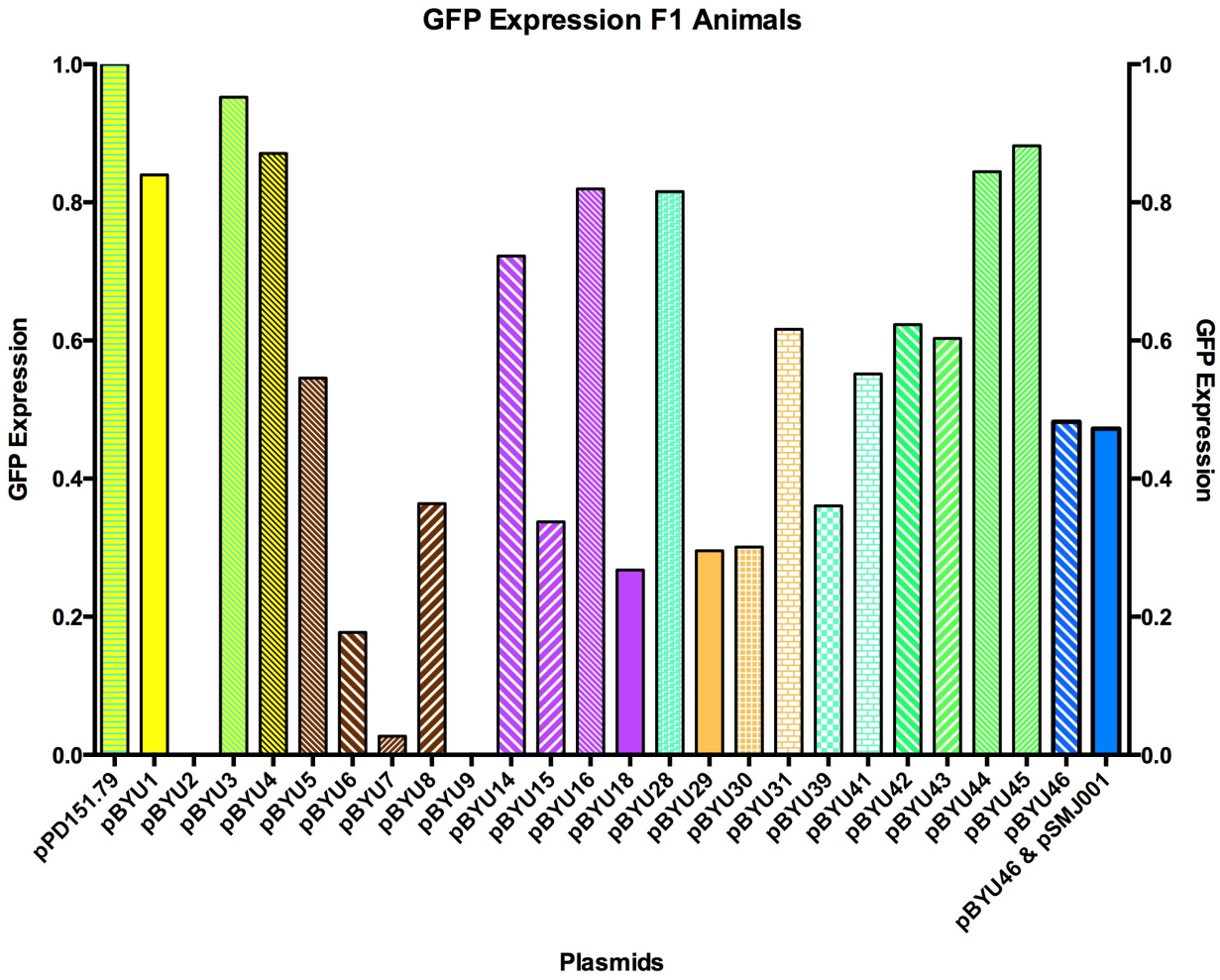


Figure 2-4 GFP expression in body-wall muscle cells for populations of F_1 animals

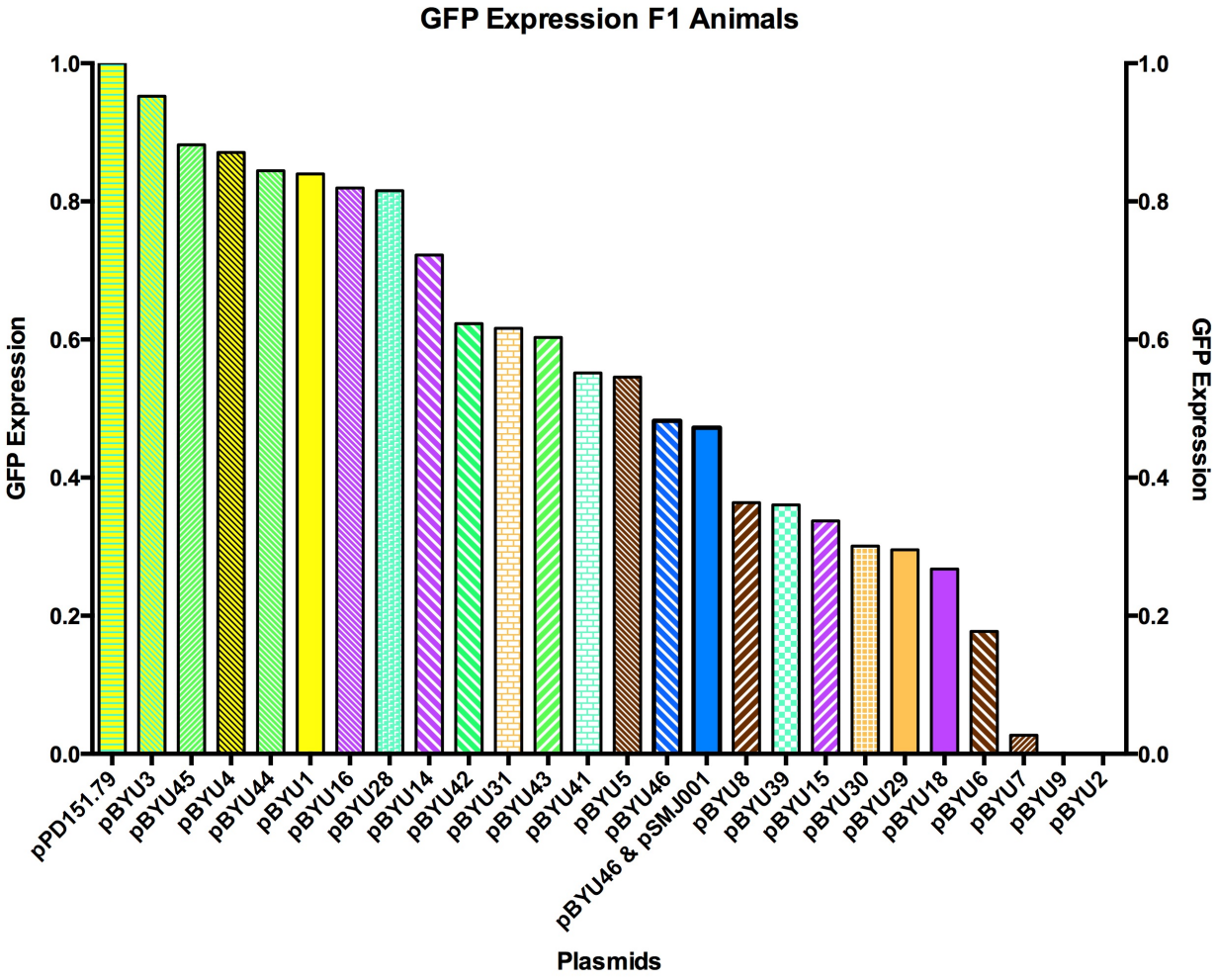


Figure 2-5 GFP expression in body-wall muscle cells in F_1 populations sorted by expression.

Table 2-1 Plasmids by insert type

Insert Type	Insert	Direction		Plasmid	F_1 N-value	% GFP F_1 Body Wall
		HindIII	NheI	pBYU1	50	84
Positioning	601		for	pBYU16	72	82
	601	for		pBYU14	108	72
	Trifonov		for	pBYU5	22	55
	Trifonov	rev		pBYU8	220	36
	601	rev		pBYU15	323	34
	601	for	for	pBYU18	71	27
	Trifonov	for		pBYU6	265	18
	Trifonov		rev	pBYU7	111	3
	Trifonov	for	for	pBYU9	90	0
Repelling	T(5)-G-T(5)		for	pBYU45	263	88
	T(5)-G-T(5)		rev	pBYU44	225	84
	A(20)	for	for	pBYU28	65	82
	T(5)-G-T(5)	rev		pBYU42	69	62
	20bp_pSJ322	for	rev	pBYU31	310	62
	T(5)-G-T(5)	for		pBYU43	388	60
	T(5)	ref	for	pBYU41	136	55
	T(5)	for	rev	pBYU39	61	36
	50bp_pSJ322	for	rev	pBYU30	206	30
	70bp_pSJ322	for	rev	pBYU29	88	30
Barrier	lacO		for	pBYU46	141	48
	lacO		for	pBYU46, pSMJ001	72	47
	lacO	for		PBYU48, pSMJ001	135	8

Location matters

Our data show that the location of the test sequence has a clear effect on GFP expression in F_1 populations. A *Hind*III restriction enzyme site is located on the 5' side of the minimal 90bp *unc-54* enhancer and a *Nhe*I restriction enzyme site is on the 3' side. See Appendix C Figure C-2. A comparison between plasmids that vary only with respect to the location of the insert (*Hind*III or *Nhe*I), show that the location of the insert can have a substantial influence on reporter expression. Variation by location was strongest for the Trifonov sequence. In the forward direction, there was a difference of 37% between the populations of F_1 GFP positive animals. Compare pBYU5 and pBYU6. For the reverse direction, this difference was 33%. Compare pBYU7 and pBYU8 Table 2-2. The repelling sequence T(5)-G-T(5) also showed a significant effect when varied only by position. Compare pBYU43 with pBYU45 and pBYU42 with pBYU44 Table 2-2. Surprisingly, the 601 sequence showed the smallest difference in its effects when varied only by location and it had the smallest suppressive effect on GFP expression in first generation animals. Compare pBYU14 and pBYU16 Table 2-2. However, this comparison could only be made for the forward orientation due to a lack of data. These results are summarized in Table 2-2. Overall these results demonstrate that the locations of nucleosome positioning signals relative to genetic regulatory elements like the *unc-54* enhancer are very important.

Effects of sequence polarity

Expression data shows that positioning and repelling sequences are polar and that these sequences vary in their effects based on location and orientation. While the 601 sequence did not show significant results based on location for the forward orientation (Table 2-2), there was a

significant difference based on orientation. A comparison of expression data for pBYU14 and pBYU15 showed a difference of 38% in expression levels. A significant difference in expression was also seen for plasmids with the Trifonov insert in different orientations. Compare pBYU6 with pBYU6 and pBYU5 with pBYU7 Table 2-3. This was especially true for pBYU5 and pBYU7 which showed a 52% difference in expression between alternate orientations at the *Nhe*I site. However, for plasmids with the T(5)-G-T(5) repelling sequence, there was no detectable significant difference based on polarity. Compare pBYU42 with pBYU43 and pBYU44 with pBYU45 Table 2-3.

It is interesting to note that difference in expression levels seen when varying the sequence polarity was not the same for *Hind*III and *Nhe*I sites. This is especially true when comparing the effects of polarity at the different insert locations for the Trifonov sequence. Plasmids pBYU5 and pBYU7 show a 52% difference in expression for the *Nhe*I site while there is only an 18% difference for pBYU6 and pBYU8 at the *Hind*III site. This indicates that expression levels can be affected to a greater degree at the *Nhe*I site than they can at the *Hind*III site. See Table 2-3 for experimental results. While this is true for the Trifonov sequence, it is still unclear if it is true for the 601 sequence as there is no data for a 601 sequence in the reverse orientation at the *Nhe*I site for comparison with other 601 vectors.

It has been shown that poly-dA:dT tracts form an asymmetric barrier to nucleosome movement in mouse, human, and yeast chromatin [23]. If there is some asymmetry with our positioning inserts like the Trifonov sequence, it would explain why sequence polarity can have such a great effect. It also means that a more accurate method for evaluating the merits of the *Hind*III insert site compared to the *Nhe*I site would be to compare plasmids whose inserts have opposite polarities at those sites. In this way, nucleosomes that were being shifted towards or

away from the enhancer at the *Nhe*I location would also undergo the same shift at the *Hind*III location. See Table 2-4. When compared this way, the differences in expression levels for Trifonov insert comparisons are still similar to each other, 19% (pBYU6 with pBYU7) and 15% (pBYU5 with pBYU8) in Table 2-4 compared to 37% (pBYU5 and PBYU6) and 33% (pBYU7 and pBYU8) in Table 2-2.

Compound effects of positioning and repelling elements

We have shown that location and polarity are important aspects to consider when using positioning/repelling to regulate gene expression. These aspects can be used together to cause greater effects than either can alone. For example, in pBYU7, the Trifonov insert in the reverse orientation at the *Nhe*I site has the strongest effect on reporter expression compared to any other single positioning or repelling element tested so far. See Table 2-4. The difference in expression at the *Nhe*I site is statistically significant when accounting for sequence polarity. See Table 2-4.

Using positioning inserts at both the *Hind*III and *Nhe*I sites has a greater effect on expression levels than using either site alone. Plasmids pBYU18 and pBYU9 which contain two 601 or Trifonov inserts respectively, both show a stronger suppression on F_1 expression levels than any of the other single 601 or Trifonov insertion plasmids. See Table 2-1. Based on observations for the 601 sequence, the construct with maximum suppression would have a reverse 601 sequence at the *Hind*III site and a forward sequence at the *Nhe*I site. For the Trifonov sequence inserts would be in the forward orientation at the *Hind*III site and the reverse orientation at the *Nhe*I site. However, since pBYU9 already has 0% GFP expression in first generation animals, this cannot be tested. However, the 601 vector could be constructed to test if those combined orientations confer the greatest suppression.

Why are these constructs contributing to the innate silencing of our transgene? With the exception of pBYU45, all current insert configurations tested resulted in either no change or increased transgene silencing in F_1 animals. Plasmid pBYU45 had slightly higher expression in the first generation but this change was not statistically significant. What factors other than DNA sequence are affecting expression? Further research needs to be done to answer this question. We chose to do a forward genetic screen to help answer this question. To that end, we needed to integrate our extrachromosomal array into the *C. elegans* genome to produce a stable transgenic line.

Table 2-2 Positional Effects of Various Inserts

Insert Type	Insert	Direction	Location HindIII	Location NheI	N-value	% GFP	Location |Difference| %	**Relative % GFP
Positioning	601	for		pBYU16	72	82	10	-2
	601	for	pBYU14		108	72		-10
	Trifonov	for	pBYU6		265	18	37*	-55
	Trifonov	for		pBYU5	22	55		-24
	Trifonov	rev	pBYU8		220	36	33*	-40
	Trifonov	rev		pBYU7	111	3		-68
Repelling	T(5)-G-T(5)	for		pBYU45	263	88	28*	3
	T(5)-G-T(5)	for	pBYU43		388	60		-20
	T(5)-G-T(5)	rev		pBYU44	225	84	22*	0
	T(5)-G-T(5)	rev	pBYU42		69	62		-18

*Statistically significant

** Relative % GFP indicates the reduction in GFP expression for F_1 animals as compared to pBYU1 levels.

Table 2-3 Effects of insert polarity

			Location				Polarity	
Insert Type	Insert	Direction	HindIII	NheI	N-value	% GFP	|Difference|	**Relative % GFP
Positioning	601	for	pBYU14		108	72	38*	-10
	601	rev	pBYU15		323	34		-42
	Trifonov	for	pBYU6		265	18	18*	-55
	Trifonov	rev	pBYU8		220	36		-40
	Trifonov	for		pBYU5	22	55	52*	-24
	Trifonov	rev		pBYU7	111	3		-68
Repelling	T(5)-G-T(5)	for	pBYU43		388	60	2	-20
	T(5)-G-T(5)	rev	pBYU42		69	62		-18
	T(5)-G-T(5)	for		pBYU45	263	88	4	3
	T(5)-G-T(5)	rev		pBYU44	225	84		0

*Statistically significant

** Relative % GFP indicates the reduction in GFP expression for F1 animals as compared to pBYU1 levels.

Table 2-4 Alternative evaluation of positional effects

			Location				Location & Polarity	
Insert Type	Insert	Orientation	HindIII	NheI	N-value	% GFP	|Difference|	Relative % GFP
Positioning	601	for		pBYU16	72	82	48*	-2
	601	rev	pBYU15		323	34		-42
	Trifonov	for	pBYU6		265	18	15*	-55
	Trifonov	rev		pBYU7	111	3		-68
	Trifonov	for		pBYU5	22	55	19	-24
	Trifonov	rev	pBYU8		220	36		-40

*Statistically significant

** Relative % GFP indicates the reduction in GFP expression for F1 animals as compared to pBYU1 levels.

2.2 INTEGRATION EXPERIMENTS

CRISPR/Cas9 an alternative to classic integration

The CRISPR/Cas9 system originally evolved as a prokaryotic defense mechanism against viruses and is a type of rudimentary adaptive immune response by prokaryotes against viruses [24]. Cas9 is an RNA-guided DNA endonuclease. Clustered regularly interspaced palindromic repeats (CRISPRs) are found in most prokaryotes but are absent from viruses and eukaryotes and are always located next to Cas9 genes [25]. The spacers in these heritable CRISPR arrays contain the sequences for targeting Cas9. Successful cutting by Cas9 complexed with a guide RNA requires sequence homology plus a proto-spacer adjacent motif or PAM sequence [26]. The CRISPR/Cas9 system has been used extensively in many model organisms in recent years for site specific genome editing. We decided to use the CRISPR/Cas9 system to integrate our extrachromosomal array.

Several techniques have been employed historically to integrate extrachromosomal arrays in *C. elegans* (*e.g.* gamma and UV radiation). These techniques cause random DNA damage in the *C. elegans* genome and rely on non-homologous end joining (NHEJ) to incorporate the extrachromosomal array when repairs are made. We chose to use the CRISPR/Cas9 system in an effort to avoid positional effects caused by random genomic insertion and to avoid damage to the extrachromosomal array.

Targeting ben-1

In *C. elegans*, *ben-1* encodes a beta-tubulin. Disruption of *ben-1* dominantly suppresses paralysis caused by benzimidizole drugs like benomyl [27]. We decided to target *ben-1* since this would allow us easily identify all heterozygous and homozygous *ben-1* disruptions caused by

CRISPR/Cas9 cutting events based on resistance to 14uM benomyl [28]. Benomyl paralyzes all N2 animals without a *ben-1* disruption. Benomyl resistant animals could then be screened for 100% transmission of the extrachromosomal array by visualizing GFP reporter expression.

Sequence for *ben-1* was obtained from UCSC genome browser [29]. We used CRISPRdirect [30] to search for potential guide sequences in exon-1 and exon-2. We got 17 hits for exon-1 and 41 hits for exon-2. We chose 3 guide sequences for exon-1 and a single sequence for exon-2. We also chose an additional sequence for exon-1 from the literature [31]. See Appendix C sgRNA Table.

sgRNA synthesis and in vitro testing

As mentioned previously, Cas9 is an RNA-guided DNA endonuclease. We obtained purified Cas9 protein as a gift from The Alder Lab (BYU). In order to target our Cas9 to the *ben-1* gene we synthesized sgRNAs using a MEGAshortscript T7^{TM} Kit (ThermoFisher). In brief, PCR was performed using oligos (IDT) and pX330 (gift from Alder Lab) to create a dsDNA template for use with the MEGAshortscript T7^{TM} Kit. This template was then used to synthesize sgRNAs. See Appendix D.

Template digestion by Cas9 complexed with an sgRNA was performed *in vitro* by incubating the sgRNA Cas9 complex in a digestion buffer with a suitable dsDNA template we made using PCR. Not all of our guides had equal cutting efficiencies. sg002, sg003 (data not shown), and sg005 [31] did not produce discreet bands post digestion. However, our GFP control digest and digests with sg004, sg006, and sg007 did produce discreet digestion products. See Figure 2-5. This information is also summarized in Table 2-5. For complete details on testing sgRNA *in vitro* See Appendix D.

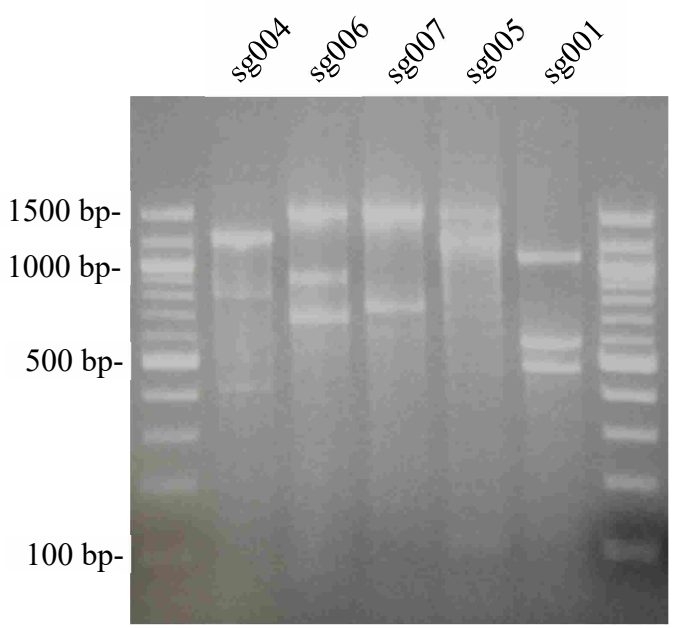


Figure 2-6. Partial *in vitro* digestion of PCR substrates by CRISPR/Cas9.

sg004 yielded fragment sizes of 434 bp and 612bp. sg006 yielded fragment sizes of 700bp and 836 bp. sg007 yielded fragments lengths of 755bp and 781 bp. sg005, a sgRNA guide from the literature, should have shown fragment sizes of 811bp and 725bp. sg001, positive control against GFP, yielded fragment sizes of 498bp and 600bp.

Table 2-5 sgRNA *in vitro* test

sgRNA	Target	Primers	PCR Template	Expected Product (bp)	Digested (bp)	Cut	Gel Image
sg001	GFP	oSMJ001 oSMJ007	L4686	1098	498, 600	yes	Figure 2-5
sg002	*C. elegans* Chr. 2	oSMJ013 oSMJ014	gDNA, AZ212	1111	596, 515	yes	Data not shown
sg003	*ben-1* exon-1	oSMJ019 oSMJ020	N2	2729	825, 1904	yes	Data not shown
sg004	*ben-1* exon-2	oSMJ029 oSMJ026	N2	1046	434, 612	yes	Figure 2-5
sg005	*ben-1* exon-1	oSMJ019 oSMJ028	N2	1536	811, 725	unclear	Figure 2-5
sg006	*ben-1* exon-1	oSMJ019 oSMJ028	N2	1536	836, 700	yes	Figure 2-5
sg007	*ben-1* exon-1	oSMJ019 oSMJ028	N2	1536	755, 781	yes	Figure 2-5

In vivo testing of ben-1 sgRNA

Based on our *in vitro* sgRNA/Cas9 digests, we concluded that it would be best to use sg004, sg006, and sg007 for *in vivo* testing. We chose to test the efficiency of sgRNA/Cas9 complexes in N2 wild-type animals first so we wouldn't need to synchronize large populations of transgenic animals in order to obtain sufficient numbers for microinjection. About 50 young adult N2 animals were taken and microinjected with an injection mix containing sg004, sg006, and sg007 in equal parts in a 1:3 ratio by mass of sgRNA to Cas9 protein. Animals were recovered slowly in recovery buffer and M9. Following this recovery, they were transferred to seeded NGM plates with 14 uM benomyl. Adults were transferred once a day for 5 days and F_1 animals were screened for benomyl resistance. P_0 animals had a high mortality rate ~50%. For complete injection conditions, see Appendix D.

This first *in vivo* test yielded a total of eight unique lines that had the benomyl resistant phenotype. Of those eight, six were chosen for Sanger sequencing. Sequence alignments for the different strains showed that sg006 was the most efficient sgRNA tested *in vivo*. Of the six samples, only Sample 1-1 and 1-6 had mutations not caused by sg006 targeting. Interestingly, only Sample 1-1 was cut more than once, once with sg004 and once with sg006. See Figure 2-2. Based on these results, we chose to use sg006 exclusively for our CRISPR/Cas9 integration experiments. It is possible that the difference in cutting efficiency observed between the different sgRNAs could be due in part to nucleosome positioning. It has been observed by others that nucleosomes can impede cutting by CRISPR/Cas9 *in vitro* and *in vivo* by occluding the requisite DNA sequences and *in vitro* this blocking effect can be overcome by the addition of chromatin remodelers [32].

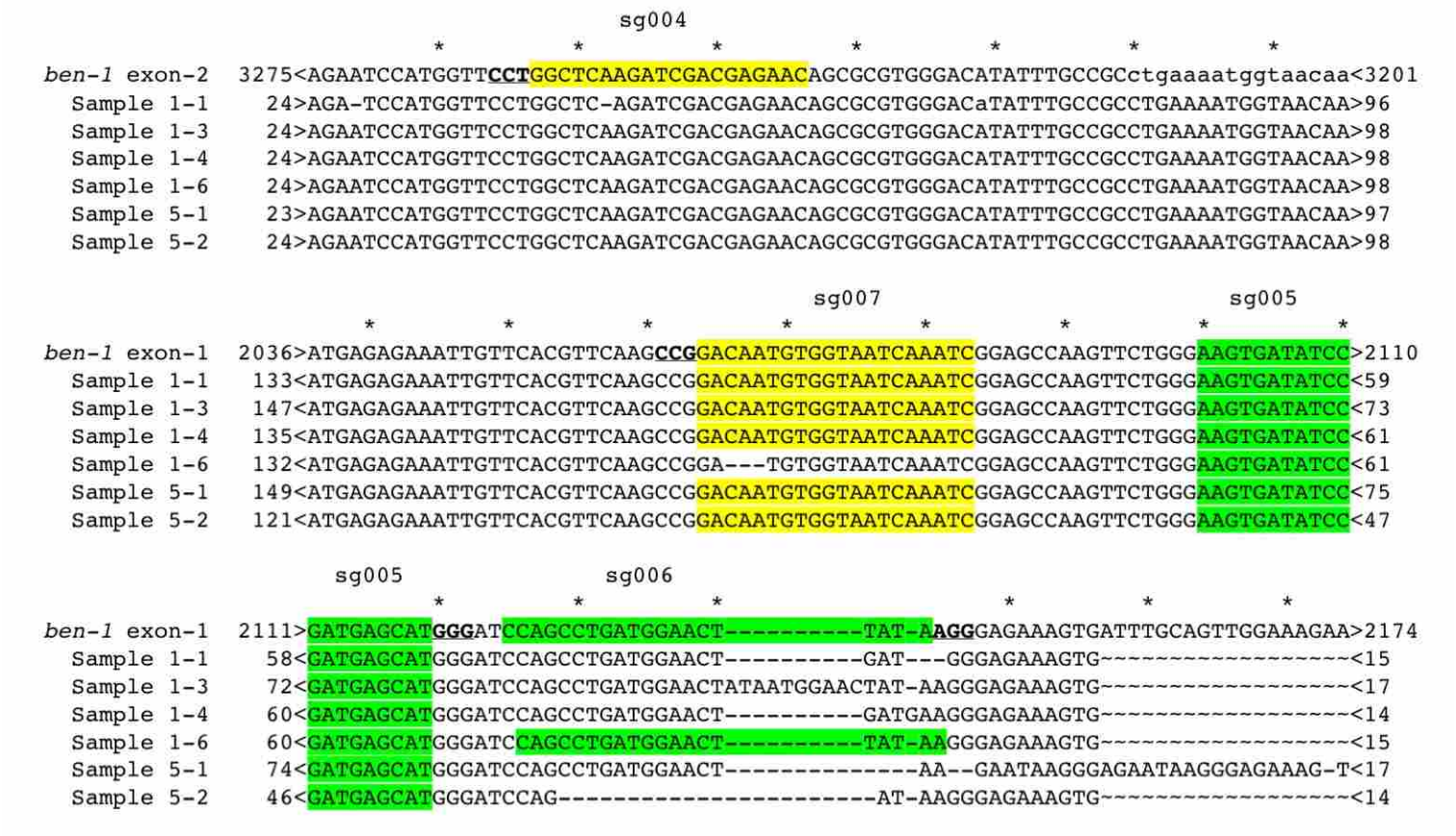


Figure 2-7. Sequence alignments confirm *in vivo* CRISPR/Cas9 cutting events.

Sequence alignments from whole worm PCR and sequencing confirm *in vivo* CRISPR/Cas9 cutting events are present in benomyl-resistant animals. PAM sites are indicated in bold. sgRNA sg004, sg006, and sg007 were used in equal parts in the injection mix.

Array integration via CRISPR/Cas9 microinjection

Transgenic lines were generated by using standard microinjection procedures for *C. elegans* [33]. A strain carrying pBYU1 and pCR4_TOPO_ben-1 was generated by injecting plasmids pBYU1 and pCR4_TOPO_ben-1 in a 10:1 ratio. See Appendix C Injection Mix Table. One line with about a 30-35% transmission rate was selected for integration. About 75 GPF-positive young adults were chosen from a synchronized population and microinjected. F_1 animals were screened for resistance to benomyl. 24 benomyl resistant lines were recovered from these injections and 8 were positive for GFP expression. However, these GFP positive lines did not have an integrated array since the array transmission remained at about 30-35%. Even though we did not obtain an integrated array from these injections, it is likely that more injections, possibly with a higher transmitting array will yield *ben-1* integrated arrays in the future. Currently, purified CRISPR/Cas9 has not been used successfully to integrate an extrachromosomal array in *C. elegans*. Presently, researchers use CRISPR/Cas9 that has been cloned into embryonic expression vectors.

Injecting purified Cas9 with its associated sgRNA has some advantages over encoded CRISPR/Cas9 under the control of an embryonic promoter. Plasmids encoding CRISPR/Cas9 can form into heritable extrachromosomal arrays that remain active and serve as a source of continual CRISPR/Cas9 activity in successive generations. This is problematic if there is any off-target activity or if Cas9 has been targeted to any part of the array that is being integrated since Cas9 will continue to cut after integration. Injecting purified ribonucleoprotein avoids this potential problem. There is also greater flexibility when using multiple sgRNAs since this method does not require sgRNA template to be cloned into a plasmid vector.

Microinjection of plasmid encoded CRISPR/Cas9 also has some advantages over injecting the purified ribonucleoprotein. Worm recovery post microinjection is much simpler since worms do not need a long time to recover. Injecting purified ribonucleoprotein puts the worms under a lot of osmotic stress compared to injecting plasmid DNA since high levels of salt need to be maintained to keep the protein soluble. Also, microinjection mixes using supercoiled plasmid DNA at a maximum concentration of 100 ng/uL form heritable arrays at a frequency of about 10% of all transgenic F_1 animals [14]. This means most of the transgenic animals expressing Cas9 should not form a heritable array. Also, co-injection markers like rol-6 or GFP variants can be used to identify F_1 animals with CRISPR/Cas9 arrays and then screened in the F_2 generation for its loss. This method is what is currently being used to integrate extrachromosomal arrays with CRISPR/Cas9 [34].

It is not known to what extent plasmids microinjected into an animal already bearing an extrachromosomal array recombine or become incorporated into the pre-existing array. However, since homologous recombination in *C. elegans* is very limited and spontaneous array integration is also quite rare [33], it seems unlikely that the incoming plasmids will combine with the pre-existing array to any significant degree. This could be very helpful for individuals who want to use the CRISPR/Cas9 system to integrate extrachromosomal arrays but lack the expertise to work with the purified ribonucleoprotein.

Gamma-ray integration

In parallel with the CRISPR/Cas9 integration experiments, we also tried other methods for array integration. While our efforts with UV did not produce any integrated lines, we were successful using gamma rays. A dose of ~3800 Rads produced 17 integrants isolated from 600 F_1 animals

for an integration rate of 2.8%. See Appendix D for complete details. These lines are currently being backcrossed to clean up any background mutations before performing our forward genetic screen.

Chapter 3. Reproducibility and consistency of *in vitro* nucleosome reconstitutions demonstrated by invitrosome isolation and sequencing

This chapter is an adaptation from the article "Reproducibility and consistency of *in vitro* nucleosome reconstitutions demonstrated by invitrosome isolation and sequencing" published in *PlosOne* August 2014.

3.1 ABSTRACT

Nucleosomes and their positions in the eukaryotic genome play an important role in regulating gene expression by influencing accessibility to DNA. Many factors influence a nucleosome's final position in the chromatin landscape including the underlying genomic sequence. One of the primary reasons for performing *in vitro* nucleosome reconstitution experiments is to identify how the underlying DNA sequence will influence a nucleosome's position in the absence of other compounding cellular factors. However, concerns have been raised about the reproducibility of data generated from these kinds of experiments. Here we present data for *in vitro* nucleosome reconstitution experiments performed on linear plasmid DNA that demonstrate that, when coverage is deep enough, these reconstitution experiments are exquisitely reproducible and highly consistent. Our data also suggests that a coverage depth of 35X be maintained for maximal confidence when assaying nucleosome positions, but lower coverage levels may be generally sufficient. These coverage depth recommendations are sufficient in the experimental system and conditions used in this study, but may vary depending on the exact parameters used in other systems.

3.2 INTRODUCTION

Nucleosomes play an important role in gene regulation. Eukaryotic genomes are highly compacted and nucleosomes are the most basic of the many levels of compaction. Nucleosomes are formed when 147 base pairs (bp) of DNA wrap about 1.7 times around a histone octamer [1]. Gene regulation in eukaryotes frequently begins at the transcriptional level with trans-acting factors binding the DNA. A nucleosome's genomic position and which DNA bases are facing towards or away from the nucleosome, often described as translational and rotational setting respectively, can affect many important processes by influencing the availability and function of binding sites encoded in the DNA [35]. Therefore, nucleosomes and their positions on the DNA are the first level of eukaryotic gene regulation.

What influences and ultimately determines a nucleosome's position within the genome is complex with many groups actively researching this question. Some of the factors that influence a nucleosome's position include the underlying DNA sequence, chromatin remodeling factors, DNA binding proteins, transcription factors, and even neighboring nucleosomes [36]. Many experiments have been done *in vitro* and *in vivo* to examine how these factors affect positions of nucleosomes and gene regulation *e.g.*[10, 13, 37-40].

In vitro nucleosome reconstitution is done by mixing naked DNA fragments and isolated or recombinant histone octamers together in a high-salt environment. While the salt is slowly dialyzed out of this solution, spontaneous interactions between the DNA and histone octamers result in the formation of nucleosomes on DNA sequences that are most thermodynamically favorable [41]. One purpose of these experiments is to observe and define the influence that underlying DNA sequence has on nucleosome formation. This is done by observing the

positioning and occupancy of reconstituted nucleosomes on defined DNA sequences or even whole genomes. Several groups, including ours, have used this method to demonstrate that nucleosome occupancy and positioning is highly dependent, at least *in vitro*, on the nature of the underlying DNA sequence [38-40]. The proclivity of a nucleosome to form and the precise positioning of a nucleosome on a DNA sequence can be two separate, yet often conflated functions of the underlying DNA sequence. In reality, these separate functions can be directed by individual elements within the DNA forming the nucleosome core and linker regions [38].

While several sequences have been shown to be highly consistent in their ability to attract and precisely position nucleosomes (e.g. 601, sea urchin 5S , container site)[10, 20, 38, 42], some researchers continue to doubt the reproducibility of *in vitro* nucleosome reconstitution across less well defined sequences and even the reproducibility of positioning across experiments. These criticisms stem from the fact that multiple nucleosome positions can be adopted on DNA fragments that are greater than 147 bp or even on a DNA fragment of only 147 bp [41, 43], which is then often interpreted to mean that *in vitro* reconstitution experiments are inconsistent in their outcomes and hence irreproducible and unreliable. Here we present evidence that *in vitro* nucleosome reconstitution experiments on plasmid DNA assayed by micrococcal nuclease (MNase) digestion and high-throughput sequencing (MNase-seq) are both reproducible and highly consistent allowing confident analysis of both nucleosome positioning and occupancy using this technique.

3.3 RESULTS AND DISCUSSION

In vitro nucleosome reconstitutions on linearized plasmid DNA and invitrosome DNA sequencing

In order to address the question of consistent positioning and occupancy in *in vitro* nucleosome (invitrosome) [12] reconstitution experiments, we analyzed the reproducibility of positioning and coverage results between multiple independent invitrosome experiments. In all experiments, invitrosomes were formed by salt dialysis using recombinant histone octamer on linearized plasmid DNA [41]. Four different linearized plasmids with identical backbones, but each harboring a different, unique ~150bp sequence (see Materials and Methods) at the same insert site (the kat-group plasmids, p4.1, p4.2, p4.3 and p4.4), were used as the DNA template in separate invitrosome experiments (Figure 3-S1 1, kat-group backbone).

To allow invitrosomes to form on the DNA templates in positions influenced only by underlying DNA preferences and to eliminate the effects of steric hindrance or positioning by neighboring nucleosomes [22, 38, 44-46], we used a reconstitution ratio of one histone octamer per 1000 bp of plasmid DNA. After *in vitro* formation, mononucleosomes were isolated by MNase digestion; and DNA from these invitrosome cores, representing their positions on the plasmid DNAs, were isolated as previously described [38, 47]. Invitrosome cores were ligated with barcoded adaptors and sequenced (Tables 1 & 2). A total of 860,741 invitrosome core DNAs were sequenced for these four plasmid reconstitutions representing ~6,600 to ~8,200 fold coverage for each experiment.

We parsed our sequence reads into individual experiments according to the embedded barcodes and mapped the reads back to their respective reference plasmids. In order to avoid any end bias [43, 48, 49] or influence from the different ~150 bp inserts in our plasmids, we filtered

our reads such that reads mapping to within 147 bp of either end of the linearized reference plasmid, as well as reads mapping to within 147 bp upstream or downstream of the insertion site or to the insertion site itself, were excluded. Any reads that overlapped these filtered areas or would overlap them when extended to 147 bp were excluded from further analysis. The resulting filtered read sets were used to create coverage plots representing nucleosome occupancy and positioning on the plasmids (Tables 1 & 2).

Table 3-1 Plasmids, Backbones, Primers and Barcodes

Invitrosome	Source plasmid	Linker pair	Barcode
p4.1	pCR4Blunt-TOPO	AF-SJ-84/ AF-SJ-99	CAGT
p4.2	pCR4Blunt-TOPO	AF-SJ-85/ AF-SJ-100	GTCT
p4.3	pCR4Blunt-TOPO	AF-SJ-86/ AF-SJ-101	TGCT
p4.4	pCR4Blunt-TOPO	AF-SJ-87/ AF-SJ-102	CCCT
p7.1	pPD149.40	AF-SJ-88/ AF-SJ-103	AACT
p7.2	pPD149.40	AF-SJ-89/ AF-SJ-104	GCAT
p7.3	pPD149.40	AF-SJ-90/ AF-SJ-105	CGAT
p7.4	pPD149.40	AF-SJ-91/ AF-SJ-106	TAAT
p7.5	pPD149.40	AF-SJ-92/ AF-SJ-107	ATAT
p7.6	pPD149.40	AF-SJ-93/ AF-SJ-108	TCTT
p7.7	pPD149.40	AF-SJ-94/ AF-SJ-109	GATT

Table 3-2 Reads for Coverage Plots

Invitrosome	Raw	Mapped	Filtered
p4.1	230867 (100%)	224453 (97.2%)	217405 (94.2%)
p4.2	230672 (100%)	229591 (99.5%)	225536 (97.8%)
p4.3	185104 (100%)	182461 (98.6%)	179213 (96.8%)
p4.4	214098 (100%)	207059 (96.7%)	204239 (95.4%)
kat-group total	860741 (100%)	843564 (98.0%)	826393 (96.0%)
p7.1	192686 (100%)	191075 (99.2%)	179666 (93.2%)
p7.2	148552 (100%)	142340 (95.8%)	134453 (90.5%)
p7.3	68977 (100%)	65821 (95.4%)	61940 (89.8%)
p7.4	113936 (100%)	109367 (96.0%)	95264 (83.6%)
p7.5	49407 (100%)	48081 (97.3%)	41552 (84.1%)
p7.6	330983 (100%)	321480 (97.1%)	188176 (56.9%)
p7.6_601	330983 (100%)	321480 (97.1%)	321480 (97.1%)
p7.7	353261 (100%)	344100 (97.4%)	330525 (93.6%)
sèt-group total*	1257802 (100%)	1222264 (97.2%)	1031576 (82.0%)
ALL	2118543 (100%)	2065828 (97.5%)	1857969 (87.7%)

*excluding p7.6_601

Consistency and reproducibility between invitrosome experiments

For each of our invitrosome experiments, we generated coverage plots by extending all mapped reads to a total length of 147 bp from the read start site. After this extension, the number of invitrosomes that occupied each site on the plasmid was calculated. Histograms of nucleosome occupancy at each site and for each plasmid, looking at both forward-mapping reads and reverse-mapping reads independently, were generated (Figure 3-1). Visual inspection and comparison of the forward-read coverage plots to the reverse-read coverage plots and between the plots of all four independent experiments showed striking near identity in their coverage and positioning patterns. For better visual comparison of these plots, we normalized the data between the four experiments by making combined forward- and reverse-read coverage plots for each experiment and then scaling the plots to the lowest coverage plot (by read count) among the four (see Materials and Methods). This allowed visual discrimination and direct comparison of all four experiments on a single plot (Figure 3-1F), further confirming the striking near identity of the results of all the experiments.

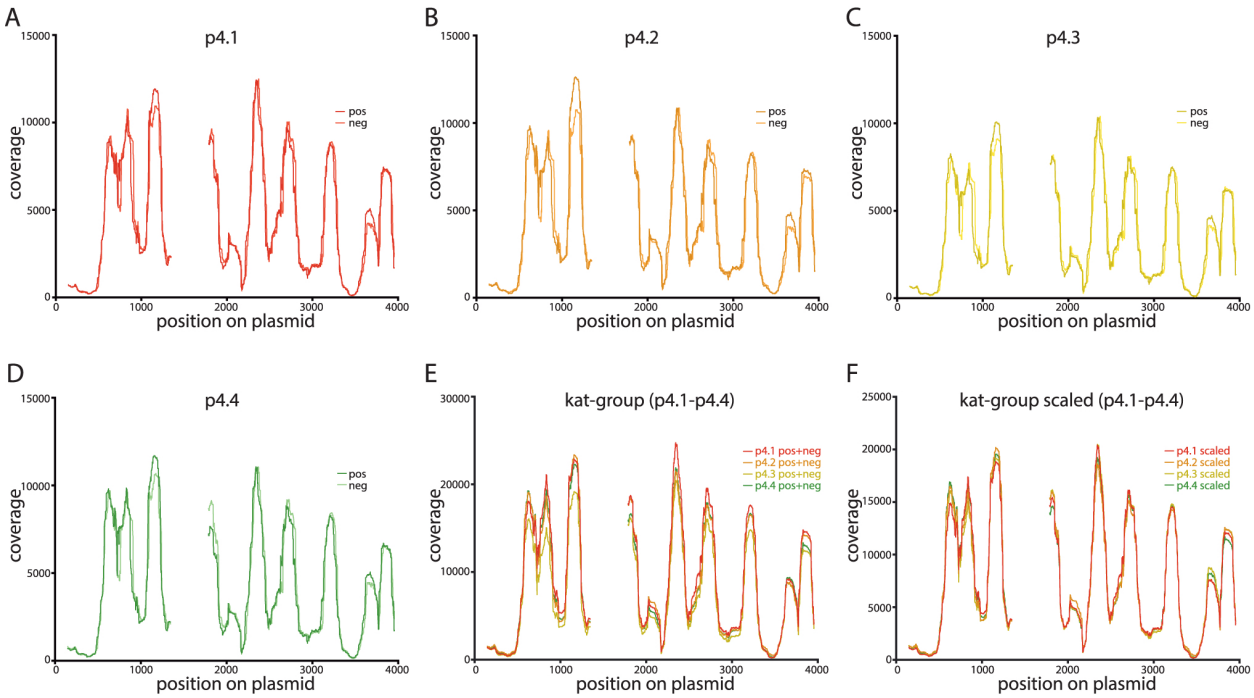


Figure 3-1. Invitrosome coverage plots for kat-group plasmids.

Invitrosome coverage plots for kat-group plasmids show near identical coverage patterns. Invitrosome coverage on each plasmid is plotted on the y-axis and the plasmid coordinates in bp are plotted on the x-axis. Gaps in the plot represent the trimmed insert site and end regions on each plasmid.1A-1D Non-scaled coverage plots with forward (pos) and reverse (neg) invitrosomes plotted separately for plasmids p4.1-p4.4. 1E Non-scaled coverage plots with forward and reverse invitrosomes (pos+neg) combined for each plasmid. All four kat-group plasmids are shown together. 1F Combined and scaled coverage plots for all four kat-group plasmids where each plot is normalized to the plasmid with the least coverage for direct comparison.

Invitrosome reproducibility on other DNAs

One concern is that the high reproducibility of the invitrosome analyses that we observed above is actually an effect unique to the plasmid backbone we used in our experiments. In order to address this question we performed *in vitro* nucleosome reconstitution experiments on a set of seven new plasmids, again all with the same plasmid backbone, but different from the backbone used in our previous experiments (the sèt-group plasmids: p7.1, p7.2, p7.3, p7.4, p7.5, p7.6 and p7.7). Like the previous set of plasmids, each of these seven plasmids harbored a different ~150

bp sequence at a unique site within the plasmid backbone (Figure 3-S1, sèt-group backbone). These invitrosome experiments were performed with reconstitution, digestion, sequencing and analysis identical to the experiments described above. Like our first set of experiments, this second set of experiments showed extremely high reproducibility between the seven plasmids both with forward- versus reverse-read coverage plots and with combined coverage plots of the seven plasmids (Figure 3-2 and Figure 3-S2). One notable feature of these data is that, unlike our previous set of experiments, the sequencing read coverage on these seven plasmids varied considerably (Figure 2B). Despite this variation, the visual patterns in the plots were entirely consistent, and after normalization by scaling, showed near identity (Figure 3-2C). After all coverage plots were scaled to the coverage plot with the least amount of coverage (Figures 3-1F and 3-2C), the effective coverages for the kat-group and sèt-group plasmids were 6660X and 1663X respectively.

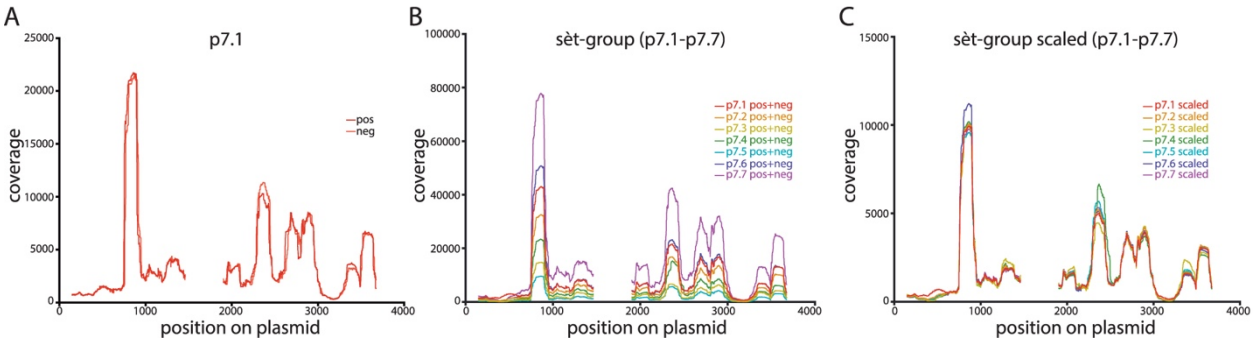


Figure 3-2. Invitrosome coverage plots for an alternative plasmid backbone

Invitrosome coverage plots for an alternative plasmid backbone (the sèt-group plasmids) also show near identical coverage patterns.2A Example coverage plot for one of the sèt-group plasmids (p7.1) with both forward (pos) and reverse (neg) invitrosomes plotted separately on the same graph. 2B Non-scaled coverage plots with forward and reverse invitrosomes (pos+neg) combined for each of the seven sèt-group plasmids plotted on the same graph. 2C Combined and scaled coverage plots for all seven sèt -group plasmids, where for direct comparison, each plot is normalized to the plasmid with the least coverage and plotted on the same graph.

Notwithstanding this qualitative visual conformation of consistency and reproducibility between invitrosome experiments, we needed a metric to quantify the similarity between our experiments. We chose to use Pearson's correlation coefficients for our metric.

Quantitative analysis of invitrosome experiment reproducibility

We calculated the Pearson's correlation between each pair of coverage plots within each group and made a Pearson's correlation coefficient matrix for the different coverage plots. We were pleased to see extremely high correlations ranging from a low of 0.974 to a near perfect correlation of 0.999 (Figure 3-3). Thus our qualitative visual and quantitative computational analyses demonstrated the extreme level of reproducibility in our invitrosome experiments. Additionally, visual inspection of the coverage plots, especially in the plots of the sèt-group plasmids, shows not only consistent relative occupancy, but also consistent positioning in individual sites with some very well positioned nucleosomes (Figure 3-1F and 3-2C). However, given the extremely high coverage levels of all our individual invitrosome experiments (non-scaled plots ranging from 1,663X to 13,225X), we wanted to know if this observed consistency and reproducibility is only possible between experiments with enormously high coverage like the ones we have here.

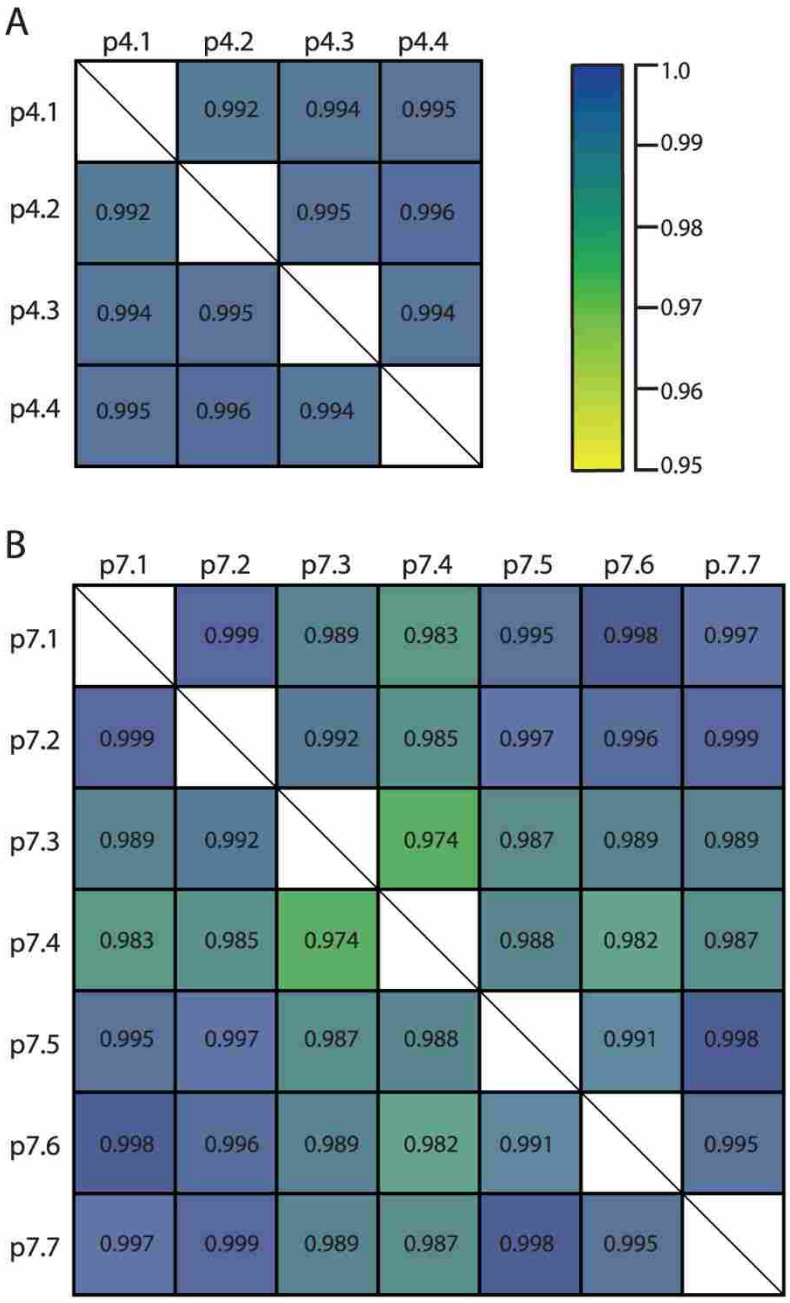


Figure 3-3. Correlation matrices for scaled coverage plots.

Correlation matrices displaying Pearson's correlation values for scaled coverage plots quantitatively demonstrate high correlations between invitrosome experiments. 3A Pearson's correlation matrix for coverage plots from plasmids p4.1-p4.4. 3B Pearson's correlation matrix for coverage plots from plasmids p7.1-p7.7. In both matrices, high to lower correlation values range in color from blue to yellow as show.

Effect of variable coverage density

Our analysis showed that our method for *in vitro* reconstitutions, mononucleosome core DNA isolation, and sequencing yielded very consistent results across samples and experiments when coverage is high. To determine the minimum level of coverage required to achieve similar or minimally acceptable Pearson's correlation coefficients, we randomly extracted different amounts of filtered reads for each plasmid corresponding to the following levels of coverage: 1X, 2.5X, 5X, 10X, 15X, 25X, 35X, 50X, 100X and 500X. We initially chose two plasmids of each backbone type for this analysis: p4.1 and p4.2 for the kat-group plasmids, and p7.1 and p7.2 for the sèt-group plasmids. We performed three replicate read extractions for each of the chosen plasmids at each of the ten coverage levels examined. We generated coverage plots as described above for each replicate at each coverage level and calculated Pearson's correlation coefficient values (S8-Supplemental data tables). Figure 3-4 is an example of this analysis for one pair of plasmids (p4.1 and p4.2) at one (35X) coverage level.

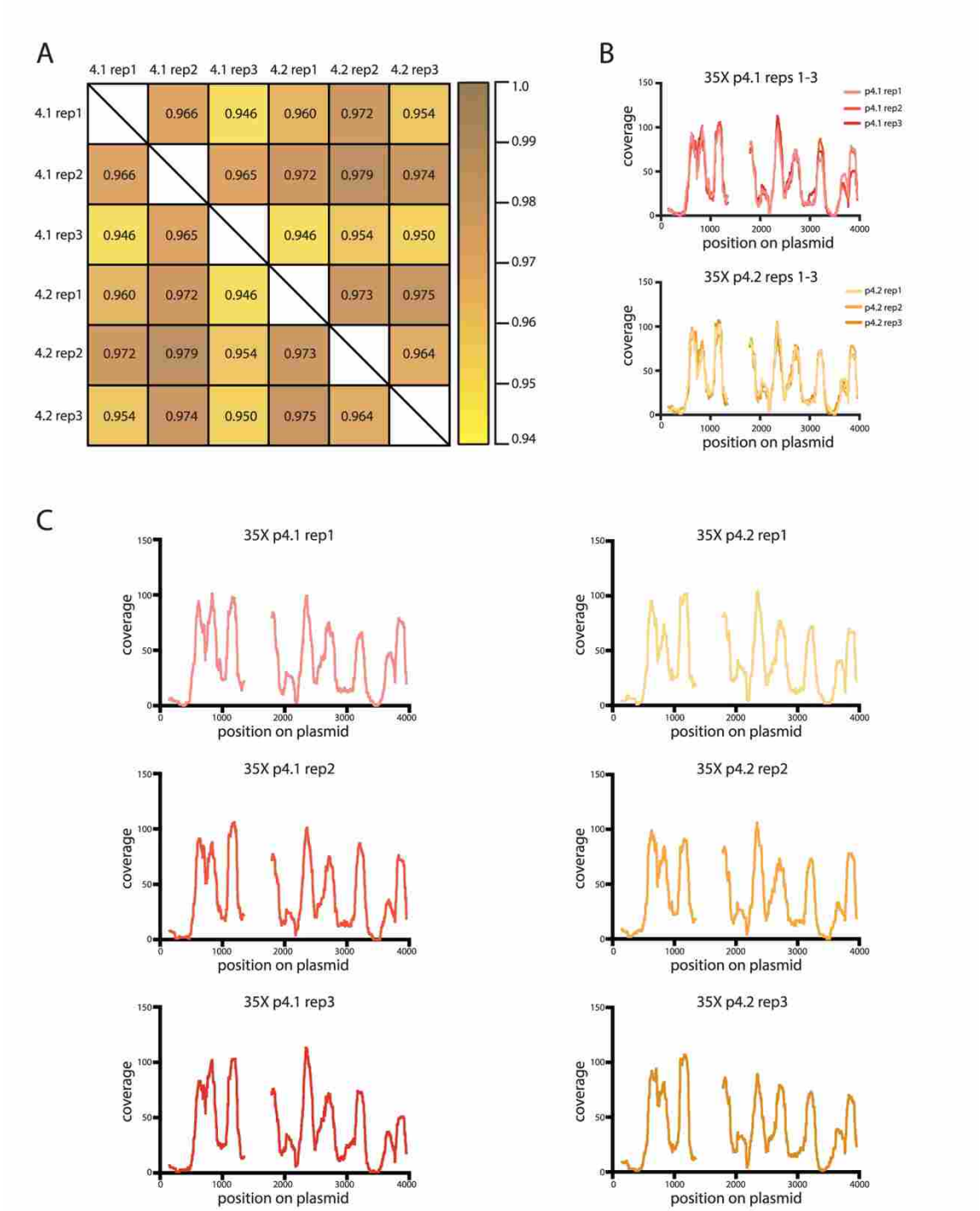


Figure 3-4. High correlations at experimentally relevant levels.

High correlations are maintained at experimentally relevant invitrosome coverage levels. 4A Pearson's correlation matrix for the coverage plots generated from three replicates at 35X coverage of plasmids p4.1 and p4.2. 4B (top) Combined coverage plots for replicates 1-3 of plasmid p4.1 at 35X and (bottom) combined coverage plots for replicates 1-3 of plasmid p4.2 at 35X. 4C All six replicates of the 35X coverage plots. (Left) coverage plots of replicates 1-3 of plasmid p4.1 at 35X. (Right) coverage plots of replicates 1-3 of plasmid p4.2 at 35X.

To visually analyze the range of Pearson's correlation coefficient values at each coverage level we plotted the Pearson's correlation coefficients as whisker plots. As expected, Pearson's correlations between plasmid replicates were inconsistent at low coverage levels and became better with increasing coverage (Figure 3-5A). As can be seen in Figure 3-5B, once coverage reached 35X, the plasmid backbone-specific pattern observed in the full, normalized-coverage experiments became apparent (Compare Figure 3-1F and Figure 3-5B last panel).

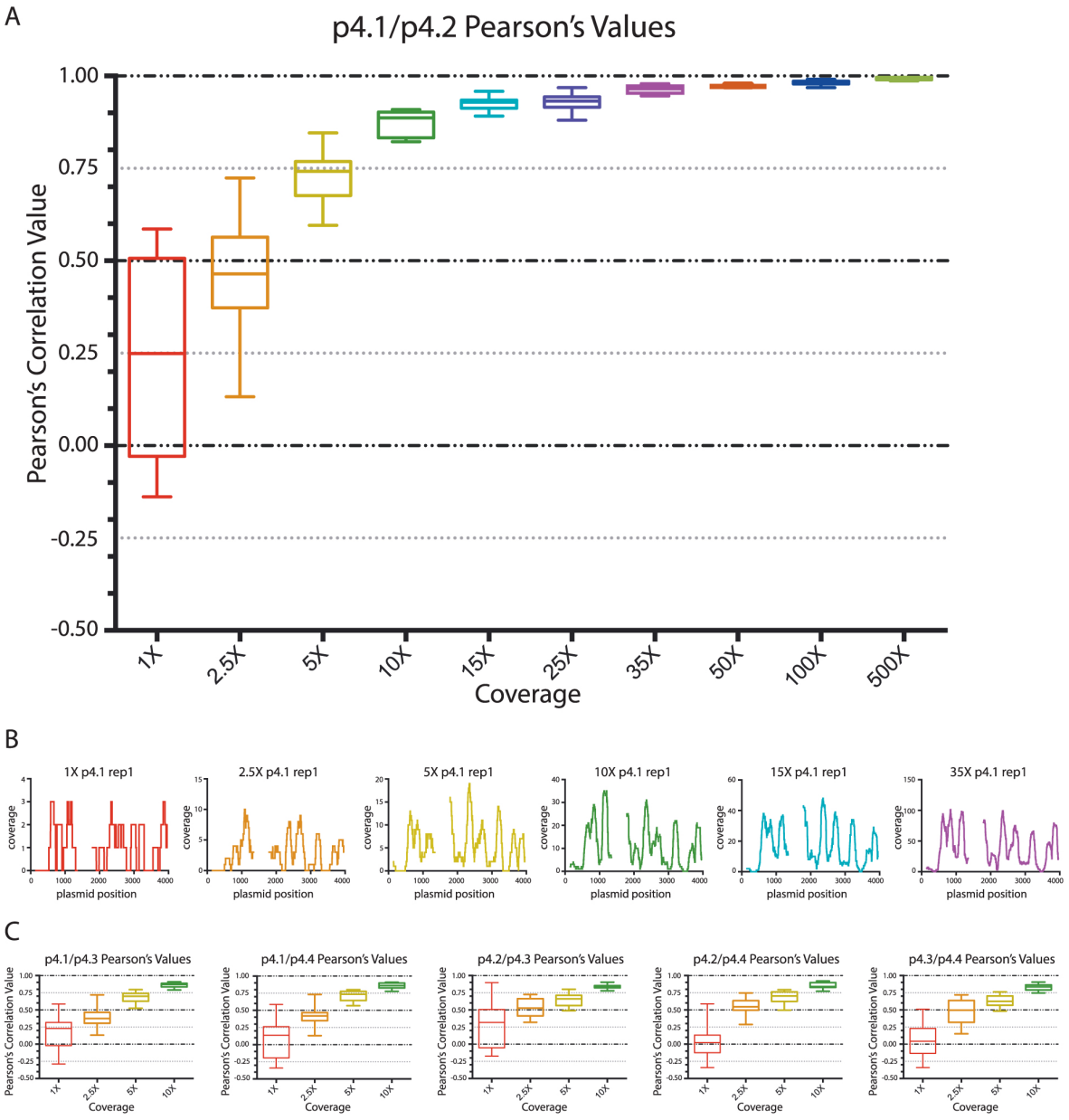


Figure 3-5. Comparison of correlation values show minimal acceptable coverage levels

Comparison of correlation values show minimal acceptable coverage levels for invitrosome experiments. 5A Range of correlation values at each coverage level for the pair-wise comparisons of the three replicates for plasmid p4.1 and the three replicates for plasmid p.4.2. The Pearson's correlation values are on the y-axis and the coverage levels are on the x-axis. The range of correlation values for each coverage level is plotted as a whisker plot composed of the 30 pair-wise comparisons at each coverage level. 5B Coverage plots for replicate 1 of plasmid p4.1 at coverage levels: 1X, 2.5X, 5X, 10X, 15X and 35X. The colors of the coverage plots correspond to the whisker-plot colors of the same coverage levels in 5A. 5C Whisker plot graphs showing the variability of correlation values at low levels of coverage (1X-10X) for all combinations of the kat-group plasmids. In 5A-5C the colors red, orange, yellow, green, indigo and purple represent data from coverage levels of 1X, 2.5X, 5X, 10X, 15X and 35X, respectively.

Due to the wide range of correlation coefficients observed at the lower coverage levels from replicates of plasmids p4.1 and p4.2, we performed the same analysis on the rest of the kat-group plasmids at the lower coverage levels (1X-10X) and compared them to one another (Figure 3-5C). We also did this low-coverage-level pair-wise analysis for all of the sèt-group plasmids (Supplemental Figures 3-S3-S7). As can be seen in Figure 3-5C, regardless of the plasmid pair, for the kat-group plasmids, correlations at very low coverage levels (1X and 2.5X) are quite variable and extremely low to nonexistent. But surprisingly, at even moderate levels of coverage (5X and 10X), correlations become modestly good (above 0.5 and 0.75 respectively). Interestingly, the Pearson's correlation coefficient values for the sèt-group plasmids at lower coverage levels (1X-10X) are strikingly higher than those of the kat-group plasmids (Figure 3-6 and Supplemental Figures 3-S4-S7). The possible cause of this will be discussed below.

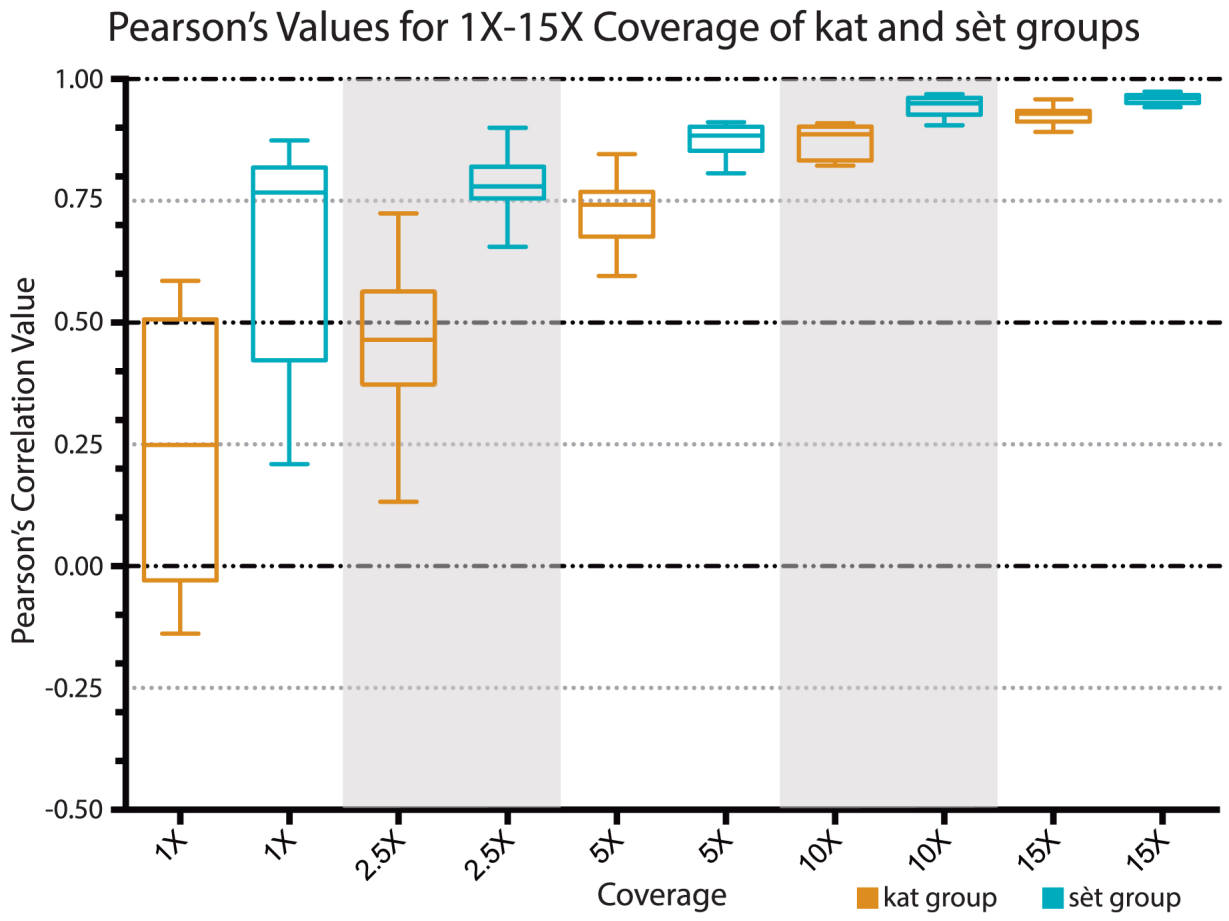


Figure 3-6. Degree of correlation between sèt-group plasmids versus kat-group plasmids.

Higher degree of correlation between low-level coverage of sèt-group plasmids compared to low-level coverage of kat-group plasmids. Range of correlation values for coverage levels 1X, 2.5X, 5X, 10X and 15X are displayed as a whisker plot. The pair-wise comparisons of the three replicates for plasmid p4.1 and the three replicates for plasmid p.4.2 from the kat-group plasmids are displayed as orange whisker-boxes and the pair-wise comparisons of the three replicates for plasmid p7.1 and the three replicates for plasmid p.7.2 from the sèt-group plasmids as blue whisker-boxes. The Pearson's correlation values are on the y-axis and the coverage levels are on the x-axis.

3.4 CONCLUSIONS

Here we have shown that *in vitro* nucleosome reconstitution experiments are reproducible and highly consistent when read coverage is sufficiently deep. We have quantified correlation coefficients between reconstitutions using Pearson's correlations, and as expected, Pearson's correlation values steadily rise and have less variation as coverage increases, with values

reaching as high as 0.999 between experiments. The most dramatic gains in increasing Pearson's correlation values with minimal variation are made once 35X coverage is reached where values are 0.946 and above.

It is important to note that in these experiments we have used nothing but the raw invitrosome reads and their coverage plots to calculate correlation values. In most studies of nucleosome positioning using high-throughput sequencing, large numbers of raw nucleosome reads are mapped and then converted to coverage plots that are used to call individually positioned nucleosome dyads using various probability statistics or smoothing algorithms which greatly decrease the variability within the data *e.g.* [37, 50]. The resulting data are then used to compare results between experiments resulting in higher correlations than if the raw data were used without such manipulation. Here, by using only the raw data we have not artificially increased our correlation values, but shown that such manipulations are not necessary to achieve even near-perfect reproducibility between invitrosome experiments using our experimental conditions.

Additionally, we found that the correlations between low coverage experiments are quite variable and at least somewhat dependent on the plasmid backbone. Specifically, the Pearson's correlation values for the kat-group plasmids at low coverage (1-10X) were strikingly lower than the values for the sèt-group plasmids (Figure 3-6). We believe that this is due to intrinsic differences in the backbones. The coverage plots for the kat-group plasmids have several more peaks representing positioned nucleosomes than those for the sèt-group plasmids (Figures 3-1F and 2C), indicating that the kat-group backbone has more places where nucleosomes are likely to form and *in toto* result in a more uniform occupancy across the entire plasmid backbone; whereas a plurality if not the majority of nucleosomes in the sèt-group plasmid backbones

occupy two or three specific sites with one of these sites being very highly occupied. We believe this to be the cause of the higher correlation values at low coverage levels for the sèt-group plasmids; with more nucleosomes in fewer sites there is less possible variation even at low levels of coverage. This is most likely due to the inherent higher affinity of these few sites. Thus when a nucleosome forms on a sèt-group plasmid backbone, it is likely to form in one of a few specific sites rather than one of the many possible sites on a kat-group plasmid backbone. Many nucleosomes in a few sites give a better correlation than the same number of nucleosomes spread over many sites.

This hypothesis can be easily tested by embedding a known strong nucleosome-positioning sequence into one of our plasmid backbones and verifying that adding such a highly attractive nucleosome-positioning sequence reduces the coverage depth required to obtain good correlation values. The unique ~150 bp insert sequence in the sèt-group plasmid p7.6 is actually the 601 nucleosome positioning sequence. The 601 sequence is the highest affinity DNA sequence known that causes occupancy and positioning of nucleosomes in *in vitro* nucleosome reconstitutions [11]. To test if the addition of the 601 sequence would result in better correlation values at lower coverage levels as hypothesized, we performed six replicate random read extractions for both the p7.6 plasmid and the p7.6 plasmid with the 601 sequence (p7.6_601). In the case of p7.6_601 we now included in the results the reads that mapped to the 601 insert site and its flanking regions that had previously been excluded from our analyses. These replicate read extractions were done for both plasmid data sets at 1X, 2.5X, 5X, 10X and 15X coverage levels. We generated coverage plots as described above for each replicate at each coverage level and calculated Pearson's correlation coefficient values between the six replicates of p7.6 and separately between the six replicates of p7.6_601 (S8-Supplemental data tables). Full-read

coverage plots for both p7.6 and p7.6_601 demonstrate that the 601 sequence is indeed highly attractive to nucleosome formation (Figure 3-7) and, that with the addition of the 601 sequence, a plurality if not a majority of reads now map to the 601 site (Figure 3-7B). As seen by whisker plots of the ranges of correlation coefficient values at each coverage level, Pearson's correlations between the p7.6_601 plasmid replicates were much higher than the Pearson's correlations between the p7.6 plasmid replicates at all coverage levels (Figure 3-8). As can be seen in Figure 3-8, surprisingly good correlations are achieved at even the 1X coverage depth in the p7.6_601 plasmid, supporting our hypothesis.

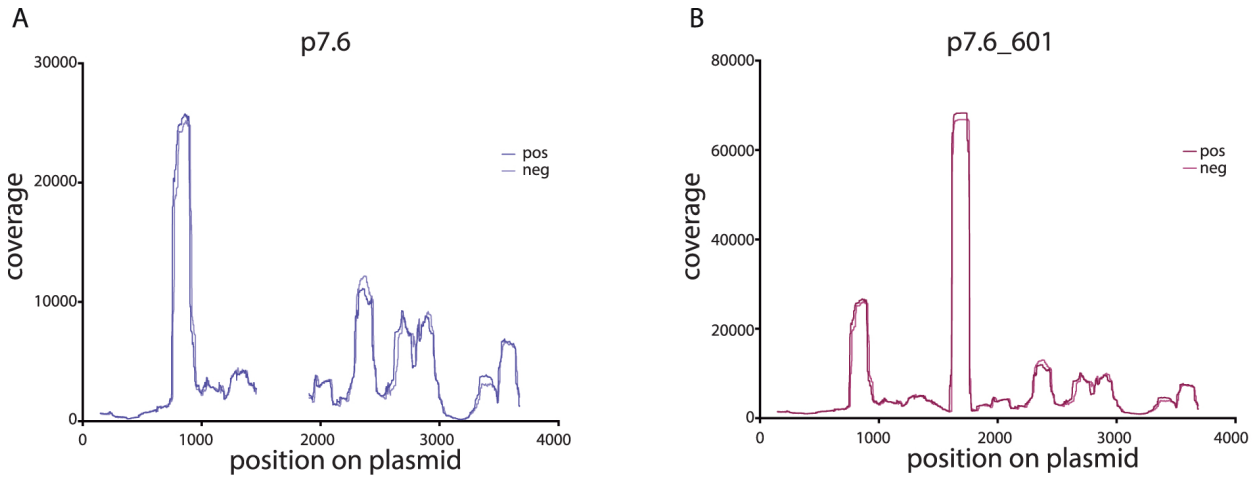


Figure 3-7. The 601 nucleosome-positioning sequence positions a plurality of invitrosomes.

7A Coverage plot for the sèt-group plasmid p7.6 with both forward (pos) and reverse (neg) invitrosomes plotted separately on the same graph. 7B Coverage plot for plasmid p7.6_601 which harbors the 601 nucleosome-positioning sequence with both forward (pos) and reverse (neg) invitrosomes plotted separately on the same graph. The highest peak starting at about base pair 1600 is where the 601 sequence begins.

The above explanation is illustrative of why the general criticisms of the reproducibility of *in vitro* nucleosome reconstitution experiments are not valid. Having multiple possible nucleosome formation sites on a given piece of DNA, and seeing these multiple outputs in invitrosome experiments, does not show that these experiments are inconsistent, but that they just have not been performed and analyzed at sufficient depth. As demonstrated by our

extremely-high-coverage experiments, and even our moderate-coverage experiments, the preference of invitrosome formation in such experiments is very consistent, and on defined stretches of DNA, is even limited to an easily quantifiable number of possible positions.

Thus we propose that studies looking at nucleosome positioning using *in vitro* reconstitution should ideally try to achieve a 35X coverage of the target genome or locus in order to have maximum confidence in the results, but also recognize that significant correlations are seen at levels as low as 5X coverage and should be used as an absolute minimum. In extreme cases (*i.e.*, p7.6_601), even 1X coverage results in a satisfactory correlation (Figure 3-8). Given the current levels of output using next-generation sequencing technologies these target coverages are easily achievable and quite reasonable.

In these experiments we have used MNase digestion of invitrosomes and high throughput sequencing of mononucleosome DNA fragment ends as our output to define individual nucleosome positions and overall nucleosome coverage. This analysis relies heavily on the patterns that are revealed by MNase digestion and assumes, as have many previous studies [38-40], that such digestions along with their known and unknown biases are representative of *in vitro* nucleosome positions. It is possible that the striking consistency between our experiments is a result of our particular technique and conditions, but this in no way detracts from our conclusions about the reproducibility of such experiments. It should be emphasized that, as exemplified by the differences in correlation values at various coverage depths between the kat-group plasmids, the sèt-group plasmids and p7.6_601 (Figures 3-6 and 3-8), the coverage depth necessary to achieve acceptable correlation values is at least dependent on the nature of the DNA sequence, and probably also dependent on the specific system used, and will vary with other factors such as octamer to DNA ratio and reconstitution method and conditions. Further

analyses using other probes and conditions to reveal *in vitro* nucleosome positions and DNA preferences, and analysis of the consistency between these techniques will be an exciting avenue for future exploration.

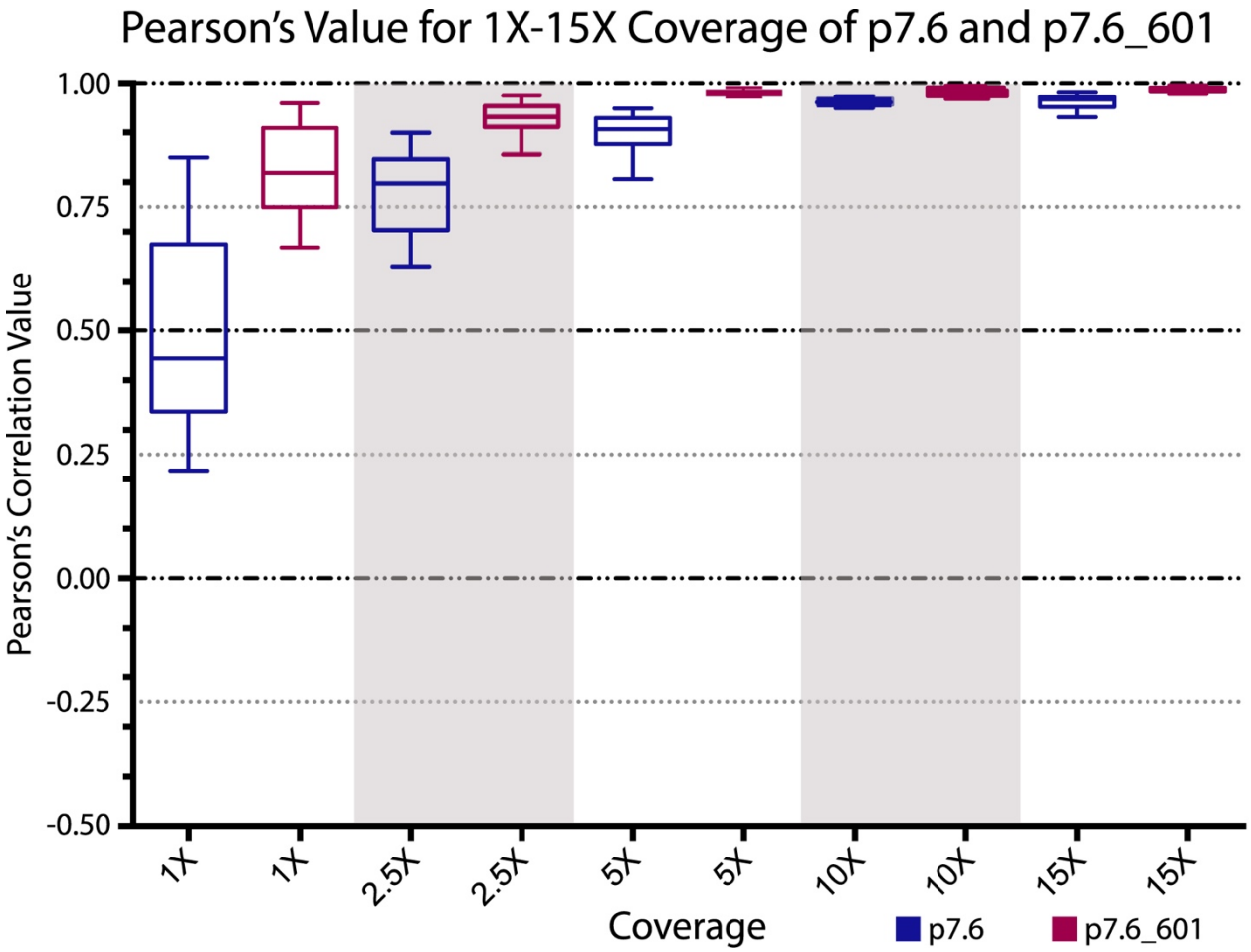


Figure 3-8. Degree of correlation between p7.6_601 versus p7.6.

Higher degree of correlation between coverage of p7.6_601 compared to coverage of p7.6. Range of correlation values for coverage levels 1X, 2.5X, 5X, 10X and 15X are displayed as a whisker plot. The correlation values from the pair-wise comparisons of the six replicates for plasmid p7.6 and the six replicates for plasmid p.7.6_601 are displayed as blue whisker boxes and purple whisker boxes respectively. The Pearson's correlation values are on the y-axis and the coverage levels are on the x-axis

3.5 MATERIALS AND METHODS

Reconstitution of in vitro nucleosomes (invitrosomes)

The 11 different plasmids used in these experiments were derived from two plasmid backbones. The four kat-group plasmids were all derived from the pCR4Blunt-TOPO plasmid (Life Technologies) and each had a different ~150 bp cloned insert in its cloning site. The seven sèt-group plasmids were derived from the pPD149.40 plasmid (a gift from Andrew Fire and the Fire Lab) (S9-Supplemental data pPD149.40) and had different ~150 bp cloned inserts in their unique *Avr* II restriction sites. The ~150 bp inserts in the plasmids were various putative nucleosome-positioning or repelling sequences that were designed for another analysis, and thus were masked and excluded from these analyses as to not obfuscate our testing of the reproducibility of invitrosome experiments, except in the case of p7.6_601 where the insert and flanking sequences were retained as described. For *in vitro* nucleosome reconstitution, all of the plasmids were linearized by restriction digestion with *Sca* I, cutting, in both groups, at a unique *Sca* I restriction site opposite of the insert sites. Invitrosomes were formed on the *Sca* I-linearized plasmid templates in separate experiments using the previously described salt dialysis technique [41]. Recombinant *Xenopus* histone octamers (a gift from Geeta Narlikar and the Narlikar Lab) and DNA templates were reconstituted at a ratio of 1 octamer per 1000 bp of linear plasmid DNA, resulting in a 1:4 molar ratio of DNA to histone octamer. Specifically, for each template 9.67 ug of DNA and 1.50 ug of histone octamer were reconstituted in a total volume of 200 ul.

Isolation of invitrosome core DNA fragments

Invitrosome core DNAs from all 11 invitrosome reconstitutions were isolated as previously described [38, 47]. Briefly, for each experiment 60 ul of invitrosomes were digested with

MNase (Roche) at 1 U/ul for 15 min at room temperature, histone proteins were digested using proteinase K (Roche), DNA was isolated using phenol/chloroform extractions and ethanol precipitation, mononucleosome core DNAs were isolated on a 2% UltraPure Agarose (Life Technologies) gel, extracted using a QIAquick Gel Extraction Kit (Qiagen), and eluted in 30 ul of EB (Qiagen).

End repair, linker ligation and library sequencing

Invitrosome core DNAs were processed and ligated with sequencing adaptors as previously described [38] with the following exceptions. For all samples the entire 30 ul of isolated invitrosome DNA cores were processed. Previously annealed duplex barcoded adaptors were added to each sample according to Table 1 (see S10-Supplemental data adaptors for adaptor sequences) and were incubated with T4 DNA ligase for 4.5 hours rather than 6.5 hours. After the ligated bands were isolated there was no amplification of the libraries, but rather 12 ul (out of 30 ul) of each of the 11 barcoded libraries were pooled together to make a single multiplexed Illumina library. This multiplexed library was sequenced on a single lane of the Illumina GAII system resulting in 2,118,543 single-end, 36-bp reads corresponding to the 11 plasmids (see Table 2 and S11-Supplemental data raw reads).

Nucleosome Mapping

Multiplexed reads were parsed by barcode using custom Perl scripts. After removal of the 4-bp barcodes, the 32-bp parsed reads were mapped back to their respective reference plasmids using a local installation of BLASTN. The BLASTN settings used were –task blastn –best_hit_overhang .1. Reads were analyzed with Fred Tan's custom Perl script summaryPsl-v2.pl

[2], and for each read with multiple hits in the BLASTN output, only the hit with the best bit-score was chosen and used in our analysis.

Nucleosome Coverage

Coverage plots were created with custom Perl scripts by informatically extending upstream (from reverse reads) or downstream (from forward reads) a nucleosome length of 147 bp from the start of the read. Every bp within the reference DNAs was given a count of 1 for each nucleosome overlapping that site. Counts were compiled and used to create a coverage plot. Combined coverage plots were made by adding the counts at identical positions from corresponding positive and negative coverage plots. In order to eliminate any positional effects due to end bias or the putative positioning sequences, BLASTN outputs went through additional filters to remove all reads that overlapped a 147 bp window on either end of the linearized plasmid and flanking the insert (except in the case of p7.6_601 where reads mapping to the insert and flanking regions were retained). These pools of filtered reads were used as inputs for extracting reads to achieve a specific coverage level.

Scaling Coverage plots

To normalize coverage plots with unequal coverage levels, all coverage plots were scaled to the plot with the lowest level of coverage before performing Pearson's correlations. Scalars were calculated by dividing the number of filtered reads (Table2 reads for coverage plots) for a coverage plot by the number of filtered reads for the coverage plot with the least coverage. The value for the coverage plot at each base pair was then divided by this scalar to yield normalized, scaled coverage plots.

Pearson's correlation coefficients

Pearson's correlation coefficients were calculated using Prism 6 version 6.0d for Mac OSX. Only values that were not in the insert and 147 bp filtering window were used in the calculations, except in the case of p7.6_601 as described. If there were no counts at a specific bp, its value was left blank.

3.6 ACKNOWLEDGMENTS

SMJ and CEK were supported by Mentoring Environment Grants from Brigham Young University's Office of Research & Creative Activities and other funding from the Department of Microbiology and Molecular Biology and the College of Life Sciences at Brigham Young University. We thank Frederick Tan for help with and use of his custom Perl scripts, Mark Clement and Justin Page for help with our custom Perl scripts, Sam Gu and Ayelet Lamm along with the Arend Sidow Lab for Illumina cluster generation and sequencing, Brent Nielsen and Jordon Richie for careful reading of the manuscript, Geeta Narlikar and her lab for help with *in vitro* reconstitution experiments and input on the manuscript, Andrew Fire for valuable ideas and support, and Vladimir Teif and Reviewer #2 for helpful suggestions. BLASTN analyses were performed on Marylou at the Fulton Supercomputer Center at BYU.

3.7 SUPPLEMENTAL MATERIAL

Supplemental Material for:

Reproducibility and Consistency of *In Vitro* Nucleosome Reconstitutions Demonstrated by Invitrosome Isolation and Sequencing

Colton E. Kempton, Justin R. Heninger, and Steven M. Johnson*

Department of Microbiology and Molecular Biology, Brigham Young University, Provo, Utah, USA

Contents

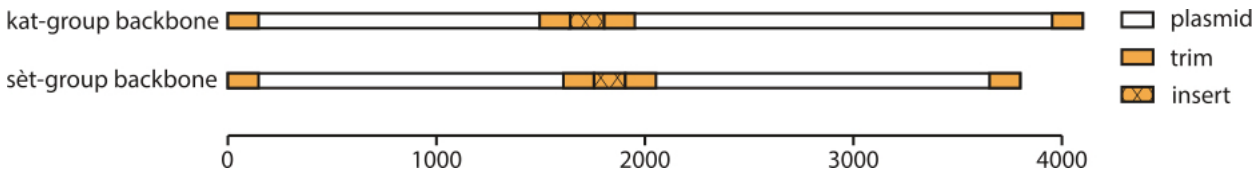


Figure 3-S1. Diagram of plasmid backbones.

Linear depiction of both the kat-group plasmid and the sèt-group plasmid backbones. The size of the plasmids is indicated in bp by the scale bar at the bottom and the areas excluded or "trimmed" from the analysis are shown in gold. The variable ~150 bp inserts in the different plasmids are shown by the cross-hatched shading within the central trimmed regions.

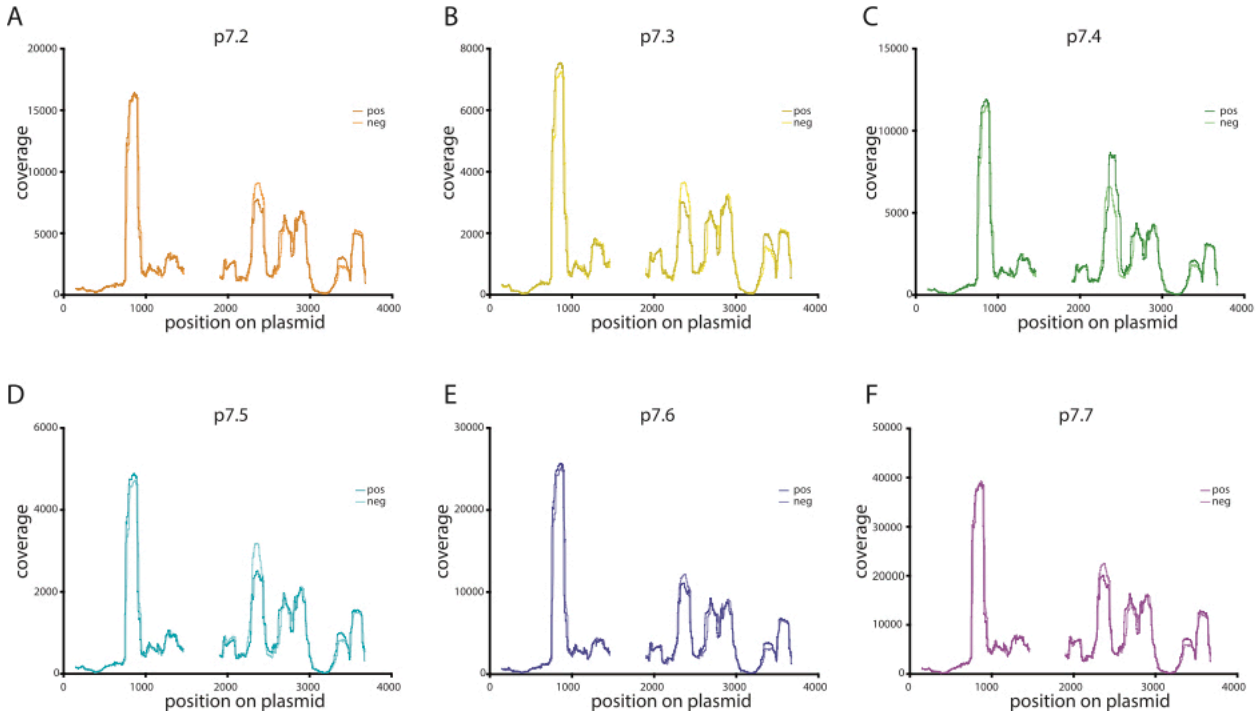


Figure 3-S2. Invitrosome coverage plots for sèt-group plasmids.

Invitrosome coverage plots for sèt-group plasmids show near identical coverage patterns. Invitrosome coverage on each plasmid is plotted on the y-axis and the plasmid coordinates in bp are plotted on the x-axis. Gaps in the plot represent the trimmed insert site and end regions on each plasmid. 1A-1F Non-scaled coverage plots with forward (pos) and reverse (neg) invitrosomes plotted separately for plasmids p7.2-p7.7

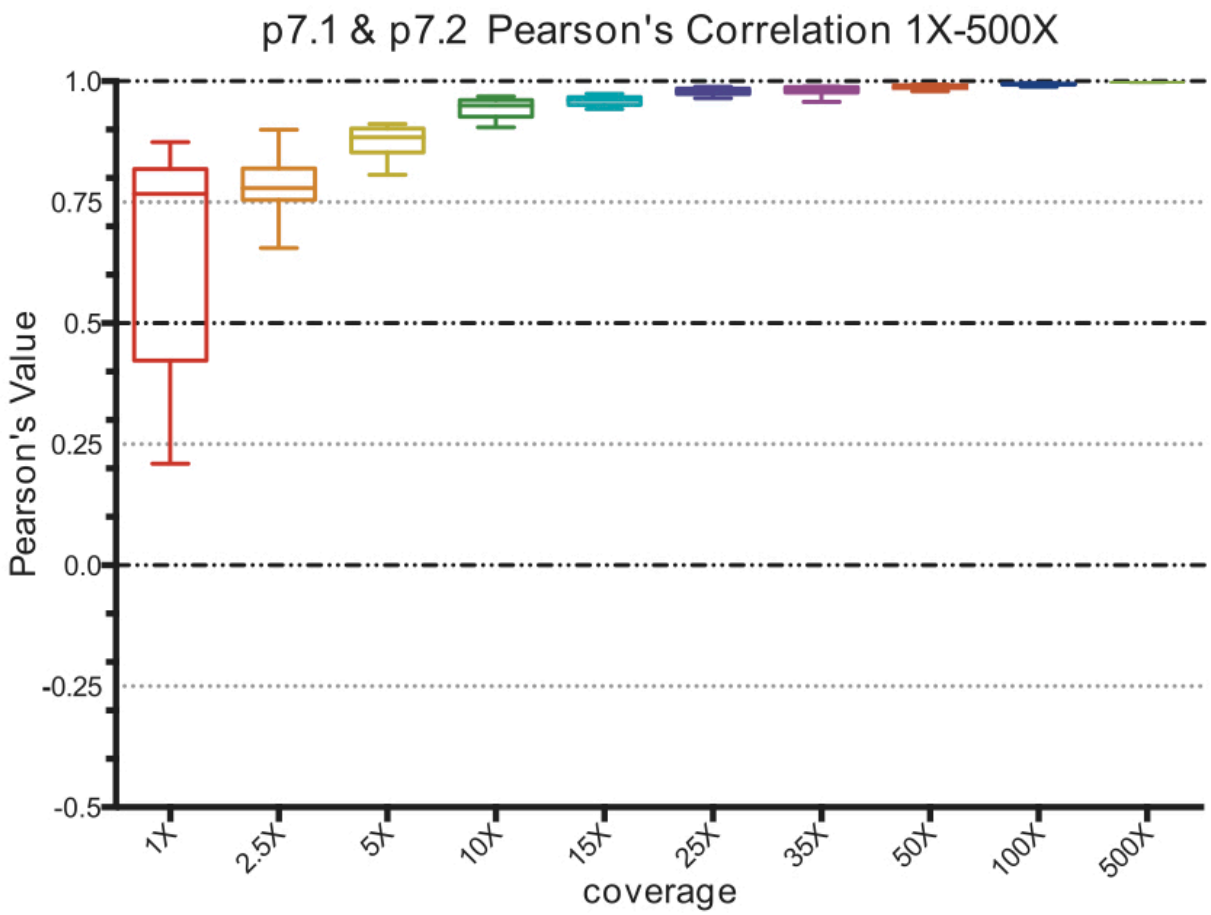


Figure 3-S3. Whisker plots for the sèt-group plasmids.

Range of correlation values at each coverage level for the pair-wise comparisons of the three replicates for plasmid p7.1 and the three replicates for plasmid p.7.2. The Pearson's correlation values are on the y-axis and the coverage levels are on the x-axis. The range of correlation values for each coverage level is plotted as a whisker plot composed of the 30 pair-wise comparisons at each coverage level.

Figures 3-S4-S7. Whisker plots for the sèt group plasmids.

Whisker plot graphs showing the variability of correlation values at low levels of coverage (1X-10X) for all combinations of the sèt -group plasmids. The Pearson's correlation values are on the y-axis and the coverage levels are on the x-axis. The range of correlation values for each coverage level is plotted as a whisker plot composed of the 30 pair-wise comparisons at each coverage level.

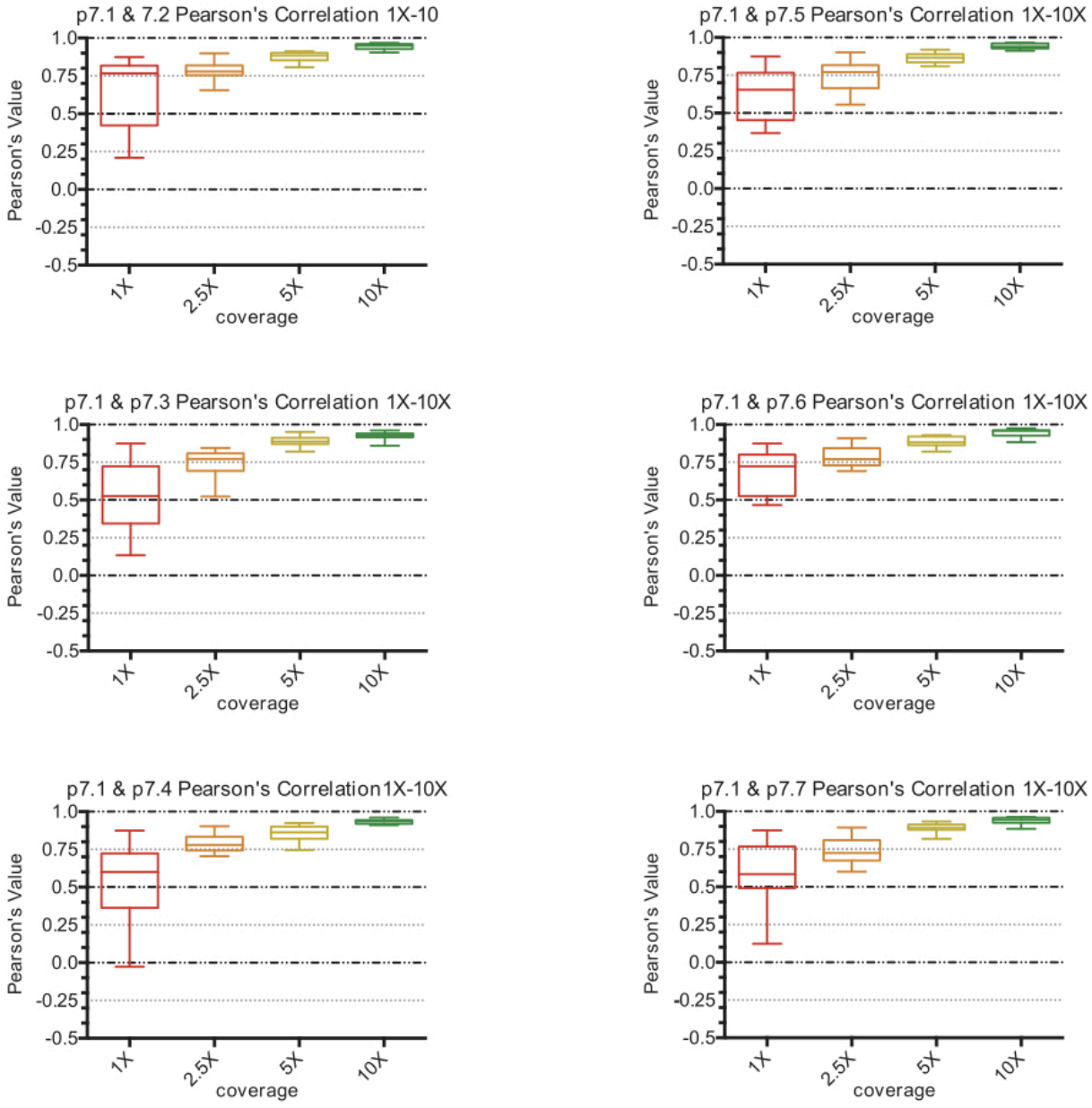


Figure 3-S4 Supplemental

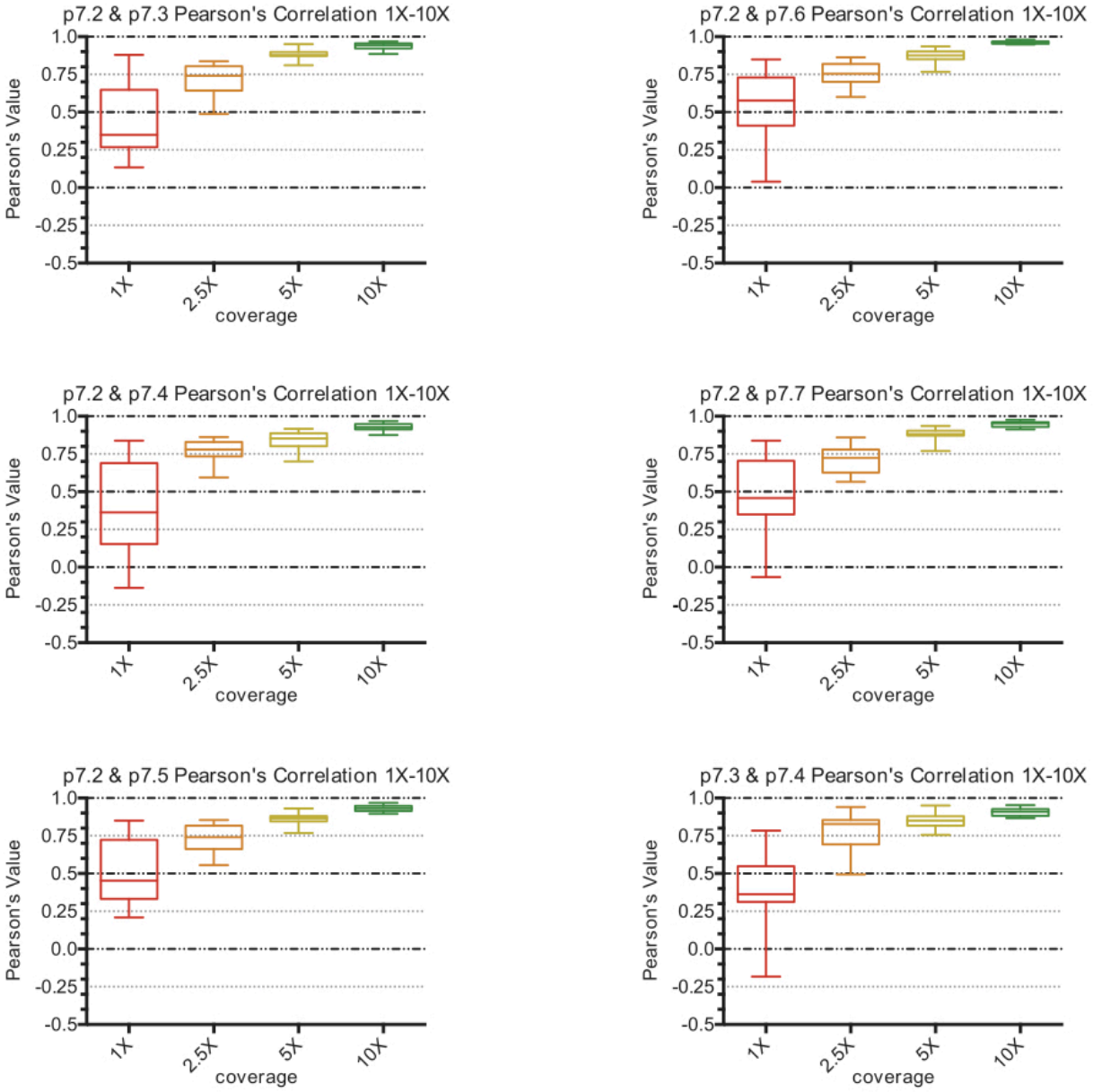
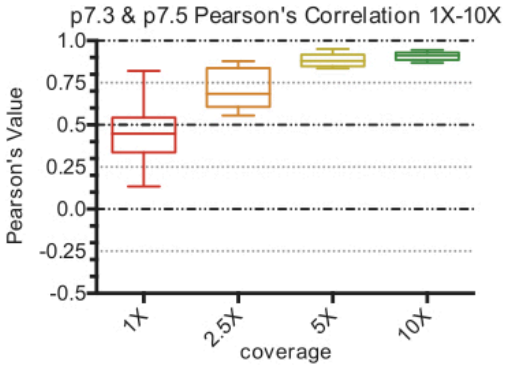
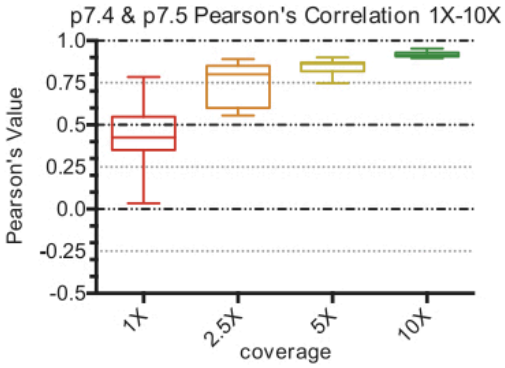
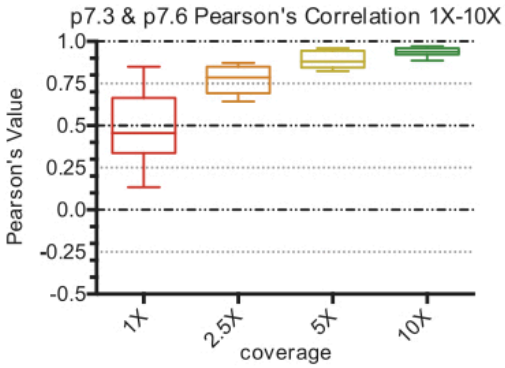
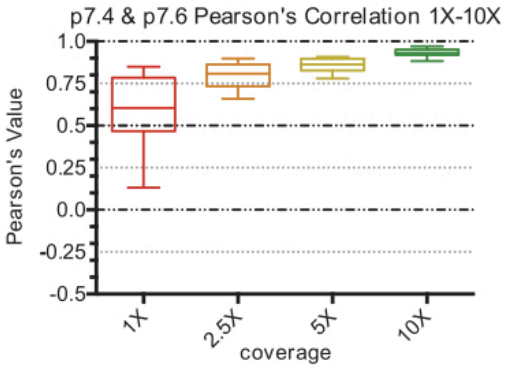


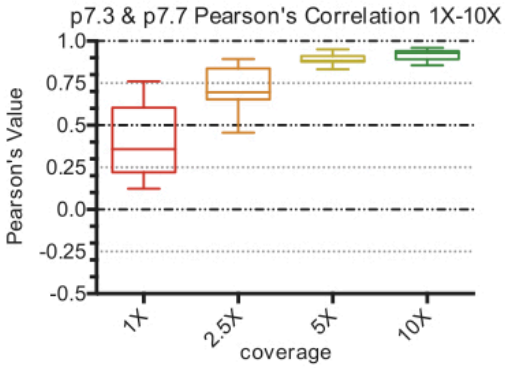
Figure 3-S5 Supplemental











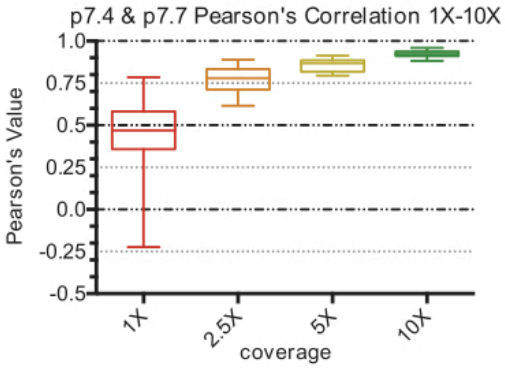


Figure 3-S6 Supplemental

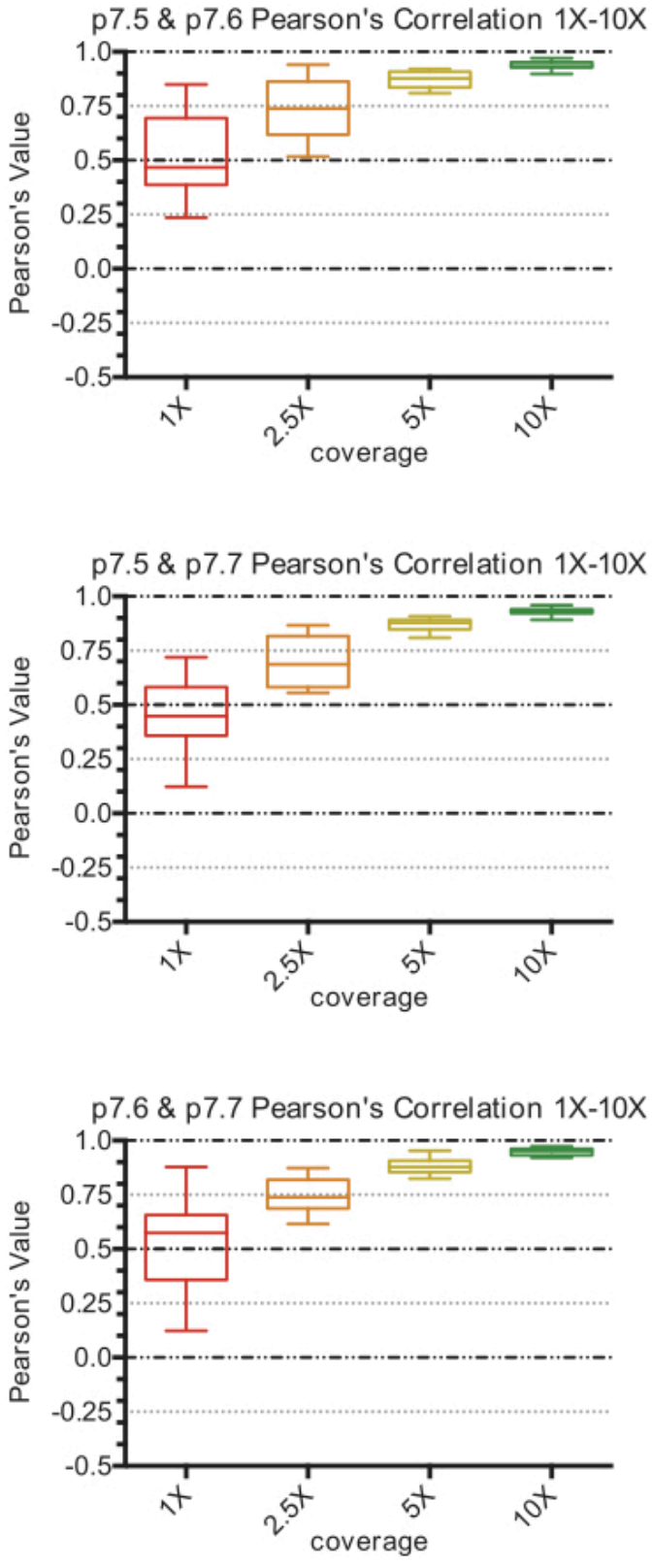


Figure 3-S7 Supplemental

Figure 3-S8. Supplemental data pPD149.40 plasmid sequence.

>pPD149.40
GGTACCCGTCGACACATACTACCGCCTGGCCATTCCTAGTGGAGTGCTTTTACCAGT
TGAGCGGAAAGCGTTTTCTTTTTAATTTACCGGACGACTGCTGAACGAGAATTGCAT
TGCCTTTATAGGGCTTGATTGTCTGCCGGGATCGAGTAGAGTGCGTCGCGACACGCT
TCGCTTGTAGCTTTCTTTCAGACGCGCGATGTATATACGGGTGGGTCCGACGGAAGC
CACTTCGCCGAGGCGCCTCGCGGCCCTGCAAGCTGGCGGAAACCGTTCCTACTGCGA
TCGCTTATCTGCTTAAATTCTGCTACATGCTTGAGCAGCAAGCACGGTCGAACCAGC
CCACCACCAGGGAGCGTACCTCCCCACTTCGCGCTTACTTATGTTACCAAGCTCGAA
GTTATTGTACTCGCGATGGCGTTACATCGAGAGCGAGCCCCTCTCCTTAAATCAGGT
GTCGCGTGGCCAGTAATTTACCAAAGCTGATAAGTCCTCTGCTAATCCCGGCGTCGC
GGCCTAGGCGCGGGAAGTCACTGTCGTTCATCCTATGCGCCGGGATGCTTACTGCGG
CTACTTGAGCCAGACTTCCCTTGAGATTCCATTTCCAACCGTATATTACTGTCCCTTA
TTCCGGTGGCCTGCGGAGTCCGGTCCGCCCCACGCTTCACGATTTTTCCGGTTGCGA
GCGCATATCGTTCCTCTTTTCCTTCAAATTTGCCCGGCCGGCGCGGGTCCTGCGCGG
AGTTTTTATCTAGCTGCCAGGCGTTGTATGATCTTCACCGGTCtctagagcggccgccaccgcggt
ggagctcCAGCTTTTGTTCCCTTTAGTGAGGGTTAATTGCGCGCTTGGCGTAATCATGGT
CATAGCTGTTTCCTGTGTGAAATTGTTATCCGCTCACAATTCCACACAACATACGAG
CCGGAAGCATAAAGTGTAAAGCCTGGGGTGCCTAATGAGTGAGCTAACTCACATTA
ATTGCGTTGCGCTCACTGCCCGCTTTCCAGTCGGGAAACCTGTCGTGCCAGCTGCAT
TAATGAATCGGCCAACGCGCGGGGAGAGGCGGTTTGCGTATTGGGCGCTCTTCCGCT
TCCTCGCTCACTGACTCGCTGCGCTCGGTCGTTCGGCTGCGGCGAGCGGTATCAGCT
CACTCAAAGGCGGTAATACGGTTATCCACAGAATCAGGGGATAACGCAGGAAAGAA
CATGTGAGCAAAAGGCCAGCAAAAGGCCAGGAACCGTAAAAAGGCCGCGTTGCTG
GCGTTTTTCCATAGGCTCCGCCCCCCTGACGAGCATCACAAAAATCGACGCTCAAGT
CAGAGGTGGCGAAACCCGACAGGACTATAAAGATACCAGGCGTTTCCCCCTGGAAG
CTCCCTCGTGCGCTCTCCTGTTCCGACCCTGCCGCTTACCGGATACCTGTCCGCCTTT
CTCCCTTCGGGAAGCGTGGCGCTTTCTCATAGCTCACGCTGTAGGTATCTCAGTTCG
GTGTAGGTCGTTCGCTCCAAGCTGGGCTGTGTGCACGAACCCCCCGTTCAGCCCGAC
CGCTGCGCCTTATCCGGTAACTATCGTCTTGAGTCCAACCCGGTAAGACACGACTTA
TCGCCACTGGCAGCAGCCACTGGTAACAGGATTAGCAGAGCGAGGTATGTAGGCGG
TGCTACAGAGTTCTTGAAGTGGTGGCCTAACTACGGCTACACTAGAAGGACAGTATT
TGGTATCTGCGCTCTGCTGAAGCCAGTTACCTTCGGAAAAAGAGTTGGTAGCTCTTG
ATCCGGCAAACAAACCACCGCTGGTAGCGGTGGTTTTTTTGTTTGCAAGCAGCAGAT
TACGCGCAGAAAAAAAGGATCTCAAGAAGATCCTTTGATCTTTTCTACGGGGTCTGA
CGCTCAGTGGAACGAAAACTCACGTTAAGGGATTTTGGTCATGAGATTATCAAAAA
GGATCTTCACCTAGATCCTTTTAAATTAAAAATGAAGTTTTAAATCAATCTAAAGTA
TATATGAGTAAACTTGGTCTGACAGTTACCAATGCTTAATCAGTGAGGCACCTATCT
CAGCGATCTGTCTATTTCGTTCATCCATAGTTGCCTGACTCCCCGTCGTGTAGATAAC
TACGATACGGGAGGGCTTACCATCTGGCCCCAGTGCTGCAATGATACCGCGAGACC
CACGCTCACCGGCTCCAGATTTATCAGCAATAAACCAGCCAGCCGGAAGGGCCGAG
CGCAGAAGTGGTCCTGCAACTTTATCCGCCTCCATCCAGTCTATTAATTGTTGCCGG
GAAGCTAGAGTAAGTAGTTCGCCAGTTAATAGTTTGCGCAACGTTGTTGCCATTGCT
ACAGGCATCGTGGTGTCACGCTCGTCGTTTGGTATGGCTTCATTCAGCTCCGGTTCCC

AACGATCAAGGCGAGTTACATGATCCCCCATGTTGTGCAAAAAAGCGGTTAGCTCCT
TCGGTCCTCCGATCGTTGTCAGAAGTAAGTTGGCCGCAGTGTTATCACTCATGGTTA
TGGCAGCACTGCATAATTCTCTTACTGTCATGCCATCCGTAAGATGCTTTTCTGTGAC
TGGTGAGTACTCAACCAAGTCATTCTGAGAATAGTGTATGCGGCGACCGAGTTGCTC
TTGCCCGGCGTCAATACGGGATAATACCGCGCCACATAGCAGAACTTTAAAAGTGCT
CATCATTGGAAAACGTTCTTCGGGGCGAAAACTCTCAAGGATCTTACCGCTGTTGAG
ATCCAGTTCGATGTAACCCACTCGTGCACCCAACTGATCTTCAGCATCTTTTACTTTC
ACCAGCGTTTCTGGGTGAGCAAAAACAGGAAGGCAAAATGCCGCAAAAAAGGGAA
TAAGGGCGACACGGAAATGTTGAATACTCATACTCTTCCTTTTTCAATATTATTGAA
GCATTTATCAGGGTTATTGTCTCATGAGCGGATACATATTTGAATGTATTTAGAAAA
ATAAACAAATAGGGGTTCCGCGCACATTTCCCCGAAAAGTGCCACCTAAATTGTAA
GCGTTAATATTTTGTTAAAATTCGCGTTAAATTTTTGTTAAATCAGCTCATTTTTTAA
CCAATAGGCCGAAATCGGCAAAATCCCTTATAAATCAAAAGAATAGACCGAGATAG
GGTTGAGTGTTGTTCCAGTTTGGAACAAGAGTCCACTATTAAAGAACGTGGACTCCA
ACGTCAAAGGGCGAAAAACCGTCTATCAGGGCGATGGCCCACTACGTGAACCATCA
CCCTAATCAAGTTTTTTGGGGTCGAGGTGCCGTAAAGCACTAAATCGGAACCCTAAA
GGGAGCCCCCGATTTAGAGCTTGACGGGGAAAGCCGGCGAACGTGGCGAGAAAGG
AAGGGAAGAAAGCGAAAGGAGCGGGCGCTAGGGCGCTGGCAAGTGTAGCGGTCAC
GCTGCGCGTAACCACCACACCCGCCGCGCTTAATGCGCCGCTACAGGGCGCGTCCC
ATTCGCCATTCAGGCTGCGCAACTGTTGGGAAGGGCGATCGGTGCGGGCCTCTTCGC
TATTACGCCAGCTGGCGAAAGGGGGATGTGCTGCAAGGCGATTAAGTTGGGTAACG
CCAGGGTTTTCCCAGTCACGACGTTGTAAAACGACGGCCAGTGAGCGCGCGTAATA
CGACTCACTATAGGGCGAATTG

Figure 3-S9. Supplemental data adapters.

Adapters used in this study.

Adaptor Sequences (barcodes are in blue for forward and red for reverse oligos)

AF-SJ-84 5'-
AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATC
TCAGT

AF-SJ-85 5'-
AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATC
TGTCT

AF-SJ-86 5'-
AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATC
TTGCT

AF-SJ-87 5'-
AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATC
TCCCT

AF-SJ-88 5'-
AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATC
TAACT

AF-SJ-89 5'-
AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATC
TGCAT

AF-SJ-90 5'-
AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATC
TCGAT

AF-SJ-91 5'-
AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATC
TTAAT

AF-SJ-92 5'-
AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATC
TATAT

AF-SJ-93 5'-
AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATC
TTCTT

AF-SJ-94 5'-
AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATC
TGATT

AF-SJ-99 5'P-CTGAGATCGGAAGTCGTATGCCGTCTTCTGCTTG

AF-SJ-100 5'P-GACAGATCGGAAGTCGTATGCCGTCTTCTGCTTG

AF-SJ-101 5'P-GCAAGATCGGAAGTCGTATGCCGTCTTCTGCTTG

AF-SJ-102 5'P-GGGAGATCGGAAGTCGTATGCCGTCTTCTGCTTG

AF-SJ-103 5'P-GTTAGATCGGAAGTCGTATGCCGTCTTCTGCTTG

AF-SJ-104 5'P-TGCAGATCGGAAGTCGTATGCCGTCTTCTGCTTG

AF-SJ-105 5'P-TCGAGATCGGAAGTCGTATGCCGTCTTCTGCTTG

AF-SJ-106 5'P-TTAAGATCGGAAGTCGTATGCCGTCTTCTGCTTG

AF-SJ-107 5'P-TATAGATCGGAAGTCGTATGCCGTCTTCTGCTTG

AF-SJ-108 5’P-AGAAGATCGGAAGTCGTATGCCGTCTTCTGCTTG

AF-SJ-109 5’P-ATCAGATCGGAAGTCGTATGCCGTCTTCTGCTTG

Chapter 4. Method to Increase Undergraduate Laboratory Student Confidence in Performing Independent Research

The following chapter is an adaptation of an article published in the *Journal of Microbiology and Biology Education* May 2017 titled "Method to Increase Undergraduate Laboratory Student Confidence in Performing Independent Research."

4.1 ABSTRACT

The goal of an undergraduate laboratory course should be not only to introduce the students to biology methodologies and techniques, but also to teach them independent analytical thinking skills and proper experiment design. This is especially true for advanced biology laboratory courses that undergraduate students typically take as a junior or senior in college. Many courses achieve the goal of teaching techniques, but fail to approach the larger goal of teaching critical thinking, experimental design, and student independence. Here we describe a study examining the application of the scaffolding instructional philosophy in which students are taught molecular techniques with decreasing guidance to force the development of analytical thinking skills and prepare undergraduate students for independent laboratory research. This method was applied to our advanced molecular biology laboratory class and resulted in an increase of confidence among the undergraduate students in their abilities to perform independent research.

Key Words

Undergraduate learning, scaffolding methodology, guided learning, laboratory class, independent research.

4.2 INTRODUCTION

Cookbook-type protocols, common to high school and undergraduate-level science classes, are a less effective means of instruction as they allow students to be passive and typically do not require critical thinking [51]. They do not accurately reflect the investigative nature of science, where there is no accompanying fill-in-the-blank, universal protocol that is used to discover new information [52] [53]. To improve the quality of science education, there has been a push to replace these cookbook-style protocols with more open-ended investigative or inquiry-type instruction that is student centered [54-57]. Our research has demonstrated increased engagement when students use, or anticipate using, data from their own genome [58]. Inquiry-based learning activities model the scientific process much better than cookbook labs and lead to increased understanding of the scientific process [59, 60].

To promote independent learning, we designed our undergraduate course to apply the scaffolding instructional methodology [61] to wean students from cookbook laboratory procedures by sequentially introducing protocols with decreasing amounts of written instructor guidance. Scaffolding originated as adults helped children develop higher psychological functioning and ability to express themselves through guided interactions [62, 63], ultimately enabling children to do things independently that normally require adult guidance and assistance [63]. We have applied this method to our undergraduate Advanced Molecular Biology Laboratory at Brigham Young University (Appendix A, Methods) with the goal of teaching the students to find and use protocols and develop scientific independence. This method enables student transition from instructor dependence to scientific independence.

General application of this method involves students performing a series of planned experiments while sequentially providing them with 1) protocols with step-by-step instructions

typed out by the professor, 2) instruction with manufacturers' protocols augmented with additional explanations inserted by the professor, 3) unaugmented manufacturers' protocols, 4) protocols received from scientists, 5) a primary literature protocol, and finally, 6) protocols found by the students themselves (Fig. 1). We applied this method to our Advanced Molecular Biology Laboratory course. Results from our student survey demonstrated significant increases in student confidence to use and adapt new protocols to carry out experiments. Students also showed greatly increased confidence in their ability to troubleshoot and to carry out independent research experiments.

4.3 PROCEDURE

Simple, professor-provided protocols

We start with simple, professor-written protocols (Fig. 1). These instructions include detailed steps to accomplish the experiments adapted from kit instructions and simplified for ease of use.

We applied this principle with our DNA fingerprinting module (Fig. 2): students isolate genomic DNA [64], perform PCR, do PCR DNA cleanup and restriction enzyme digests [65], and analyze DNA on gels. Each of the protocols is step-by-step instructions typed out by the professor.

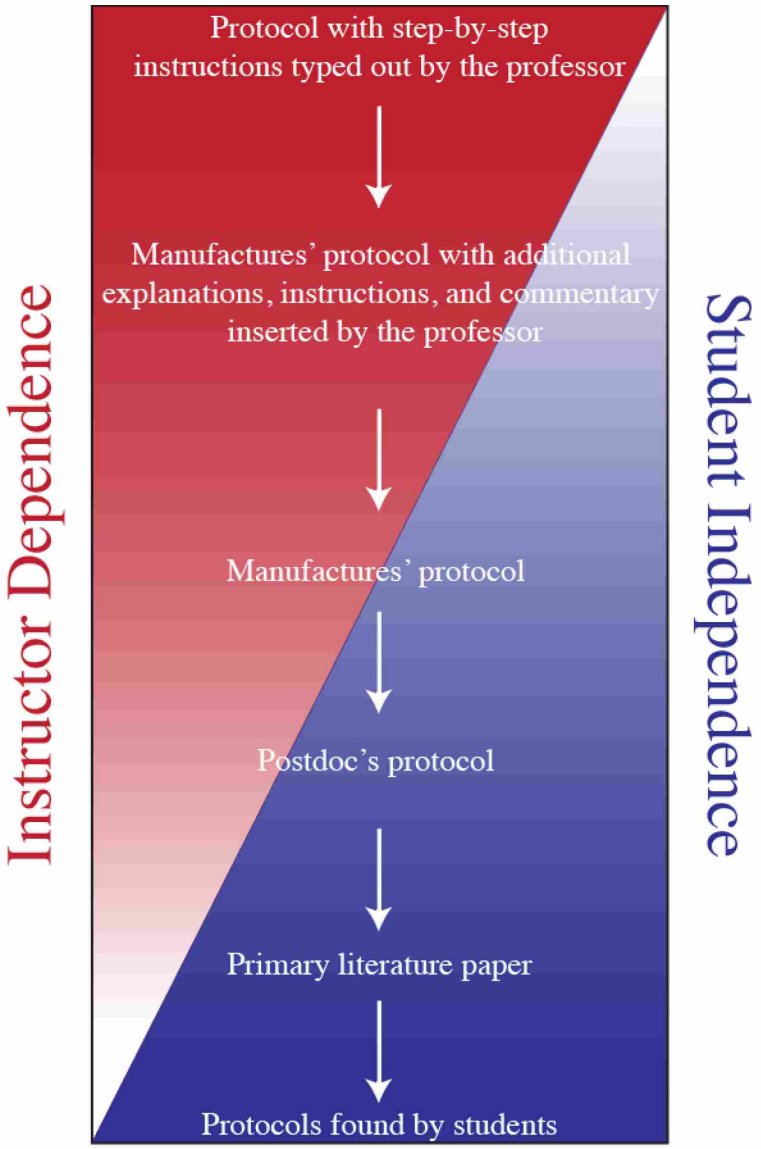


Figure 4-1. Overview of the application of the weaning philosophy and approach.

The weaning approach is applied to any laboratory class by initially providing students with protocols that are highly modified by the professor (1st and 2nd), followed by protocols with decreasing amounts of professor modifications and protocols with no professor modifications (3rd), and finally resources from which the students must extrapolate protocols (4th and 5th). Ultimately, students are not provided with protocols, but instead find protocols on their own (last). The red color on the left that decreases from top to bottom represents the amount of student dependence on the written instructions from the professor, and the blue color on the right that increases from top to bottom represents the amount of student independence at each stage of the weaning.

Module	Experiment	Resource
DNA Fingerprinting	DNA ioslation PCR PCR DNA clean-up Agarose gel electrophoresis Restriction enzyme digestion	Protocol and step-by-step instructions typed out by the professor
Site-directed Mutagenesis	Plasmid isolation Restriction digest and electrophoresis Site-directed mutagenesis Bacterial transformation Colony selection and colony PCR DNA sequencing submission & analysis	Manufactures' protocol with additional explanations, instructions and commentary inserted by the professor
Northern Blotting	RNA isolation Northern blotting Chemiluminescent detection	Manufactures' protocol
EMSA	Nuclear protein isolation Electrophoretic mobility shift assay Chemiluminescent detection	Postdoc's protocol Primary literature paper
Independent Projects	Variable	Protocols found by students

Figure 4-2. Specific application of the weaning philosophy and approach.

The name of each specific module (left) is listed with its accompanying experiments (middle) and the type of resources that are provided for those experiments (right). The color and intensity of the background fields of the modules represents the amount of professor dependence (red) or student independence (blue) in each module (see Fig. 1). Experiments listed in green print are procedures that the students have learned in previous modules. The other color print (black or white) differs only for ease of reading.

Manufacturers' protocols with added instructions

In the second phase we use protocols/instructions that come with kits, supplemented with additional instructions by the professor (Fig. 1).

Our Site-Directed Mutagenesis module (Fig. 2) applies this principle. Students isolate plasmid DNA, perform site-directed mutagenesis, bacterial transformations, colony selection and colony PCR and sequence PCR products. We use supplemented protocols from the QIAprep

Miniprep [66], Phusion™ Site-Directed Mutagenesis [67], and the Zero Blunt® TOPO® PCR Cloning Kits [68]. Students apply first module protocols as they perform restriction digests, gel electrophoresis, and colony PCR in preparation for sequencing to confirm the success of their mutagenesis.

Manufacturers' protocols

In the third phase we provide students with only the protocols/instructions that come from the kit (Fig. 1). We use three protocols (short, long, and average-length) students might actually experience in the real world. Students must glean what is necessary from the protocol for them to do the experiment.

Our Northern Blotting module applies this principle (Fig. 2). Students isolate RNA using TRIzol® Reagent [69] with a two-page protocol outlining multiple procedures. This is followed by northern blotting using a detailed 42-page NorthernMax®-Gly kit and protocol [70]. Students look through the protocol and decide which steps to include for their application. Finally, we use the Chemiluminescent Nucleic Acid Detection Module and protocol, a straightforward kit and instructions, to visualize the probe on their blots.

Real-life protocols

The final phase in our methodology toward independence is to use protocols the students might receive from other researchers when trying to reproduce published techniques. Students receive a protocol sent from a postdoctoral fellow and a primary-literature paper from which they need to reproduce an experiment. Students follow the postdoc protocol and read and understand the primary literature paper to glean what they need to replicate the experiments contained therein. These are the types of protocols they might encounter in a research career. Using and

applying them in a carefully controlled laboratory experience prepares them for independent research.

We applied this principle with our Electroporation Mobility Shift Assay (EMSA) module. The instructional resources for these experiments are a primary research paper [71] and a protocol from a postdoc [72]. Students determine how to perform the EMSA in the paper from the materials and methods section and additional outside resources online. The terse protocol provided by the postdoc has each step for nuclear protein isolation, but no logistical commentary. The chemiluminescence kit and protocol used in the third module is again used here to reinforce the skills they previously acquired.

Independent application

Having experienced a range of instructional materials and performed several molecular techniques, students are asked to directly apply what they have learned throughout the semester. The culminating event is when students choose, design, and perform independent projects for the last four weeks of the semester. Students independently come up with their own scientific questions, plan the procedures, find the necessary protocols, and perform the experiments. Instructors only approve their projects and provide the necessary reagents.

The pinnacle event is the last day of class when students present their independent projects, complete with background, hypothesis, experimental procedures, data, results, and conclusions to the entire class. With the final independent project, the students have moved from preplanned, instructor-dependent, results-controlled experiments to independently conceived, designed, and executed projects that succeed or fail based on the student.

4.4 CONCLUSION

Here we present the application of a scaffolding pedagogical method to transform undergraduate laboratory students into independent researchers. We surveyed student attitudes about their abilities to perform independent research. Student abilities to independently plan and execute appropriate experiments increased, as did their confidence to do independent research (Appendix B, Measuring Learning). This methodology is likely applicable to any lab course in life sciences striving to develop independent undergraduate researchers. Consistent results between three sections taught by three different professors suggest that this method is not instructor specific, but generally applicable.

4.5 ACKNOWLEDGEMENTS

The authors declare that there are no conflicts of interest.

4.6 SUPPLEMENTAL MATERIALS

Supplemental Materials for:

Method to Increase Undergraduate Laboratory Student Confidence in Performing Independent Research[†]

Colton E. Kempton, Scott Weber, and Steven M. Johnson*

Department of Microbiology and Molecular Biology, Brigham Young University, Provo, Utah 84602

1. Appendix A: Methods, Course sections, Survey instrument, Data analysis, Advanced molecular biology laboratory protocols, Laboratory safety procedures
2. Appendix B: Measuring learning, Figure 4-S1, Figure 4-S2

Appendix A: Methods

Course Sections

Students were third- and fourth-year undergraduate students enrolled in a 15-week advanced molecular biology laboratory course that met for three hours, twice a week (Advanced Molecular Biology Lab, MMBIO 442) offered Fall 2015 and Winter 2016 semesters at Brigham Young University. 21 students who were enrolled in one section of MMBIO 442 in the fall semester, and 28 students who were enrolled in two separate sections of MMBIO 442 winter semester participated in the study. The students were informed that participation in the survey was purely voluntary and did not affect their course grades in any way. The single fall-semester and two winter-semester sections of MMBIO 442 were taught by three different professors. This study design was reviewed and approved by the Brigham Young University Institutional Review Board (study approval number E15320)

Survey Instrument

On the first day and last day of the course, we administered the same survey (Pre- and Post-Survey). This survey consisted of statements to gauge student attitudes and perceptions about research and their own research capabilities. Student attitudes were measured according to their level of agreement with the survey statements on a 5-point Likert scale. The students were informed that this survey, while scored for curriculum assessment, did not contribute to their grade in the course and, despite not contributing any points in the course, were encouraged to do their best. The pre- and post-surveys were not scored until after final grades were assigned for the course.

Data Analysis

The pre- and post-surveys were scored according to the level of agreement with each statement. 21 students took the pre-survey and post-survey for fall semester and 28 students took both the pre- and post-surveys winter semester. Only scores from students who took both pre- and post-surveys were used in the analysis. The change in student pre-course and post-course confidence in their ability to independently perform scientific research was assayed by the pre-survey and post-survey scores. These

were evaluated by paired t-tests. Statistical analysis and graphs were done using Prism 6 version 6.0d for Mac OSX.

Advanced Molecular Biology Laboratory Protocols

Protocols with professor-typed step-by-step instructions, manufactures' protocols with additional instructions, and the postdoctoral fellow-provided protocol used in our Advanced Molecular Biology Laboratory course are available upon request.

Laboratory Safety Procedures

All organisms use in the experiments are BSL-1 organisms, i.e., DH5a *E. coli*, *C. elegans* and cheek epithelial skin cells. All students received laboratory safety training the first day of class before starting any experiments and the ASM *Guidelines for Biosafety in Teaching Laboratories* were fully applied http://www.asm.org/index.php/educators/laboratory-safety-guidelines.

Appendix B:

Measuring Learning

The main goal of our method presented here is to transition students from instructor-dependent learners to independent and confident researchers capable of designing and carrying out their own experiments. We surveyed student attitudes and beliefs about their ability to do research at the start and end of the course. After completion of the course, students had greater confidence in their abilities to perform independent research (Supplemental Materials, Appendix B, Supplemental Figures 4-S1 and 4-S2).

Each training phase builds on previous phases to prepare students for their independent projects. Students should gradually become comfortable with new protocols and eventually be able to adapt them to fit the needs of their independent projects. To test this hypothesis, we compared pre-survey and post-survey scores and quantified student attitudes toward research (Figs S1 and S2). The post-survey showed a statistically significant decrease in anxiety experienced by the students when presented with new protocols (Figs S1A and S1B question 2). Responses to questions 4 and 5 indicate students are

significantly more comfortable following protocols that are not explicit and can adapt these protocols to meet their needs (Figs S1A and S1B questions 4 and 5). Students also gained greater confidence identifying and using protocols found in primary literature (Figs S1A and S1B question 3).

Independent research projects are a great opportunity for students to implement what they have learned and practiced during the course. Initial experiments are often unsuccessful or only partially successful, meaning no data were generated or they were inconclusive. To proceed, students have to identify potential sources of error or adjust the parameters of their experiment. Our surveys found that after completing their independent projects, students are more confident in their abilities to interpret experimental results, identify sources of experimental error, troubleshoot, and perform independent scientific research (Figs S2A and S2B questions 7, 8, 9, and 10). These results indicate that the students indeed improved in their abilities to independently plan appropriate experiments, and are evidence that the students increased in their abilities to do independent research.

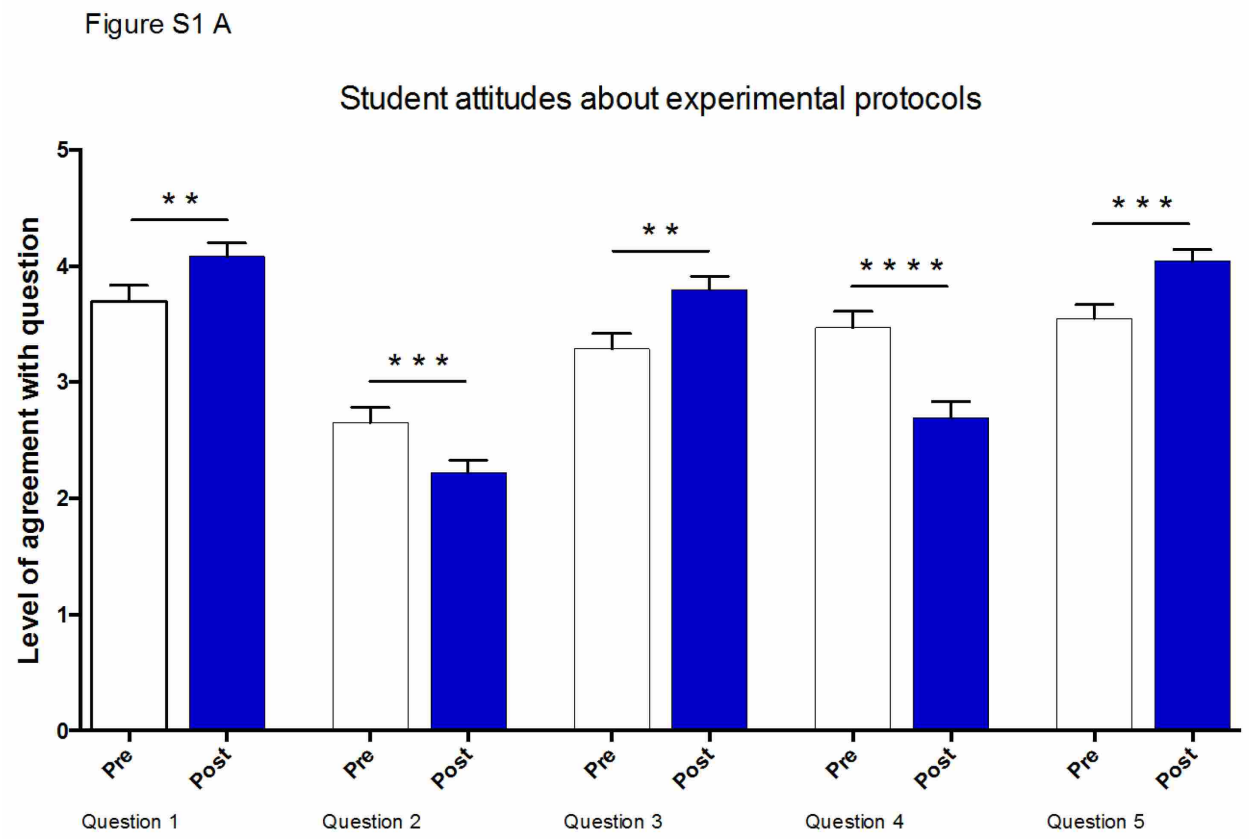


Figure 4-S1A. Student attitudes about experimental protocols.

S1A, student responses to questions 1-5 on a 5-point Likert scale. Statistical analysis was done using a paired t-test. All values are mean with ± SEM and an n=49 for pre-survey (white) and n=49 for post-survey (blue) questions. The p-values for each question were: 1. **=P 0.0022, 2. ***=P 0.0008, 3. **=P 0.0014, 4. ****=P<0.0001, and 5. ***=P 0.0004.

Figure 4-S1B. Student survey questions.

1. I have done a lot of scientific research experiments compared to the average undergraduate student.
2. I get anxious when presented with a new experimental protocol.
3. I am comfortable synthesizing and using protocols described in primary literature.
4. I am only comfortable following new protocols when the directions are explicit.
5. I can adapt general and or generic protocols to test a hypothesis.

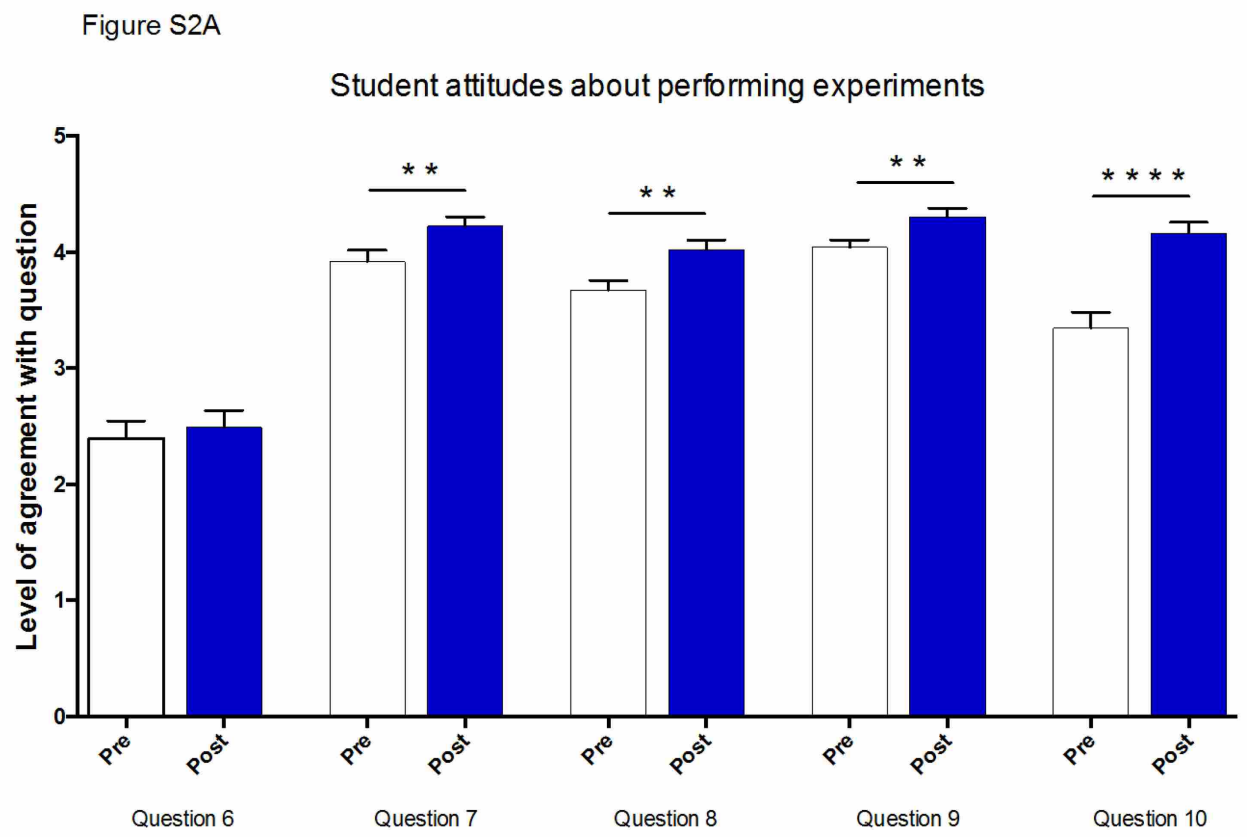


Figure 4-S2A. Student attitudes about performing experiments.

S2A, student responses to questions 6-10 on a 5-point Likert scale. Statistical analysis was done using a paired t-test. All values are mean with ± SEM and an n=49 for pre-survey (white) and n=49 for post-survey (blue) questions. The p-values for each question were: 6. not significant, 7. **=P 0.0029, 8. **=P 0.0015, 9. **=P 0.0055, and 10. ****=P <0.0001.

Figure 4-S2B. Student survey questions.

6. Experiments are usually successful the first time they are attempted.
7. I am comfortable interpreting experimental results and drawing conslusions form them.
8. If an experiment doesn't work as expected or fails to yield results, I can figure out what happened.
9. I can identify potential sources of error in my experiments.
10. I consider myself capable of doing independent scientific research.

Chapter 5. Conclusions about this work Future Directions

5.1 NUCLEOSOME POSITIONING VIA DNA SEQUENCE MANIPULATION

Innate silencing of transgene expression remains a barrier to effective gene therapy. This innate silencing is most likely an interplay between many different factors. We have shown that one element, DNA sequence identity, can have a powerful effect on the behavior of transgene expression *in vivo.* While our results were not what we expected, we were able to affect expression by manipulating DNA sequence alone. We also identified some important principles that will help us in our future experiments. We have shown that both location and polarity of the DNA sequence can affect expression levels. We have also shown that nucleosome positioning sequences and repelling sequences have the ability to alter transgene expression from our extrachromosomal arrays. However, our understanding is still incomplete.

Our initial strategy was to attempt to form well positioned nucleosomes upstream and downstream of our enhancer leaving the enhancer to interact with the requisite transcription factors in the linker DNA. However, as mentioned previously nearly all of our experiments actually increased the silencing effect. However, our observations that location and polarity of the DNA sequence can affect expression mean that it will be important to try out all combinations for a particular DNA sequence to more fully understand its effect on the system. Developing a protocol to assay the locations of the nucleosomes on our reporter construct will also be very important.

These putative nucleosome positioning and repelling DNA sequences can have a greater effect when used together rather than singly. This suggests that we may get better results if we mix these elements in different combinations in our reporter construct pBYU1. As mentioned previously, poly-dA:dT tracts form an asymmetric barrier to nucleosome movement in mouse,

human, and yeast chromatin. For this reason we chose to test multiple combinations of poly-dA:dT tracts. (*e.g.* pBYU41 and pBYU45). If nucleosome positioning sequences like the 601 and Trifonov sequences are positioning nucleosomes but are doing it asymmetrically, it would also be worth investigating whether moving them a little further away from the enhancer will improve expression. It is possible if the nucleosome is too close to any transcription factor binding sites, it could interfere with binding. Indeed the 90bp *unc-54* enhancer is composed of multiple signals that together confer tissue specificity for *unc-54*. One of these signals, designated site III, is on the 3' end of the enhancer and is the strongest of the 4 sites [18]. This could explain why some of our inserts had stronger effects when located at the *Nhe*I site. However, complete transgene silencing from our extrachromosomal arrays is likely due to the influences of several factors including chromatin remodelers. We have yet to identify these factors. A forward genetic screen in the integrated pBYU1 lines we have created should help identify some of those factors.

While generating transgenic worm lines in *C. elegans* is easier than it is for some other model organisms, it is still a labor intensive process. Since sequence location, polarity, and the combinations in which they are used can all affect expression patterns, many microinjections will need to be performed to test them all. However, it might be possible to test multiple constructs in a single injection. It has been shown that plasmids are assembled into extrachromosomal arrays proportional to the ratio of the plasmids in the injection mix and that expression levels from an active reporter are proportional to the copy number [34]. However, we don't know what the minimum detectable level of GFP expression is from our reporter. Since pBYU9 had no detectable GFP body wall expression, it should be possible to use different ratios of pBYU1 and pBYU9 to identify the minimum amount of active reporter necessary in order to detect it. However, using this approach would not help identify which sequences contribute to silencing in

their respective constructs. This approach should however allow us to identify activating sequences more rapidly.

5.2 CONSIDERATIONS FOR WORKING WITH CRISPR/CAS9

The area of CRISPR/Cas9 technology is developing rapidly. It is an efficient way to perform genome editing *in vivo.* Expressing both CRISPR and Cas9 from expression vectors remains the most common practice, but the ability to make multiple edits simultaneously with the purified ribonucleoprotein without having to screen for the retention of the expression vectors makes the later method very attractive. However, working with the purified ribonucleoprotein is more difficult than working with expression vectors in *C. elegans* because of the high levels of salt required to maintain Cas9 solubility. CRISPR/Cas9 is a very useful technique for genome editing. While others have successfully used CRISPR/Cas9 expression vectors followed by positive and negative selection to integrate extrachromosomal arrays. No one has successfully integrated an extrachromosomal array using a purified ribonucleoprotein approach. Although we have been unsuccessful to date, we remain confident that additional microinjections will yield positive results. However, it is still possible to integrate extrachromosomal arrays using CRISPR/Cas9 expression vectors microinjected into parent strains that have the extrachromosomal array to be integrated [34]. Also, even if that fails there are still more traditional methods like gamma irradiation that are quite effective.

5.3 INVITROSOME REPRODUCIBILITY AND SEQUENCING COVERAGE.

A major reason for conducting *in vitro* nucleosome reconstitution experiments is to determine the effects of the underlying DNA sequence on nucleosome formation. It has been implied previously that nucleosome reconstitutions are stochastic and not very reproducible. We have

shown that using our reaction conditions nucleosome reconstitution experiments are highly reproducible. Furthermore, we show that the most dramatic gains in Pearson's coefficients are made when sequence coverage is maintained at a minimum depth of 35X. Therefore, we recommend that individuals conducting *In vitro* nucleosome reconstitution experiments maintain a minimum of 35X sequencing coverage.

5.4 MENTORING AND TRAINING YOUNG SCIENTISTS

An important part of research and science is communicating it with others. Experienced researchers are needed to mentor, train, and inspire the next generation of researchers. We need to improve our techniques so the rising generation of scientists will be prepared for the future. Towards that end, we have devised a scaffolding pedagogical method for transforming undergraduate students in to independent thinkers and researchers. Application of our method by multiple instructors showed increased student ability and confidence to plan and carry out independent research. Our data also suggests that this approach is not instructor specific but is generally applicable.

References

1. Luger, K., et al., *Crystal structure of the nucleosome core particle at 2.8 A resolution.* Nature, 1997. **389**(6648): p. 251-60.
2. Radman-Livaja, M. and O.J. Rando, *Nucleosome positioning: how is it established, and why does it matter?* Dev Biol, 2010. **339**(2): p. 258-66.
3. Tyagi, M., et al., *Chromatin remodelers: We are the drivers!!* Nucleus, 2016. **7**(4): p. 388-404.
4. Simcikova, M., et al., *Towards effective non-viral gene delivery vector.* Biotechnol Genet Eng Rev, 2015. **31**(1-2): p. 82-107.
5. Kazuki, Y. and M. Oshimura, *Human artificial chromosomes for gene delivery and the development of animal models.* Mol Ther, 2011. **19**(9): p. 1591-601.
6. Challita, P.M. and D.B. Kohn, *Lack of expression from a retroviral vector after transduction of murine hematopoietic stem cells is associated with methylation in vivo.* Proc Natl Acad Sci U S A, 1994. **91**(7): p. 2567-71.
7. Klug, C.A., S. Cheshier, and I.L. Weissman, *Inactivation of a GFP retrovirus occurs at multiple levels in long-term repopulating stem cells and their differentiated progeny.* Blood, 2000. **96**(3): p. 894-901.
8. Dietz, W.M., et al., *Minicircle DNA is superior to plasmid DNA in eliciting antigen-specific CD8+ T-cell responses.* Mol Ther, 2013. **21**(8): p. 1526-35.
9. Trifonov, E.N., *Sequence-dependent deformational anisotropy of chromatin DNA.* Nucleic Acids Res, 1980. **8**(17): p. 4041-53.
10. Thastrom, A., et al., *Sequence motifs and free energies of selected natural and non-natural nucleosome positioning DNA sequences.* J Mol Biol, 1999. **288**(2): p. 213-29.
11. Kunkel, G.R. and H.G. Martinson, *Nucleosomes will not form on double-stranded RNa or over poly(dA).poly(dT) tracts in recombinant DNA.* Nucleic Acids Res, 1981. **9**(24): p. 6869-88.
12. Gabdank, I., D. Barash, and E.N. Trifonov, *Nucleosome DNA bendability matrix (C. elegans).* J Biomol Struct Dyn, 2009. **26**(4): p. 403-11.
13. Johnson, S.M., et al., *Flexibility and constraint in the nucleosome core landscape of Caenorhabditis elegans chromatin.* Genome Res, 2006. **16**(12): p. 1505-16.
14. Mello, C.C., et al., *Efficient gene transfer in C.elegans: extrachromosomal maintenance and integration of transforming sequences.* EMBO J, 1991. **10**(12): p. 3959-70.
15. Stinchcomb, D.T., et al., *Extrachromosomal DNA transformation of Caenorhabditis elegans.* Mol Cell Biol, 1985. **5**(12): p. 3484-96.
16. Riddle DL, B.T., Meyer BJ, et al., *C. elegans II*, in *Section II The Organization, Structure, and Function of Muscle*, B.T. Riddle DL, Meyer BJ, Priess JR, Editor. 1997, Cold Spring Harbor Laboratory Press: Cold Spring Harbor, NY.

17. Okkema, P.G., et al., *Sequence requirements for myosin gene expression and regulation in Caenorhabditis elegans.* Genetics, 1993. **135**(2): p. 385-404.

18. Jantsch-Plunger, V. and A. Fire, *Combinatorial structure of a body muscle-specific transcriptional enhancer in Caenorhabditis elegans.* J Biol Chem, 1994. **269**(43): p. 27021-8.

19. Jin, Y., *Transformation*, in *C. elgans A Practical Approach*, I.A. Hope, Editor. 1999, Oxford University Press: Oxford, NY. p. 70-96.

20. Lowary, P.T. and J. Widom, *New DNA sequence rules for high affinity binding to histone octamer and sequence-directed nucleosome positioning.* J Mol Biol, 1998. **276**(1): p. 19-42.

21. Gracey, L.E., et al., *An in vitro-identified high-affinity nucleosome-positioning signal is capable of transiently positioning a nucleosome in vivo.* Epigenetics Chromatin, 2010. **3**(1): p. 13.

22. Kornberg, R.D. and L. Stryer, *Statistical distributions of nucleosomes: nonrandom locations by a stochastic mechanism.* Nucleic Acids Res, 1988. **16**(14A): p. 6677-90.

23. de Boer, C.G. and T.R. Hughes, *Poly-dA:dT tracts form an in vivo nucleosomal turnstile.* PLoS One, 2014. **9**(10): p. e110479.

24. Barrangou, R., et al., *CRISPR provides acquired resistance against viruses in prokaryotes.* Science, 2007. **315**(5819): p. 1709-12.

25. Jansen, R., et al., *Identification of genes that are associated with DNA repeats in prokaryotes.* Mol Microbiol, 2002. **43**(6): p. 1565-75.

26. Mojica, F.J., et al., *Short motif sequences determine the targets of the prokaryotic CRISPR defence system.* Microbiology, 2009. **155**(3): p. 733-740.

27. Driscoll, M., et al., *Genetic and molecular analysis of a Caenorhabditis elegans beta-tubulin that conveys benzimidazole sensitivity.* J Cell Biol, 1989. **109**(6 Pt 1): p. 2993-3003.

28. Zhao, P., et al., *Oligonucleotide-based targeted gene editing in C. elegans via the CRISPR/Cas9 system.* Cell Res, 2014. **24**(2): p. 247-50.

29. Kent, W.J., et al., *The human genome browser at UCSC.* Genome Res, 2002. **12**(6): p. 996-1006.

30. Naito, Y., et al., *CRISPRdirect: software for designing CRISPR/Cas guide RNA with reduced off-target sites.* Bioinformatics, 2015. **31**(7): p. 1120-3.

31. Chen, C., L.A. Fenk, and M. de Bono, *Efficient genome editing in Caenorhabditis elegans by CRISPR-targeted homologous recombination.* Nucleic Acids Res, 2013. **41**(20): p. e193.

32. Horlbeck, M.A., et al., *Nucleosomes impede Cas9 access to DNA in vivo and in vitro.* Elife, 2016. **5**.

33. Hope, I.A., *C. elegans : a practical approach*. The practical approach series. 1999, Oxford ; New York: Oxford University Press. xxi, 281 p.

34. Yoshina, S., et al., *Locus-specific integration of extrachromosomal transgenes in C. elegans with the CRISPR/Cas9 system.* Biochemistry and Biophysics Reports, 2016. **5**: p. 70-76.

35. Lu, Q., L.L. Wallrath, and S.C. Elgin, *Nucleosome positioning and gene regulation.* J Cell Biochem, 1994. **55**(1): p. 83-92.

36. Struhl, K. and E. Segal, *Determinants of nucleosome positioning.* Nat Struct Mol Biol, 2013. **20**(3): p. 267-73.

37. Valouev, A., et al., *A high-resolution, nucleosome position map of C. elegans reveals a lack of universal sequence-dictated positioning.* Genome Res, 2008. **18**(7): p. 1051-63.

38. Valouev, A., et al., *Determinants of nucleosome organization in primary human cells.* Nature, 2011.

39. Kaplan, N., et al., *The DNA-encoded nucleosome organization of a eukaryotic genome.* Nature, 2009. **458**(7236): p. 362-6.

40. Zhang, Y., et al., *Intrinsic histone-DNA interactions are not the major determinant of nucleosome positions in vivo.* Nat Struct Mol Biol, 2009. **16**(8): p. 847-52.

41. Luger, K., T.J. Rechsteiner, and T.J. Richmond, *Expression and purification of recombinant histones and nucleosome reconstitution.* Methods Mol Biol, 1999. **119**: p. 1-16.

42. Simpson, R.T. and D.W. Stafford, *Structural features of a phased nucleosome core particle.* Proc Natl Acad Sci U S A, 1983. **80**(1): p. 51-5.

43. Flaus, A. and T.J. Richmond, *Positioning and stability of nucleosomes on MMTV 3'LTR sequences.* J Mol Biol, 1998. **275**(3): p. 427-41.

44. Johnson, S.M., *Painting a perspective on the landscape of nucleosome positioning.* J Biomol Struct Dyn, 2010. **27**(6): p. 795-802.

45. Mavrich, T.N., et al., *A barrier nucleosome model for statistical positioning of nucleosomes throughout the yeast genome.* Genome Res, 2008. **18**(7): p. 1073-83.

46. Kornberg, R., *The location of nucleosomes in chromatin: specific or statistical.* Nature, 1981. **292**(5824): p. 579-80.

47. Locke, G., et al., *Global remodeling of nucleosome positions in C. elegans.* BMC Genomics, 2013. **14**: p. 284.

48. Meersseman, G., S. Pennings, and E.M. Bradbury, *Mobile nucleosomes--a general behavior.* EMBO J, 1992. **11**(8): p. 2951-9.

49. Flaus, A. and T.J. Richmond, *Base-pair resolution mapping of nucleosomes in vitro.* Methods Mol Biol, 1999. **119**: p. 45-60.

50. Brogaard, K., et al., *A map of nucleosome positions in yeast at base-pair resolution.* Nature, 2012. **486**(7404): p. 496-501.

51. Modell, H.I. and J.A. Michael, *Promoting Active Learning in the Life-Science Classroom - Defining the Issues.* Promoting Active Learning in the Life Science Classroom, 1993. **701**: p. 1-7.

52. Cox, D.D. and L.V. Davis, *The context of biological education: the case for change [by] Donald D. Cox and Lary V. Davis*. Commission on Undergraduate Education in the Biological Sciences Publication,. 1972, Washington. ix, 147 p.

53. McComas, W.F., *The nature of science in science education : rationales and strategies*. Science & technology education library. 1998, Dordrecht ; Boston: Kluwer Academic Publishers. xxi, 365 p.

54. Boyer Commission on Educating Undergraduates in the Research University., *Reinventing Undergraduate Education a blueprint for America's Research Universities*. 1998, Stony Brook, NY?: State University of New York at Stony Brook for the Carnegie Foundation for the Advancement of Teaching.

55. National Research Council (USA). Committee on Development of an Addendum to the National Science Education Standards on Scientific Inquiry, *Inquiry and the National science education standards : a guide for teaching and learning*. 2000, Washington, D.C.: National Academy Press. xviii, 202 s.

56. National Research Council (U.S.). Committee on Undergraduate Biology Education to Prepare Research Scientists for the 21st Century., *Bio 2010 : transforming undergraduate education for future research biologists*. 2003, Washington, D.C.: National Academies Press. xv, 191 p.

57. Woodin, T., V.C. Carter, and L. Fletcher, *Vision and Change in Biology Undergraduate Education, A Call for Action-Initial Responses.* Cbe-Life Sciences Education, 2010. **9**(2): p. 71-73.

58. Weber, K.S., J.L. Jensen, and S.M. Johnson, *Anticipation of Personal Genomics Data Enhances Interest and Learning Environment in Genomics and Molecular Biology Undergraduate Courses.* PLoS One, 2015. **10**(8): p. e0133486.

59. DiPasquale, D.M., C.L. Mason, and F.W. Kolkhorst, *Exercise in inquiry.* Journal of College Science Teaching, 2003. **32**(6): p. 388-393.

60. Handelsman, J., et al., *Education. Scientific teaching.* Science, 2004. **304**(5670): p. 521-2.

61. Beed, P.L., E.M. Hawkins, and C.M. Roller, *Moving Learners toward Independence - the Power of Scaffolded Instruction.* Reading Teacher, 1991. **44**(9): p. 648-655.

62. Bruner, J.S., *The ontogenesis of speech acts.* Journal of Child Language, 1975. **2**(01): p. 1-19.

63. Vygotsky, L.S., *Mind in Society: The Development of Higher Psychological Processes*. 1978, Cambridge, MA: Harvard University Press.

64. Invitrogen, *PureLink™ Genomic DNA Kits, For purification of genomic DNA, User Manual*. 2007, Life Technologies Corporation.

65. Invitrogen, *PureLink® Quick Gel Extraction and PCR Purification Combo Kit, For purification of DNA fragments from agarose gels and rapid, efficient purification of PCR products, User Manual*. 2010, Life Technologies Corporation.

66. Qiagen, *QIAprep® Miniprep Handbook*. 2002, Qiagen.

67. Finnzymes, *Phusion™ Site-Directed Mutagenesis Kit, Instructions*. 2008, New England Biolabs, Inc.

68. Invitrogen, *Zero Blunt® TOPO® PCR Cloning Kit, Five-minute cloning of blunt-end PCR products, User Manual*. 1999, Invitrogen Corporation.

69. Ambion, *TRIzol® Reagent, Instructions*. 2007, Life Technologies Corporation.

70. Ambion, *Instruction Manual, NorthernMax® -Gly, Glyoxal-Based System for Northern Blots*. 2008, Applied Biosystems.

71. Johnson, S.M., S.Y. Lin, and F.J. Slack, *The time of appearance of the C. elegans let-7 microRNA is transcriptionally controlled utilizing a temporal regulatory element in its promoter.* Dev Biol, 2003. **259**(2): p. 364-79.

72. Ramaswamy, G., *Gopi's nuclear extract protocol*, S.M. Johnson, Editor. 2003. p. 1.

Appendix C

Table C-1 Nucleosome Positioning Inserts

Element name	Sequence 5' > 3'	Length (bp)
***unc-54* enhancer**	gtcttcttctaaattcccataaaatcccgaaactccttccctctatcttctttttcttctcgttttcaaatgtt tctctctATCCCATTCTCTCATCAATTGAGTGGGATGAGGCTAT CTCTGCCTCTCTTCTGAATCTCTGAACCATCTTACATTACA CTGTGGATGACGagccccacaggctcccttgcatcagatactgccattggggatggc aaagaagagagaaggtattgtgaggatatattttctaagaaaaaacgtttgaagaaaagaagatg aagaa	286
***unc-54* minimal enhancer**	ATCCCATTCTCTCATCAATTGAGTGGGATGAGGCTATCTCT GCCTCTCTTCTGAATCTCTGAACCATCTTACATTACACTGT GGATGACG	90
Trifonov	GGAGATCCCTCGAAAATTTCCTTTCCGGAAATTCCCCGGG AATTCTCGAAGATTCTCCGAGGATCCCCGAAGATCCTTAG GGATTCTTTAAGGATTCTCGGAGATTCCCGAGGGATCTTC GGGAATCTCCGGAGAATTTCCGGAAAGGAAATTCCCCGGA GATTTCC	167
601	ACTGGAGAATCCCGGTCTGCAGGCCGCTCAATTGGTCGTA GACAGCTCTAGCACCGCTTAAACGCACGTACGCGCTGTCC CCCGCGTTTTAACCGCCAAGGGGATTACTCCCTAGTCTCCA GGCACGTGTCAGATATATACATCCTGT	148
601_5bp	tgcagACTGGAGAATCCCGGTCTGCAGGCCGCTCAATTGGTC GTAGACAGCTCTAGCACCGCTTAAACGCACGTACGCGCTG TCCCCCGCGTTTTAACCGCCAAGGGGATTACTCCCTAGTCT CCAGGCACGTGTCAGATATATACATCCTGT	153
601_3bp	tgcACTGGAGAATCCCGGTCTGCAGGCCGCTCAATTGGTCG TAGACAGCTCTAGCACCGCTTAAACGCACGTACGCGCTGT CCCCCGCGTTTTAACCGCCAAGGGGATTACTCCCTAGTCTC CAGGCACGTGTCAGATATATACATCCTGT	151
20bp_pSJ322	ACAGCAAAAAACATTAAAAA	20
50bp_pSJ322	ATAATTTTTCTAGTATAATTCTCATGAGAAACAGCAAAAA ACATTAAAAA	50
70bp_pSJ322	TATTTCAAAAATAATTTTTCTAGTATAATTCTCATGAGAAA CaGCAAAAAACATTAAAAACCAATAATTT	70
polyA(20)	AAAAAAAAAAAAAAAAAAAA	20
polyA(5)	AAAAA	5
polyA(5)_c_A(5)	AAAAAcAAAAA	11
lac operator	ACCTTAACACTCGCCTATTGTTAA	24

Table C-2 Nucleosome Positioning Primers

HBSNN1	AGCTTGGATCCAGGCCTGCGGCCG
HBSNN2	CTAGCGGCCGCAGGCCTGGATCCA
CK-SJ-3	GGAAGCTTGGAGATCCCTCGAAAATTTCCTCGGAAA TTCCCCGGGAATTCTCG
CK-SJ-4	CGCTAGCGGAAATCTCCGGGGAATTTCCTTTCCGGA AATTCTCCGGAGATTCCCG
CK-SJ-5	ACTAGCAAGCTTGGAGATCCCTCGAAAATTTC
CK-SJ-6	GTGCTCAAGCTTGGAAATCTCCGGGGAATTTC
CK-SJ-7	ACTAGCGCTAGCGGAGATCCCTCGAAAATTTC
CK-SJ-8	GTGCTCGCTAGCGGAAATCTCCGGGGAATTTC
SD-Triv for	/5PHOS/CTCCGGAGAATTTcCGGAAAGGAAATTCC
SD-Triv Rev	/PHOS/ATTCCCGAAGATCCCTCGGGAATC
HindIII601 R	AAGCTTACTGGAGAATCCCGGTCTGC
HindIII601 F	AAGCTTACAGGATGTATATATCTGACACG
NheI601 R	GCTAGCACTGGAGAATCCCGGTCTGC
NheI601 F	GCTAGCACAGGATGTATATATCTGACACG
N5bp601F	gctagcTGCAGactggagaatcccggtctg
N5bp601R	gctagcTGCAGacaggatgtatatatctgacacg
H5bp601F	aagctTCCGGAactggagaatcccggtctgc
H5bp601R	aagctTCCGGaacaggatgtatatatctgacacg
N3bp601F	gctagcTGCactggagaatcccggtctgc
N3bp601R	gctagcTGCacaggatgtatatatctgacacg
H3bp601F	aagctTCCGactggagaatcccggtctgc
H3bp601R	aagctTCCGacaggatgtatatatctgacacg
NheI_5t	/5Phos/CTAGCTTTTTG
NheI_5a	/5Phos/CTAGCAAAAAG
HindIII_5t	/5Phos/AGCTTTTTTTA
HindIII_5a	/5Phos/AGCTTAAAAAA
N_R_polyT	/5Phos/CTAGCAAAAAAAAAAAAAAAAAAG
N_F_polyT	/5Phos/CTAGCTTTTTTTTTTTTTTTTTTG
H_R_polyT	/5Phos/AGCTTAAAAAAAAAAAAAAAAAAA
H_F_polyT	/5Phos/AGCTTTTTTTTTTTTTTTTTTTA
20R_N_pSJ322	/5Phos/CTAGCTTTTTAATGTTTTTTGCTGTG
20F_N_pSJS32	/5Phos/CTAGCACAGCAAAAAACATTAAAAAG
20R_H_pSJ322	/5Phos/AGCTTTTTTTAATGTTTTTTGCTGTA
20F_H_pSJ322	/5Phos/AGCTTACAGCAAAAAACATTAAAAAA
50R_N_pSJ322	GCTAGCTTTTTAATGTTTTTTGCTGTTTC
50F_N_pSJ322	GCTAGCATAATTTTTCTAGTATAATTCTC

50R_H_pSJ322	AAGCTTTTTTTAATGTTTTTTGCTGTTTC
50F_H_pSJ322	AAGCTTATAATTTTTCTAGTATAATTCTC
70R_N_pSJ322	GCTAGCAAATTATTGGTTTTTAATGTTTTTTGCTG
70F_N_pSJ322	GCTAGCTATTTCAAAAATAATTTTTCTAGTATAATTC
70R_H_pSJ322	AAGCTTAAATTATTGGTTTTTAATGTTTTTTGCTG
70F_H_pSJ322	AAGCTTTATTTCAAAAATAATTTTTCTAGTATAATTC
oSMJ021	/5phos/agctttttgttttta
oSMJ022	/5phos/agcttaaaaacaaaa
oSMJ023	/5phos/ctagctttttgttttg
oSMJ024	/5phos/ctagcaaaaacaaaaag
oSMJ034	/phos/ agcttTGGAATTGTGAGCGGATAACAATTa
oSMJ035	/phos/ aACCTTAACACTCGCCTATTGTTAAttcga
oSMJ036	/phos/ ctagcTGGAATTGTGAGCGGATAACAATTg
oSMJ037	/phos/ gACCTTAACACTCGCCTATTGTTAAcgatc
oSMJ038	/phos/ agcttAATTGTTATCCGCTCACAATTCCAa
oSMJ039	/phos/ ctagcAATTGTTATCCGCTCACAATTCCAg

Table C-3 CRISPR Primers

oSMJ008	tgtaatacgactcactataggCTTAAATTTATTTGCACTACgttttagagctagaaatagc
oSMJ009	aaaaaagcaccgactcggtgccact
oSMJ010	tgtaatacgactcactataggGACTCTGAAATAACGTTGCGgttttagagctagaaatagc
oSMJ011	GACTCTGAAATAACGTTGCG
oSMJ012	AAGGTGAAAGGTGTAAATAATC
oSMJ013	atggacaaactattgctgatgcg
oSMJ014	TCAATGGAAGGTCGTAAATTTGG
oSMJ015	tctcggaagacttaagtggtgcc
oSMJ016	TGGTGACTCCGGACATTGTAACGG
oSMJ017	tgtaatacgactcactataggGAGCATGGGATCCAGCCTGAgttttagagctagaaatagc
oSMJ018	tgtaatacgactcactataggGTTCTCGTCGATCTTGAGCCgttttagagctagaaatagc
oSMJ019	GCTGCAGCCGCATACAATCG
oSMJ020	ATCTGCTGGGTAAGCTCGGC
oSMJ025	CGTCCCCTGACCTTCTTCAG
oSMJ026	CAGTCAGGTAACGTCCGTGT
oSMJ027	tgtaatacgactcactataggAAGTGATATCCGATGAGCATgttttagagctagaaatagc
oSMJ028	atacgccggtggttttgtctc
oSMJ029	gacaaaaccaccggcgtatac
oSMJ030	tgtaatacgactcactataggCCAGCCTGATGGAACTTATAgttttagagctagaaatagc
oSMJ031	tgtaatacgactcactataggGATTTGATTACCACATTGTCgttttagagctagaaatagc
oSMJ032	CATTAGCCTCATTATAGTAGACATTG

oSMJ033	CGAACACAAAGTTATCTGGAC

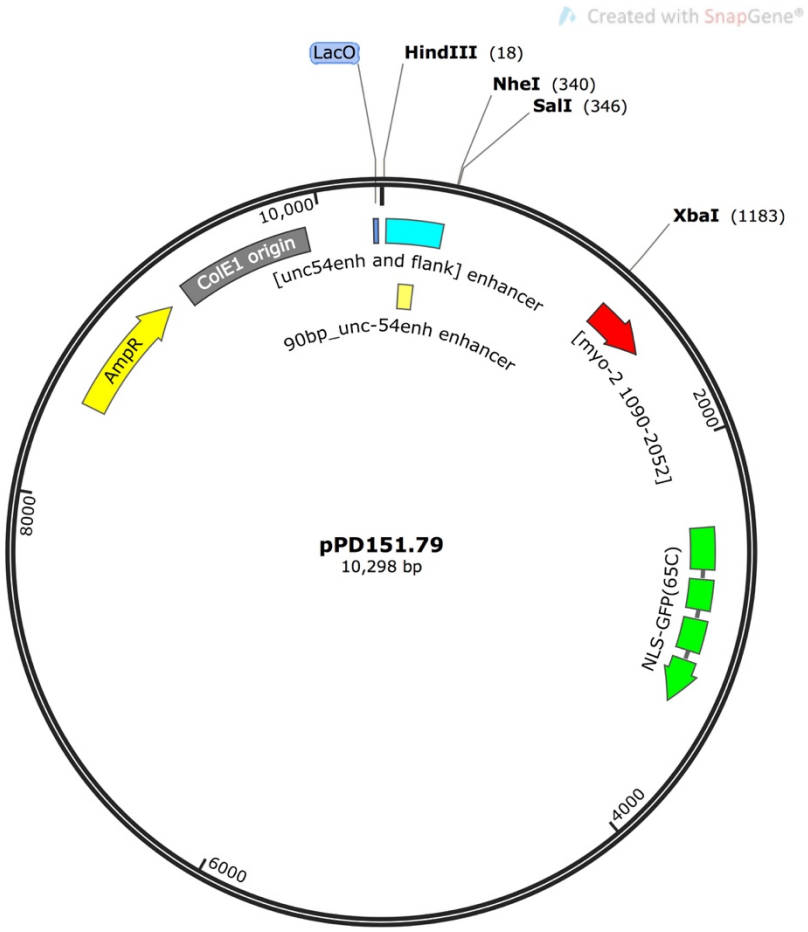
Table C-4 sgRNAs

sgRNA Guide	Forward Primer	CRISPRdirect Sequence 5' . 3'	Target
sg001	oSMJ008	CTTAAATTTATTTGCACTAC	GFP
sg002	oSMJ010	GACTCTGAAATAACGTTGCG	*C. elegans* Chr2, 6,665,000
sg003	oSMJ017	GAGCATGGGATCCAGCCTGA	*ben-1*, exon-1
sg004	oSMJ018	GTTCTCGTCGATCTTGAGCC	*ben-1*, exon-2
sg005	oSMJ027	AAGTGATATCCGATGAGCAT	*ben-1*, exon-1 [31]
sg006	oSMJ030	CCAGCCTGATGGAACTTATA	*ben-1*, exon-1
sg007	oSMJ031	GATTTGATTACCACATTGTC	*ben-1,* exon-1
All forward primers used oSMJ009 as the reverse primer (px330)			

Table C-5 Injection Mix Table

Injection Mix	**Plasmids**	**Concentrations**
pBYU1, pCR4_TOPO_ben-1	pBYU1 pCR4_TOPO_ben-1	93.5 ng/uL 6.25 ng/uL
pBYU46, pSMJ001	pBYU46 pSMJ001 pRF4	30 ng/uL 30 ng/uL 40ng /uL
All other plasmids were injected at 50 ng/uL with pRF4 (rol-6) as a co-injection marker at 50 ng/uL		

Figure C-1 pPD151.79



ATGACCATGATTACGCCAAGCTTGTCTTCTTCTAAATTCCCATAAAATCCCGAAACT
CCTTCCCTCTATCTTCTTTTTCTTCTCGTTTTCAAATGTTTCTCTCTATCCCATTCTCTC
ATCAATTGAGTGGGATGAGGCTATCTCTGCCTCTCTTCTGAATCTCTGAACCATCTTA
CATTACACTGTGGATGACGAGCCCCACAGGCTCCCTTGCATCAGATACTGCCATTGG
GGATGGCAAAGAAGAGAGAAGGTATTGTGAGGATATATTTTTCTAAGAAAAAACGT
TTGAAGAAAAGAAGATGAAGAAgatccccgggattggccaaaggacccttggCTAGCGTCGACTCCG
GCGGAGAAAAGCGGCTTGGATGAAAGGACGATCAAATAAAAGCAGAATCTCGGTA
CGCCGGCGGCGAGTTAAGGATATAAGAGTTAAGGGACTCCACTGAATCGGGAAGAA
AGGACGGCGATGCAGGTAAGCAAATGCAAAAGATACAAATACGGATTCCAGTAATG
GATGACGTGACAGAAGACCGCCAGGACGCTGCGAGATAGAAACACTGGACTATTGG
TCTCCCGGCCAAAAAGAAATAAATAATAATGTAAACTACTCCCGGGCGAGAGAACC
GCGGTCAGTACAAGATTAAGGAATGGTTATTCTCGCCCGGGGAAATGCGGTCGGCC
TCTCGGCCAGAAACCGATGGCGGAAGAAAGAAACTGGATCTGCTCCGAGCAGGGAA
ACCCCCGCGACTGGAAACCACGGGGAGGCCCCCGGGAGGCGAATGTATACCGTCCT
CATATTGAAGCTCCTGGATAGGGAAATATCGGGGCCGGCCGGATGGGCATCGGTGC
TCCGGGTATATCAGACCGGGAGATAGGAGTCGGCGGCTACCCGTGCATGAGAGGAG
GGACGCATGACCTAGGATCAACTCGACGGCCGCCTTATTAAATTGGTCCCAGGGCA

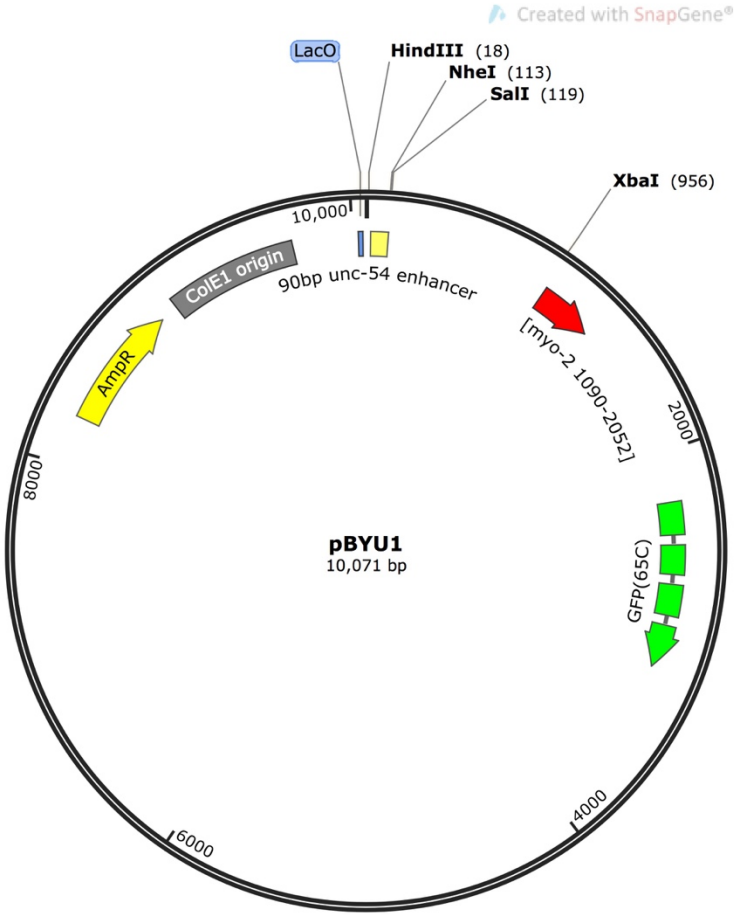
GTGGAAGGGCGAGCATGCTGGATGGTATAGCTAGAGGGACAGAAAACTTTCGAGAC
CTCATGCCAGAAAAGAATAAGCGACGAACGCTGGACGACGACTTACAGCGGGAAA
CGATCAAATACCATCCAGACTACTTTTTTGTATCTTGAGGAAGTAATAACGAATCAG
GTGCAGGCACACTAACTAAGTTGGTTGTAGAGTTGCGGTTGAAATCTAGAGGATCG
AGGCATTTGAATTGGGGGTGGTGGACAGTAACTGTCTGTAATAATAATTACTCCTGA
CCAGGTTGCAATTCGAGTTTTGATAAGCATAATTATACCTTGTACATTGTGGGTTTTG
TGCTGTGGACGTTTTATTGTGGACATCCCCATAAGCTACAAGAAACCAAAAATGAAA
TTAAAAGTATTGAAAAACGTCGTAACATTTTATATCTGAGTAGTATCCTTTGCTTTAA
ATGTCCATAAAAATAATTTTATAATCAATAAAACAACGTTTGTAAATCAACTGAGTT
TACAAGTAGAGACATTGAGGGATACTTTCACTATGCTAAAGTGAATAATCGACCAA
ATAATAACTCACTTTGGTATTTATTCCTGTCTTATAATGTTATGTATGAATTAAATTC
ATATGCATATGGCTCACTCTGACAAAAAAAAATAATCTTCCAGATCAATATTGACTA
CCGATGCGGGTGGTCTTTTGCTTTGAATTCTGCTGAACTTTACACCCCGAACAGCAA
TGTGTGCTTCAGCTAAAAAAAAGTAAGTGTGTTAATCAGTCCCCCCGATTCTTCATTT
TTTGCCCCTCTCTCCCGTTTCGTCGGCAAAAGAAGAGAAAATAAAGATAAGTCTCAA
GATAGGTTGGTAATCGCTAAAGTGGTTGTGTGGATAAGAGTAGCAAAATGGCAGGA
AGAGCACTTTGCGCGCACACACTGTACTCATTGTTCTGGATAAAATTCTCTCGTTGTT
TGCCGTCGGATGTCTGCCTCTCTGCCATTGAGCCGGCTTCTTCACTATCTTTAGTTAA
CCTAAAATGCCGTTTCTTTTCTCGTATCCCACTATCCGTTGAGGTTCTCTGCTCTCTTC
GCTCCCTTACCGCCAGCGAGCAACTATCCGTGGGGGCGCCTTGCTCGGAAGATGGG
GGGGAAGAAAGAAGATTTTTGCTATTTGCACTTGAGAAAGAGACTTTTCCTGCGTCG
ATGGTTAGAGAACAGTGTGCAGACACTTTTCAGCTACCTAGATACATGGATATCCCC
GCCTCCCAATCCACCCACCCAGGGAAAAAGAAGGGCTCGCCGAAAAATCAAAGTTA
TCTCCAGGCTCGCGCATCCCACCGAGCGGTTGACTTCTCTCCACCACTTTTCATTTTA
ACCCTCGGGGTACGGGATTGGCCAAAGGACCCAAAGGTATGTTTCGAATGATACTA
ACATAACATAGAACATTTTCAGGAGGACCCTTGGAGGGTACCGAGCTCAGAAAAAA
TGACTGCTCCAAAGAAGAAGCGTAAGGTACCGGTAGAAAAAATGAGTAAAGGAGA
AGAACTTTTCACTGGAGTTGTCCCAATTCTTGTTGAATTAGATGGTGATGTTAATGG
GCACAAATTTTCTGTCAGTGGAGAGGGTGAAGGTGATGCAACATACGGAAAACTTA
CCCTTAAATTTATTTGCACTACTGGAAAACTACCTGTTCCATGGgtaagtttaaacatatatatacta
actaaccctgattatttaaattttcagCCAACACTTGTCACTACTTTCTgTTATGGTGTTCAATGCTTcTC
gAGATACCCAGATCATATGAAACgGCATGACTTTTTCAAGAGTGCCATGCCCGAAGG
TTATGTACAGGAAAGAACTATATTTTTCAAAGATGACGGGAACTACAAGACACgtaagt
ttaaacagttcggtactaactaaccatacatatttaaattttcagGTGCTGAAGTCAAGTTTGAAGGTGATACCCT
TGTTAATAGAATCGAGTTAAAAGGTATTGATTTTAAAGAAGATGGAAACATTCTTGG
ACACAAATTGGAATACAACTATAACTCACACAATGTATACATCATGGCAGACAAAC
AAAAGAATGGAATCAAAGTTgtaagtttaaacatgattttactaactaactaatctgatttaaattttcagAACTTCAA
AATTAGACACAACATTGAAGATGGAAGCGTTCAACTAGCAGACCATTATCAACAAA
ATACTCCAATTGGCGATGGCCCTGTCCTTTTACCAGACAACCATTACCTGTCCACAC
AATCTGCCCTTTCGAAAGATCCCAACGAAAAGAGAGACCACATGGTCCTTCTTGAGT
TTGTAACAGCTGCTGGGATTACACATGGCATGGATGAACTATACAAATgcCCGGTGG
GTGAAGACCAGAAACAGCACCTCGAACTGAGCCGCGATATTGCCCAGCGTTTCAAC
GCGCTGTATGGCGAGATCGATCCCGTCGTTTTACAACGTCGTGACTGGGAAAACCCT
GGCGTTACCCAACTTAATCGCCTTGCAGCACATCCCCCTTTCGCCAGCTGGCGTAAT
AGCGAAGAGGCCCGCACCGATCGCCCTTCCCAACAGTTGCGaAGgtaagtttaaacagatccatac
taactaacttgttctgacataattttcagCtTGAATGGCGAATGGCGCTTTGCCTGGTTTCCGGCACCAG

AAGCGGTGCCGGAAAGCTGGCTGGAGTGCGATCTTCCTGAGGCCGATACTGTCGTC
GTCCCCTCAAACTGGCAGATGCACGGTTACGATGCGCCCATCTACACCAACGTAACC
TATCCCATTACGGTCAATCCGCCGTTTGTTCCCACGGAGAATCCGACGGGTTGTTAC
TCGCTCACATTTAATGTTGATGAAAGCTGGCTACAGGAAGGCCAGACGCGAATTATT
TTTGATGGCGTTAACTCGGCGTTTCATCTGTGGTGCAACGGGCGCTGGGTCGGTTAC
GGCCAGgtaagtttaattaagttgatactaactaacaaagatctgattaattttcagGACAGTCGTTTGCCGTCTGAAT
TTGACCTGAGCGCATTTTTACGCGCCGGAGAAAACCGCCTCGCGGTGATGGTGCTGC
GTTGGAGTGACGGCAGTTATCTGGAAGATCAGGATATGTGGCGGATGAGCGGCATT
TTCCGTGACGTCTCGTTGCTGCATAAACCGACTACACAAATCAGCGATTTCCATGTT
GCCACTCGCTTTAATGATGATTTCtccCGCGCTGTACTGGAGGCTGAAGTcCAGgtaagttta
aacaggatcttactaactaacatgctaacactgaattttcagATGTGCGGCGAGTTGCGTGACTACCTACGGGT
AACAGTTTCTTTATGGCAGGGTGAAACGCAGGTCGCCAGCGGCACCGCGCCTTTCGG
CGGTGAAATTATCGATGAGCGTGGTGGTTATGCCGATCGCGTCACACTACGTCTGAA
CGTCGAAAACCCGAAACTGTGGAGCGCCGAAATCCCGAATCTCTATCGTGCGGTGG
TTGAACTGCACACCGCCGACGGCACGCTGATTGAAGCAGAAGCCTGCGATGTCGGT
TTCCGCGAGGTGCGGATTGAAAATGGTCTGCTGCTGCTGAACGGCAAGCCGTTGCTG
ATTCGAGGCGTTAACCGTCACGAGCATCATCCTCTGCATGGTCAGGTCATGGATGAG
CAGACGATGGTGCAGGATgtaagtttaaactattcgttactaactaactttaaacatttaaattttcagATCCTGCTGA
TGAAGCAGAACAACTTTAACGCCGTGCGCTGTTCGCATTATCCGAACCATCCGCTGT
GGTACACGCTGTGCGACCGCTACGGCCTGTATGTGGTGGATGAAGCCAATATTGAA
ACCCACGGCATGGTGCCAATGAATCGTCTGACCGATGATCCGCGCTGGCTACCGGC
GATGAGCGAACGCGTAACGCGAATGGTGCAGCGCGATCGTAATCACCCGAGTGTGA
TCATCTGGTCGCTGGGGAATGgtaagtttaaacagttgaatactaactaacggagatctttgaaattttcagAATCAG
GCCACGGCGCTAATCACGACGCGCTGTATCGCTGGATCAAATCTGTCGATCCTTCCC
GCCCGGTGCAGTATGAAGGCGGCGGAGCCGACACCACGGCCACCGATATTATTTGC
CCGATGTACGCGCGCGTGGATGAAGACCAGCCCTTCCCGGCTGTGCCGAAATGGTC
CATCAAAAAATGGCTTTCGCTACCTGGAGAGACGCGCCCGCTGATtCTTTGCGAGgtaa
gtttaaacagaactctactaactaacacattagatcctaattttcagTACGCtCACGCGATGGGcAACAGTCTTGGC
GGTTTCGCTAAATACTGGCAGGCGTTTCGTCAGTATCCCCGTTTACAGGGCGGCTTC
GTCTGGGACTGGGTGGATCAGTCGCTGATTAAATATGATGAAAACGGCAACCCGTG
GTCGGCTTACGGCGGTGATTTTGGCGATACGCCGAACGATCGCCAGTTCTGTATGAA
CGGTCTGGTCTTTGCCGACCGCACGCCGCATCCAGCgtaagtttaaacaataacctactaactaacgtagat
aatttaaattttcagGCTGACGGAAGCAAAACACCAGCAGCAGTTTTTCCAGTTCCGTTTATCC
GGGCAAACCATCGAAGTGACCAGCGAATACCTGTTCCGTCATAGCGATAACGAGCT
CCTGCACTGGATGGTGGCGCTGGATGGTAAGCCGCTGGCAAGCGGTGAAGTGCCTC
TGGATGTCGCTCCACAAGGTAAACAGTTGATTGAACTGCCTGAACTACCGCAGCCGG
AGAGCGCCGGGCAACTCTGGCTCACAGTACGCGTAGTGCAACCGAACGCGACCGCA
TGGTCAGAAGCCGGGCACATCAGCGCaTGGCAGCAGTGGAGgtaagtttaaacaagatcctactaac
taactctacattgatgaattttcagACTGGCGGAAAACCTCAGTGTGACGCTCCCCGCCGCGTCCCAC
GCCATCCCGCATCTGACCACCAGCGAAATGGATTTTTGCATCGAGCTGGGTAATAAG
CGTTGGCAATTTAACCGCCAGTCAGGCTTTCTTTCACAGATGTGGATTGGCGATAAA
AAACAACTGCTGACGCCGCTGCGCGATCAGTTCACCCGTGCACCGCTGGATAACGA
CATTGGCGTAAGTGAAGCGACCCGCATTGACCCTAACGCCTGGGTCGgtaagtttaaacaaag
ttgtactaactaacgaagatcttgataattttcagAACGCTGGAAGGCGGCGGGCCATTACCAGGCCGAAG
CAGCGTTGTTGCAGTGCACGGCAGATACACTTGCTGATGCGGTGCTGATTACGACCG
CTCACGCGTGGCAGCATCAGGGGAAAACCTTATTTATCAGCCGGAAAACCTACCGG

ATTGATGGTAGTGGTCAAATGGCGATTACCGTTGATGTTGAAGTGGCGAGCGATACA
CCGCATCCGGCGCGGATTGGCCTGAACTGCCAGCTGGCGCAGGTAGCAGAGCGGGT
AAACTGGCTCGGATTAGGGCCGCAAGAAAACTATCCCGACCGCCTTACTGCCGCCT
GTTTTGACCGCTGGGATCTGCCATTGTCAGACATGTAgtaagtttaaacttgatagtactaactaacatgttt
catttaaattttcagTACCCCGTACGTCTTCCCGAGCGAAAACGGTCTGCGCTGCGGGACGCG
CGAATTGAATTATGGCCCACACCAGTGGCGCGGCGACTTCCAGTTCAACATCAGCCG
CTACAGTCAACAGCAACTGATGGAAACCAGCCATCGCCATCTGCTGCACGCGGAAG
AAGGCACATGGCTGAATATCGACGGTTTCCATATGGGGATTGGTGGCGACGACTCCT
GGAGCCCGTCAGTATCGGCGGAATTCCAACTGAGCGCCGGTCGCTACCATTACCAA
CTTGTCTGGTGTCAAAAATAATAGGGGCCGCTGTCATCAGAgtaagtttaaactgagttctactaacta
acgagtaatatttaaattttcagCATCTCGCGCCCGTGCCTCTGACTTCTAAGTCCAATTACTCTTCA
ACATCCCTACATGCTCTTTCTCCCTGTGCTCCCACCCCCTATTTTTGTTATTATCAAAA
AAACTTCTTCTTAATTTCTTTGTTTTTTAGCTTCTTTTAAGTCACCTCTAACAATGAAA
TTGTGTAGATTCAAAAATAGAATTAATTCGTAATAAAAAGTCGAAAAAAATTGTGCT
CCCTCCCCCCATTAATAATAATTCTATCCCAAAATCTACACAATGTTCTGTGTACACT
TCTTATGTTTTTTTTACTTCTGATAAATTTTTTTTGAAACATCATAGAAAAAACCGCA
CACAAAATACCTTATCATATGTTACGTTTCAGTTTATGACCGCAATTTTTATTTCTTC
GCACGTCTGGGCCTCTCATGACGTCAAATCATGCTCATCGTGAAAAAGTTTTGGAGT
ATTTTTGGAATTTTTCAATCAAGTGAAAGTTTATGAAATTAATTTTCCTGCTTTTGCT
TTTTGGGGGTTTCCCCTATTGTTTGTCAAGAGTTTCGAGGACGGCGTTTTTCTTGCTA
AAATCACAAGTATTGATGAGCACGATGCAAGAAAGATCGGAAGAAGGTTTGGGTTT
GAGGCTCAGTGGAAGGTGAGTAGAAGTTGATAATTTGAAAGTGGAGTAGTGTCTAT
GGGGTTTTTGCCTTAAATGACAGAATACATTCCCAATATACCAAACATAACTGTTTC
CTACTAGTCGGCCGTACGGGCCCTTTCGTCTCGCGCGTTTCGGTGATGACGGTGAAA
ACCTCTGACACATGCAGCTCCCGGAGACGGTCACAGCTTGTCTGTAAGCGGATGCCG
GGAGCAGACAAGCCCGTCAGGGCGCGTCAGCGGGTGTTGGCGGGTGTCGGGGCTGG
CTTAACTATGCGGCATCAGAGCAGATTGTACTGAGAGTGCACCATATGCGGTGTGAA
ATACCGCACAGATGCGTAAGGAGAAAATACCGCATCAGGCGGCCTTAAGGGCCTCG
TGATACGCCTATTTTTATAGGTTAATGTCATGATAATAATGGTTTCTTAGACGTCAGG
TGGCACTTTTCGGGGAAATGTGCGCGGAACCCCTATTTGTTTATTTTTCTAAATACAT
TCAAATATGTATCCGCTCATGAGACAATAACCCTGATAAATGCTTCAATAATATTGA
AAAAGGAAGAGTATGAGTATTCAACATTTCCGTGTCGCCCTTATTCCCTTTTTTGCGG
CATTTTGCCTTCCTGTTTTTGCTCACCCAGAAACGCTGGTGAAAGTAAAAGATGCTG
AAGATCAGTTGGGTGCACGAGTGGGTTACATCGAACTGGATCTCAACAGCGGTAAG
ATCCTTGAGAGTTTTCGCCCCGAAGAACGTTTTCCAATGATGAGCACTTTTAAAGTT
CTGCTATGTGGCGCGGTATTATCCCGTATTGACGCCGGGCAAGAGCAACTCGGTCGC
CGCATACACTATTCTCAGAATGACTTGGTTGAGTACTCACCAGTCACAGAAAAGCAT
CTTACGGATGGCATGACAGTAAGAGAATTATGCAGTGCTGCCATAACCATGAGTGA
TAACACTGCGGCCAACTTACTTCTGACAACGATCGGAGGACCGAAGGAGCTAACCG
CTTTTTTGCACAACATGGGGGATCATGTAACTCGCCTTGATCGTTGGGAACCGGAGC
TGAATGAAGCCATACCAAACGACGAGCGTGACACCACGATGCCTGTAGCAATGGCA
ACAACGTTGCGCAAACTATTAACTGGCGAACTACTTACTCTAGCTTCCCGGCAACAA
TTAATAGACTGGATGGAGGCGGATAAAGTTGCAGGACCACTTCTGCGCTCGGCCCTT
CCGGCTGGCTGGTTTATTGCTGATAAATCTGGAGCCGGTGAGCGTGGGTCTCGCGGT
ATCATTGCAGCACTGGGGCCAGATGGTAAGCCCTCCCGTATCGTAGTTATCTACACG
ACGGGGAGTCAGGCAACTATGGATGAACGAAATAGACAGATCGCTGAGATAGGTGC

CTCACTGATTAAGCATTGGTAACTGTCAGACCAAGTTTACTCATATATACTTTAGATT
GATTTAAAACTTCATTTTTAATTTAAAAGGATCTAGGTGAAGATCCTTTTTGATAATC
TCATGACCAAAATCCCTTAACGTGAGTTTTCGTTCCACTGAGCGTCAGACCCCGTAG
AAAAGATCAAAGGATCTTCTTGAGATCCTTTTTTTCTGCGCGTAATCTGCTGCTTGCA
AACAAAAAAACCACCGCTACCAGCGGTGGTTTGTTTGCCGGATCAAGAGCTACCAA
CTCTTTTTCCGAAGGTAACTGGCTTCAGCAGAGCGCAGATACCAAATACTGTCCTTC
TAGTGTAGCCGTAGTTAGGCCACCACTTCAAGAACTCTGTAGCACCGCCTACATACC
TCGCTCTGCTAATCCTGTTACCAGTGGCTGCTGCCAGTGGCGATAAGTCGTGTCTTAC
CGGGTTGGACTCAAGACGATAGTTACCGGATAAGGCGCAGCGGTCGGGCTGAACGG
GGGGTTCGTGCACACAGCCCAGCTTGGAGCGAACGACCTACACCGAACTGAGATAC
CTACAGCGTGAGCATTGAGAAAGCGCCACGCTTCCCGAAGGGAGAAAGGCGGACA
GGTATCCGGTAAGCGGCAGGGTCGGAACAGGAGAGCGCACGAGGGAGCTTCCAGG
GGGAAACGCCTGGTATCTTTATAGTCCTGTCGGGTTTCGCCACCTCTGACTTGAGCG
TCGATTTTTGTGATGCTCGTCAGGGGGGCGGAGCCTATGGAAAAACGCCAGCAACG
CGGCCTTTTTACGGTTCCTGGCCTTTTGCTGGCCTTTTGCTCACATGTTCTTTCCTGCG
TTATCCCCTGATTCTGTGGATAACCGTATTACCGCCTTTGAGTGAGCTGATACCGCTC
GCCGCAGCCGAACGACCGAGCGCAGCGAGTCAGTGAGCGAGGAAGCGGAAGAGCG
CCCAATACGCAAACCGCCTCTCCCCGCGCGTTGGCCGATTCATTAATGCAGCTGGCA
CGACAGGTTTCCCGACTGGAAAGCGGGCAGTGAGCGCAACGCAATTAATGTGAGTT
AGCTCACTCATTAGGCACCCCAGGCTTTACACTTTATGCTTCCGGCTCGTATGTTGTG
TGGAATTGTGAGCGGATAACAATTTCACACAGGAAACAGC

Figure C-2 pBYU1



ATGACCATGATTACGCCAAGCTTATCCCATTCTCTCATCAATTGAGTGGGATGAGGC
TATCTCTGCCTCTCTTCTGAATCTCTGAACCATCTTACATTACACTGTGGATGACGgC
TAGCGTCGACTCCGGCGGAGAAAAGCGGCTTGGATGAAAGGACGATCAAATAAAAG
CAGAATCTCGGTACGCCGGCGGCGAGTTAAGGATATAAGAGTTAAGGGACTCCACT
GAATCGGGAAGAAAGGACGGCGATGCAGGTAAGCAAATGCAAAAGATACAAATAC
GGATTCCAGTAATGGATGACGTGACAGAAGACCGCCAGGACGCTGCGAGATAGAAA
CACTGGACTATTGGTCTCCCGGCCAAAAAGAAATAAATAATAATGTAAACTACTCCC
GGGCGAGAGAACCGCGGTCAGTACAAGATTAAGGAATGGTTATTCTCGCCCGGGGA
AATGCGGTCGGCCTCTCGGCCAGAAACCGATGGCGGAAGAAAGAAACTGGATCTGC
TCCGAGCAGGGAAACCCCCGCGACTGGAAACCACGGGGAGGCCCCCGGGAGGCGA
ATGTATACCGTCCTCATATTGAAGCTCCTGGATAGGGAAATATCGGGGCCGGCCGGA
TGGGCATCGGTGCTCCGGGTATATCAGACCGGGAGATAGGAGTCGGCGGCTACCCG
TGCATGAGAGGAGGGACGCATGACCTAGGATCAACTCGACGGCCGCCTTATTAAAT
TGGTCCCAGGGCAGTGGAAGGGCGAGCATGCTGGATGGTATAGCTAGAGGGACAGA
AAACTTTCGAGACCTCATGCCAGAAAAGAATAAGCGACGAACGCTGGACGACGACT
TACAGCGGGAAACGATCAAATACCATCCAGACTACTTTTTTGTATCTTGAGGAAGTA
ATAACGAATCAGGTGCAGGCACACTAACTAAGTTGGTTGTAGAGTTGCGGTTGAAA
TCTAGAGGATCGAGGCATTTGAATTGGGGGTGGTGGACAGTAACTGTCTGTAATAAT
AATTACTCCTGACCAGGTTGCAATTCGAGTTTTGATAAGCATAATTATACCTTGTAC

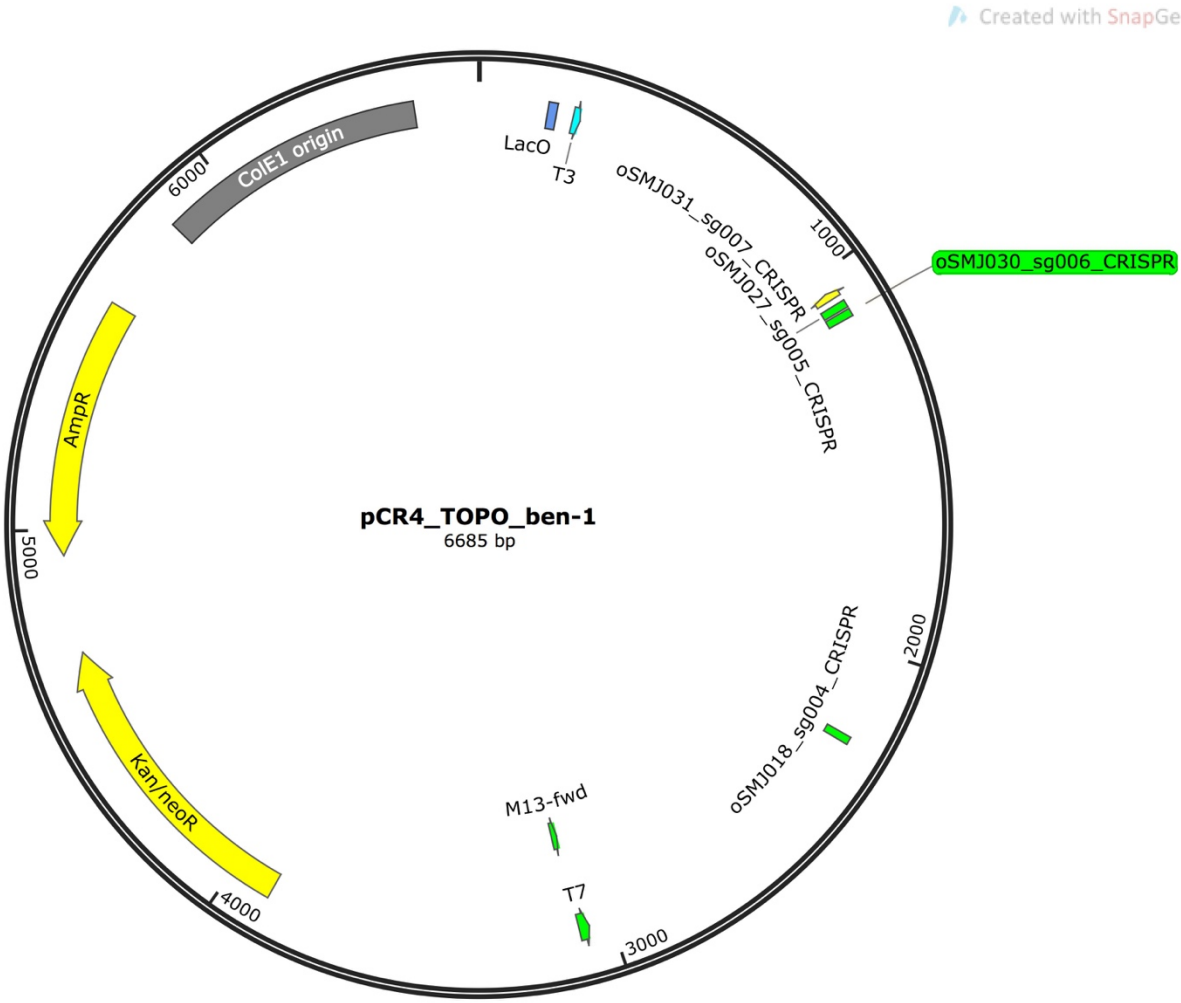
ATTGTGGGTTTTGTGCTGTGGACGTTTTATTGTGGACATCCCCATAAGCTACAAGAA
ACCAAAAATGAAATTAAAAGTATTGAAAAACGTCGTAACATTTTATATCTGAGTAGT
ATCCTTTGCTTTAAATGTCCATAAAAATAATTTTATAATCAATAAAACAACGTTTGTA
AATCAACTGAGTTTACAAGTAGAGACATTGAGGGATACTTTCACTATGCTAAAGTGA
ATAATCGACCAAATAATAACTCACTTTGGTATTTATTCCTGTCTTATAATGTTATGTA
TGAATTAAATTCATATGCATATGGCTCACTCTGACAAAAAAAAAATAATCTTCCAGAT
CAATATTGACTACCGATGCGGGTGGTCTTTTGCTTTGAATTCTGCTGAACTTTACACC
CCGAACAGCAATGTGTGCTTCAGCTAAAAAAAAGTAAGTGTGTTAATCAGTCCCCCC
GATTCTTCATTTTTTGCCCCTCTCTCCCGTTTCGTCGGCAAAAGAAGAGAAAATAAA
GATAAGTCTCAAGATAGGTTGGTAATCGCTAAAGTGGTTGTGTGGATAAGAGTAGC
AAAATGGCAGGAAGAGCACTTTGCGCGCACACACTGTACTCATTGTTCTGGATAAA
ATTCTCTCGTTGTTTGCCGTCGGATGTCTGCCTCTCTGCCATTGAGCCGGCTTCTTCA
CTATCTTTAGTTAACCTAAAATGCCGTTTCTTTTCTCGTATCCCACTATCCGTTGAGG
TTCTCTGCTCTCTTCGCTCCCTTACCGCCAGCGAGCAACTATCCGTGGGGGCGCCTTG
CTCGGAAGATGGGGGGGAAGAAAGAAGATTTTTGCTATTTGCACTTGAGAAAGAGA
CTTTTCCTGCGTCGATGGTTAGAGAACAGTGTGCAGACACTTTTCAGCTACCTAGAT
ACATGGATATCCCCGCCTCCCAATCCACCCACCCAGGGAAAAAGAAGGGCTCGCCG
AAAAATCAAAGTTATCTCCAGGCTCGCGCATCCCACCGAGCGGTTGACTTCTCTCCA
CCACTTTTCATTTTAACCCTCGGGGTACGGGATTGGCCAAAGGACCCAAAGGTATGT
TTCGAATGATACTAACATAACATAGAACATTTTCAGGAGGACCCTTGGAGGGTACCG
AGCTCAGAAAAAATGACTGCTCCAAAGAAGAAGCGTAAGGTACCGGTAGAAAAAA
TGAGTAAAGGAGAAGAACTTTTCACTGGAGTTGTCCCAATTCTTGTTGAATTAGATG
GTGATGTTAATGGGCACAAATTTTCTGTCAGTGGAGAGGGTGAAGGTGATGCAACA
TACGGAAAACTTACCCTTAAATTTATTTGCACTACTGGAAAACTACCTGTTCCATGGg
taagtttaaacatatatactaactaaccctgattatttaaattttcagCCAACACTTGTCACTACTTTCTgTTATGGT
GTTCAATGCTTcTCgAGATACCCAGATCATATGAAACgGCATGACTTTTTCAAGAGTG
CCATGCCCGAAGGTTATGTACAGGAAAGAACTATATTTTTCAAAGATGACGGGAAC
TACAAGACACgtaagtttaaacagttcggtactaactaaccatacatatttaaattttcagGTGCTGAAGTCAAGTTT
GAAGGTGATACCCTTGTTAATAGAATCGAGTTAAAAGGTATTGATTTTAAAGAAGAT
GGAAACATTCTTGGACACAAATTGGAATACAACTATAACTCACACAATGTATACATC
ATGGCAGACAAACAAAAGAATGGAATCAAAGTTgtaagtttaaacatgattttactaactaactaatctgattt
aaattttcagAACTTCAAAATTAGACACAACATTGAAGATGGAAGCGTTCAACTAGCAGAC
CATTATCAACAAAATACTCCAATTGGCGATGGCCCTGTCCTTTTACCAGACAACCAT
TACCTGTCCACACAATCTGCCCTTTCGAAAGATCCCAACGAAAAGAGAGACCACAT
GGTCCTTCTTGAGTTTGTAACAGCTGCTGGGATTACACATGGCATGGATGAACTATA
CAAATgcCCGGTGGGTGAAGACCAGAAACAGCACCTCGAACTGAGCCGCGATATTGC
CCAGCGTTTCAACGCGCTGTATGGCGAGATCGATCCCGTCGTTTTACAACGTCGTGA
CTGGGAAAACCCTGGCGTTACCCAACTTAATCGCCTTGCAGCACATCCCCCTTTCGC
CAGCTGGCGTAATAGCGAAGAGGCCCGCACCGATCGCCCTTCCCAACAGTTGCGaAG
gtaagtttaaacagatccatactaactaacttgttctgacataattttcagCtTGAATGGCGAATGGCGCTTTGCCTGG
TTTCCGGCACCAGAAGCGGTGCCGGAAAGCTGGCTGGAGTGCGATCTTCCTGAGGC
CGATACTGTCGTCGTCCCCTCAAACTGGCAGATGCACGGTTACGATGCGCCCATCTA
CACCAACGTAACCTATCCCATTACGGTCAATCCGCCGTTTGTTCCCACGGAGAATCC
GACGGGTTGTTACTCGCTCACATTTAATGTTGATGAAAGCTGGCTACAGGAAGGCCA
GACGCGAATTATTTTTGATGGCGTTAACTCGGCGTTTCATCTGTGGTGCAACGGGCG
CTGGGTCGGTTACGGCCAGgtaagtttaattaagttgatactaactaacaaagatctgattaattttcagGACAGTCG

TTTGCCGTCTGAATTTGACCTGAGCGCATTTTTACGCGCCGGAGAAAACCGCCTCGC
GGTGATGGTGCTGCGTTGGAGTGACGGCAGTTATCTGGAAGATCAGGATATGTGGC
GGATGAGCGGCATTTTCCGTGACGTCTCGTTGCTGCATAAACCGACTACACAAATCA
GCGATTTCCATGTTGCCACTCGCTTTAATGATGATTTCtccCGCGCTGTACTGGAGGCT
GAAGTcCAGgtaagtttaaacaggatcttactaactaacatgctaacactgaattttcagATGTGCGGCGAGTTGCGT
GACTACCTACGGGTAACAGTTTCTTTATGGCAGGGTGAAACGCAGGTCGCCAGCGG
CACCGCGCCTTTCGGCGGTGAAATTATCGATGAGCGTGGTGGTTATGCCGATCGCGT
CACACTACGTCTGAACGTCGAAAACCCGAAACTGTGGAGCGCCGAAATCCCGAATC
TCTATCGTGCGGTGGTTGAACTGCACACCGCCGACGGCACGCTGATTGAAGCAGAA
GCCTGCGATGTCGGTTTCCGCGAGGTGCGGATTGAAAATGGTCTGCTGCTGCTGAAC
GGCAAGCCGTTGCTGATTCGAGGCGTTAACCGTCACGAGCATCATCCTCTGCATGGT
CAGGTCATGGATGAGCAGACGATGGTGCAGGATgtaagtttaaactattcgttactaactaactttaaacattta
aattttcagATCCTGCTGATGAAGCAGAACAACTTTAACGCCGTGCGCTGTTCGCATTATC
CGAACCATCCGCTGTGGTACACGCTGTGCGACCGCTACGGCCTGTATGTGGTGGATG
AAGCCAATATTGAAACCCACGGCATGGTGCCAATGAATCGTCTGACCGATGATCCG
CGCTGGCTACCGGCGATGAGCGAACGCGTAACGCGAATGGTGCAGCGCGATCGTAA
TCACCCGAGTGTGATCATCTGGTCGCTGGGGAATGgtaagtttaaacagttgaatactaactaacggagatc
tttgaaattttcagAATCAGGCCACGGCGCTAATCACGACGCGCTGTATCGCTGGATCAAATC
TGTCGATCCTTCCCGCCCGGTGCAGTATGAAGGCGGCGGAGCCGACACCACGGCCA
CCGATATTATTTGCCCGATGTACGCGCGCGTGGATGAAGACCAGCCCTTCCCGGCTG
TGCCGAAATGGTCCATCAAAAAATGGCTTTCGCTACCTGGAGAGACGCGCCCGCTG
ATtCTTTGCGAGgtaagtttaaacagaactctactaactaacacattagatcctaattttcagTACGCtCACGCGATGG
GcAACAGTCTTGGCGGTTTCGCTAAATACTGGCAGGCGTTTCGTCAGTATCCCCGTTT
ACAGGGCGGCTTCGTCTGGGACTGGGTGGATCAGTCGCTGATTAAATATGATGAAA
ACGGCAACCCGTGGTCGGCTTACGGCGGTGATTTTGGCGATACGCCGAACGATCGC
CAGTTCTGTATGAACGGTCTGGTCTTTGCCGACCGCACGCCGCATCCAGCgtaagtttaaac
aataacctactaactaacgtagataatttaaattttcagGCTGACGGAAGCAAAACACCAGCAGCAGTTTTTC
CAGTTCCGTTTATCCGGGCAAACCATCGAAGTGACCAGCGAATACCTGTTCCGTCAT
AGCGATAACGAGCTCCTGCACTGGATGGTGGCGCTGGATGGTAAGCCGCTGGCAAG
CGGTGAAGTGCCTCTGGATGTCGCTCCACAAGGTAAACAGTTGATTGAACTGCCTGA
ACTACCGCAGCCGGAGAGCGCCGGGCAACTCTGGCTCACAGTACGCGTAGTGCAAC
CGAACGCGACCGCATGGTCAGAAGCCGGGCACATCAGCGCaTGGCAGCAGTGGAGgt
aagtttaaacaagatcctactaactaactctacattgatgaattttcagACTGGCGGAAAACCTCAGTGTGACGCTC
CCCGCCGCGTCCCACGCCATCCCGCATCTGACCACCAGCGAAATGGATTTTTGCATC
GAGCTGGGTAATAAGCGTTGGCAATTTAACCGCCAGTCAGGCTTTCTTTCACAGATG
TGGATTGGCGATAAAAAACAACTGCTGACGCCGCTGCGCGATCAGTTCACCCGTGC
ACCGCTGGATAACGACATTGGCGTAAGTGAAGCGACCCGCATTGACCCTAACGCCT
GGGTCGgtaagtttaaacaaagttgtactaactaacgaagatcttgataattttcagAACGCTGGAAGGCGGCGGGC
CATTACCAGGCCGAAGCAGCGTTGTTGCAGTGCACGGCAGATACACTTGCTGATGCG
GTGCTGATTACGACCGCTCACGCGTGGCAGCATCAGGGGAAAACCTTATTTATCAGC
CGGAAAACCTACCGGATTGATGGTAGTGGTCAAATGGCGATTACCGTTGATGTTGAA
GTGGCGAGCGATACACCGCATCCGGCGCGGATTGGCCTGAACTGCCAGCTGGCGCA
GGTAGCAGAGCGGGTAAACTGGCTCGGATTAGGGCCGCAAGAAAACTATCCCGACC
GCCTTACTGCCGCCTGTTTTGACCGCTGGGATCTGCCATTGTCAGACATGTAgtaagtttaa
acttgatagtactaactaacatgtttcatttaaattttcagTACCCCGTACGTCTTCCCGAGCGAAAACGGTCTG
CGCTGCGGGACGCGCGAATTGAATTATGGCCCACACCAGTGGCGCGGCGACTTCCA

GTTCAACATCAGCCGCTACAGTCAACAGCAACTGATGGAAACCAGCCATCGCCATC
TGCTGCACGCGGAAGAAGGCACATGGCTGAATATCGACGGTTTCCATATGGGGATT
GGTGGCGACGACTCCTGGAGCCCGTCAGTATCGGCGGAATTCCAACTGAGCGCCGG
TCGCTACCATTACCAACTTGTCTGGTGTCAAAAATAATAGGGGCCGCTGTCATCAGA
gtaagtttaaactgagttctactaactaacgagtaatatttaaattttcagCATCTCGCGCCCGTGCCTCTGACTTCTA
AGTCCAATTACTCTTCAACATCCCTACATGCTCTTTCTCCCTGTGCTCCCACCCCCTA
TTTTTGTTATTATCAAAAAAACTTCTTCTTAATTTCTTTGTTTTTTAGCTTCTTTTAAG
TCACCTCTAACAATGAAATTGTGTAGATTCAAAAATAGAATTAATTCGTAATAAAAA
GTCGAAAAAAATTGTGCTCCCTCCCCCCATTAATAATAATTCTATCCCAAAATCTAC
ACAATGTTCTGTGTACACTTCTTATGTTTTTTTTACTTCTGATAAATTTTTTTTGAAAC
ATCATAGAAAAAACCGCACACAAAATACCTTATCATATGTTACGTTTCAGTTTATGA
CCGCAATTTTTATTTCTTCGCACGTCTGGGCCTCTCATGACGTCAAATCATGCTCATC
GTGAAAAAGTTTTGGAGTATTTTTGGAATTTTTCAATCAAGTGAAAGTTTATGAAAT
TAATTTTCCTGCTTTTGCTTTTTGGGGGTTTCCCCTATTGTTTGTCAAGAGTTTCGAGG
ACGGCGTTTTTCTTGCTAAAATCACAAGTATTGATGAGCACGATGCAAGAAAGATCG
GAAGAAGGTTTGGGTTTGAGGCTCAGTGGAAGGTGAGTAGAAGTTGATAATTTGAA
AGTGGAGTAGTGTCTATGGGGTTTTTGCCTTAAATGACAGAATACATTCCCAATATA
CCAAACATAACTGTTTCCTACTAGTCGGCCGTACGGGCCCTTTCGTCTCGCGCGTTTC
GGTGATGACGGTGAAAACCTCTGACACATGCAGCTCCCGGAGACGGTCACAGCTTG
TCTGTAAGCGGATGCCGGGAGCAGACAAGCCCGTCAGGGCGCGTCAGCGGGTGTTG
GCGGGTGTCGGGGCTGGCTTAACTATGCGGCATCAGAGCAGATTGTACTGAGAGTG
CACCATATGCGGTGTGAAATACCGCACAGATGCGTAAGGAGAAAATACCGCATCAG
GCGGCCTTAAGGGCCTCGTGATACGCCTATTTTTATAGGTTAATGTCATGATAATAA
TGGTTTCTTAGACGTCAGGTGGCACTTTTCGGGGAAATGTGCGCGGAACCCCTATTT
GTTTATTTTTCTAAATACATTCAAATATGTATCCGCTCATGAGACAATAACCCTGATA
AATGCTTCAATAATATTGAAAAAGGAAGAGTATGAGTATTCAACATTTCCGTGTCGC
CCTTATTCCCTTTTTTGCGGCATTTTGCCTTCCTGTTTTTGCTCACCCAGAAACGCTGG
TGAAAGTAAAAGATGCTGAAGATCAGTTGGGTGCACGAGTGGGTTACATCGAACTG
GATCTCAACAGCGGTAAGATCCTTGAGAGTTTTCGCCCCGAAGAACGTTTTCCAATG
ATGAGCACTTTTAAAGTTCTGCTATGTGGCGCGGTATTATCCCGTATTGACGCCGGG
CAAGAGCAACTCGGTCGCCGCATACACTATTCTCAGAATGACTTGGTTGAGTACTCA
CCAGTCACAGAAAAGCATCTTACGGATGGCATGACAGTAAGAGAATTATGCAGTGC
TGCCATAACCATGAGTGATAACACTGCGGCCAACTTACTTCTGACAACGATCGGAGG
ACCGAAGGAGCTAACCGCTTTTTTGCACAACATGGGGGATCATGTAACTCGCCTTGA
TCGTTGGGAACCGGAGCTGAATGAAGCCATACCAAACGACGAGCGTGACACCACGA
TGCCTGTAGCAATGGCAACAACGTTGCGCAAACTATTAACTGGCGAACTACTTACTC
TAGCTTCCCGGCAACAATTAATAGACTGGATGGAGGCGGATAAAGTTGCAGGACCA
CTTCTGCGCTCGGCCCTTCCGGCTGGCTGGTTTATTGCTGATAAATCTGGAGCCGGTG
AGCGTGGGTCTCGCGGTATCATTGCAGCACTGGGGCCAGATGGTAAGCCCTCCCGTA
TCGTAGTTATCTACACGACGGGGAGTCAGGCAACTATGGATGAACGAAATAGACAG
ATCGCTGAGATAGGTGCCTCACTGATTAAGCATTGGTAACTGTCAGACCAAGTTTAC
TCATATATACTTTAGATTGATTTAAAACTTCATTTTTAATTTAAAAGGATCTAGGTGA
AGATCCTTTTTGATAATCTCATGACCAAAATCCCTTAACGTGAGTTTTCGTTCCACTG
AGCGTCAGACCCCGTAGAAAAGATCAAAGGATCTTCTTGAGATCCTTTTTTTCTGCG
CGTAATCTGCTGCTTGCAAACAAAAAAACCACCGCTACCAGCGGTGGTTTGTTTGCC
GGATCAAGAGCTACCAACTCTTTTTCCGAAGGTAACTGGCTTCAGCAGAGCGCAGAT

ACCAAATACTGTCCTTCTAGTGTAGCCGTAGTTAGGCCACCACTTCAAGAACTCTGT
AGCACCGCCTACATACCTCGCTCTGCTAATCCTGTTACCAGTGGCTGCTGCCAGTGG
CGATAAGTCGTGTCTTACCGGGTTGGACTCAAGACGATAGTTACCGGATAAGGCGC
AGCGGTCGGGCTGAACGGGGGGTTCGTGCACACAGCCCAGCTTGGAGCGAACGACC
TACACCGAACTGAGATACCTACAGCGTGAGCATTGAGAAAGCGCCACGCTTCCCGA
AGGGAGAAAGGCGGACAGGTATCCGGTAAGCGGCAGGGTCGGAACAGGAGAGCGC
ACGAGGGAGCTTCCAGGGGGAAACGCCTGGTATCTTTATAGTCCTGTCGGGTTTCGC
CACCTCTGACTTGAGCGTCGATTTTTGTGATGCTCGTCAGGGGGGCGGAGCCTATGG
AAAAACGCCAGCAACGCGGCCTTTTTACGGTTCCTGGCCTTTTGCTGGCCTTTTGCTC
ACATGTTCTTTCCTGCGTTATCCCCTGATTCTGTGGATAACCGTATTACCGCCTTTGA
GTGAGCTGATACCGCTCGCCGCAGCCGAACGACCGAGCGCAGCGAGTCAGTGAGCG
AGGAAGCGGAAGAGCGCCCAATACGCAAACCGCCTCTCCCCGCGCGTTGGCCGATT
CATTAATGCAGCTGGCACGACAGGTTTCCCGACTGGAAAGCGGGCAGTGAGCGCAA
CGCAATTAATGTGAGTTAGCTCACTCATTAGGCACCCCAGGCTTTACACTTTATGCTT
CCGGCTCGTATGTTGTGTGGAATTGTGAGCGGATAACAATTTCACACAGGAAACAGC

Figure C-3 pCR4_TOPO_ben-1



AGCGCCCAATACGCAAACCGCCTCTCCCCGCGCGTTGGCCGATTCATTAATGCAGCT
GGCACGACAGGTTTCCCGACTGGAAAGCGGGCAGTGAGCGCAACGCAATTAATGTG
AGTTAGCTCACTCATTAGGCACCCCAGGCTTTACACTTTATGCTTCCGGCTCGTATGT
TGTGTGGAATTGTGAGCGGATAACAATTTCACACAGGAAACAGCTATGACCATGATT
ACGCCAAGCTCAGAATTAACCCTCACTAAAGGGACTAGTCCTGCAGGTTTAAACGA
ATTCGCCCTgctgcagccgcatacaatcgattctgaagaagcggcggcgccccgccgatggctccgcgagccaattgtgctctcc
gagtgcattccgacgaaaggacgaagacaacaacaacagagacacagagtataattgacgagaattggaagagcgcgtcccctgaccttt
cttcagtattcctatcctcctatttcataattctagctgtaaaatttcgggatttatgggccctatttctcggaagacttaagtggtgccaggctgtct
cattgcagtttggtctacaaaaaatgcgggaattttcgcccgaaatgcaaaatcagttgagaactatgcctctcgtctcccgcatttttgtagat
ctacgtagatcaaaccaaaatgagacactctgacaccatgtgaagattggtcattgttttgtcccccgtccttgtaaaagtaatttagccgcgcg
gcggtgtcaccgcgaccccatggccatctgccaattttggcccctggctagttcaaacgaagagaaggacggagctcccaatttccaatcg
ctttttttccctttttgtttcacattctcactcgttgtttgcgagtttaaaatttaaaattcaacgaatttttacctttttgtttcccctctacgtgacccttctc
gttttgcactccggttacaaaatattcaatattcaatatttcagttctcatataacttcaaaaagaacttggaaaaaATGAGAGAAATT
GTTCACGTTCAAGCCGGACAATGTGGTAATCAAATCGGAGCCAAGTTCTGGGAAGT
GATATCCGATGAGCATGGGATCCAGCCTGATGGAACTTATAAGGGAGAAAGTGATT
TGCAGTTGGAAAGAATCAATGTCTACTATAATGAGGCTAATGgtgagaaatttagctttttattcgat

tttcagattctgtttgaatataataaagcaaaatgttccgaaattttctttaacttccaattttcaatatttattgctcaatcgcaaaatttgttcgagtgt
tgcgcaataaatacggtgcccggtctcgacacgacttttgttgacaacgaaaggtcgtgcgcctttaaaggatactgtagcttcaaacttttgtg
gcagcgggctttgcattagttttcatagttttttgattgataaatgtgtacttatgtatttttttattaaaaactcaaatatttattacacattttaacaaatt
aattctgcaaattatgagaatgaacggaagatatattgccagagacactattaccggtacagagagtgtagatagttagagagtgacagacat
acgggaacctatggggcggggcgcgcggaagagaagatttgtgtcgatttacgaaatgatgacaacgaggaaaatttcgtaaatcgacac
aaatcgtctcttccgcgcgccccgccccattgggtcccggatgtctgtcactctctaactatatacactctctgtaccggtattacgaacgtagg
aatcgtggtattttgagagacaaaaccaccggcgtatactagtttcttgcacacttttcggttgaattaggattttagttagtcaaacaaaagctc
aaagacagctttctgaaattcacgtttacaaattcacgaatagttatttttatagacccattctgatgaaatttcagtatatctgtttctcccatttttcg
attctataaacgtgtggtgtctcttgcgcattcgcgctctatagcaaccaatttttttcaatttttattttttctttccaaagataaatcttcgattaagac
ctcatttttgttgggttttaaaaaaaaaaattttttaaattaatttaaaatttaaaatttaattctctcaatgtgaattaccaaaattcactaattgttaccatt
ttcagGCGGCAAATATGTCCCACGCGCTGTTCTCGTCGATCTTGAGCC**AGG**AACCATGG
ATTCTGTCCGCTCTGGACCATTCGGCCAGCTGTTCCGTCCAGATAACTTTGTGTTCGG
ACAATCCGGAGCCGGAAACAACTGGGCCAAGGGTCACTACACCGAAGGAGCCGAA
CTTGTCGATAATGTGCTCGACGTAGTTCGAAAAGAGGCTGAAGGATGTGATTGTCTT
CAAGGATTCCAGTTGACTCATTCACTTGGAGGAGGAACTGGATCTGGAATGGGAAC
TCTTCTCATTTCGAAAATCCGTGAAGAGTATCCAGATAGAATTATGAGTTCTTTCTCG
GTTGTTCCGTCGCCAAAGgttggattaattgaatttaatgaatattttaaaactaataattaaaattcagGTCTCGGACA
CAGTCGTCGAGCCATACAACGCTACTCTTTCTGTCCACCAGCTCGTTGAAAATACCG
ATGAGACTTTCTGCATTGACAACGAGGCTCTTTATGATATCTGCTTCAGAACCCTCA
AGCTTTCAAATCCAACTTATGGAGATCTCAATCATCTTGgtaagattttcttttatttttattatttttctatttt
aactttcaattatttcagTTTCCGTTACAATGTCCGGAGTCACCACGTGCCTCCGCTTCCCAGGA
CAACTCAATGCTGATCTCCGCAAACTTGCAGTCAACATGGTTCCATTCCCACGTCTT
CACTTCTTTATGCCAGGATTTGCTCCATTGTCAGCTAAAGGAGCACAAGCGTACCGT
GCACTTACGGTCGCCGAGCTTACCCAGCAGATTAAGGGCGAATTCGCGGCCGCTAA
ATTCAATTCGCCCTATAGTGAGTCGTATTACAATTCACTGGCCGTCGTTTTACAACGT
CGTGACTGGGAAAACCCTGGCGTTACCCAACTTAATCGCCTTGCAGCACATCCCCCT
TTCGCCAGCTGGCGTAATAGCGAAGAGGCCCGCACCGATCGCCCTTCCCAACAGTTG
CGCAGCCTATACGTACGGCAGTTTAAGGTTTACACCTATAAAAGAGAGAGCCGTTAT
CGTCTGTTTGTGGATGTACAGAGTGATATTATTGACACGCCGGGGCGACGGATGGTG
ATCCCCCTGGCCAGTGCACGTCTGCTGTCAGATAAAGTCTCCCGTGAACTTTACCCG
GTGGTGCATATCGGGGATGAAAGCTGGCGCATGATGACCACCGATATGGCCAGTGT
GCCGGTCTCCGTTATCGGGGAAGAAGTGGCTGATCTCAGCCACCGCGAAAATGACA
TCAAAAACGCCATTAACCTGATGTTCTGGGGAATATAAATGTCAGGCATGAGATTAT
CAAAAAGGATCTTCACCTAGATCCTTTTCACGTAGAAAGCCAGTCCGCAGAAACGGT
GCTGACCCCGGATGAATGTCAGCTACTGGGCTATCTGGACAAGGGAAAACGCAAGC
GCAAAGAGAAAGCAGGTAGCTTGCAGTGGGCTTACATGGCGATAGCTAGACTGGGC
GGTTTTATGGACAGCAAGCGAACCGGAATTGCCAGCTGGGGCGCCCTCTGGTAAGG
TTGGGAAGCCCTGCAAAGTAAACTGGATGGCTTTCTTGCCGCCAAGGATCTGATGGC
GCAGGGGATCAAGCTCTGATCAAGAGACAGGATGAGGATCGTTTCGCATGATTGAA
CAAGATGGATTGCACGCAGGTTCTCCGGCCGCTTGGGTGGAGAGGCTATTCGGCTAT
GACTGGGCACAACAGACAATCGGCTGCTCTGATGCCGCCGTGTTCCGGCTGTCAGCG
CAGGGGCGCCCGGTTCTTTTTGTCAAGACCGACCTGTCCGGTGCCCTGAATGAACTG
CAAGACGAGGCAGCGCGGCTATCGTGGCTGGCCACGACGGGCGTTCCTTGCGCAGC
TGTGCTCGACGTTGTCACTGAAGCGGGAAGGGACTGGCTGCTATTGGGCGAAGTGC
CGGGGCAGGATCTCCTGTCATCTCACCTTGCTCCTGCCGAGAAAGTATCCATCATGG
CTGATGCAATGCGGCGGCTGCATACGCTTGATCCGGCTACCTGCCCATTCGACCACC

AAGCGAAACATCGCATCGAGCGAGCACGTACTCGGATGGAAGCCGGTCTTGTCGAT
CAGGATGATCTGGACGAAGAGCATCAGGGGCTCGCGCCAGCCGAACTGTTCGCCAG
GCTCAAGGCGAGCATGCCCGACGGCGAGGATCTCGTCGTGACCCATGGCGATGCCT
GCTTGCCGAATATCATGGTGGAAAATGGCCGCTTTTCTGGATTCATCGACTGTGGCC
GGCTGGGTGTGGCGGACCGCTATCAGGACATAGCGTTGGCTACCCGTGATATTGCTG
AAGAGCTTGGCGGCGAATGGGCTGACCGCTTCCTCGTGCTTTACGGTATCGCCGCTC
CCGATTCGCAGCGCATCGCCTTCTATCGCCTTCTTGACGAGTTCTTCTGAATTATTAA
CGCTTACAATTTCCTGATGCGGTATTTTCTCCTTACGCATCTGTGCGGTATTTCACAC
CGCATCAGGTGGCACTTTTCGGGGAAATGTGCGCGGAACCCCTATTTGTTTATTTTTC
TAAATACATTCAAATATGTATCCGCTCATGAGATTATCAAAAAGGATCTTCACCTAG
ATCCTTTTAAATTAAAAATGAAGTTTTAAATCAATCTAAAGTATATATGAGTAAACT
TGGTCTGACAGTTACCAATGCTTAATCAGTGAGGCACCTATCTCAGCGATCTGTCTA
TTTCGTTCATCCATAGTTGCCTGACTCCCCGTCGTGTAGATAACTACGATACGGGAG
GGCTTACCATCTGGCCCCAGTGCTGCAATGATACCGCGAGACCCACGCTCACCGGCT
CCAGATTTATCAGCAATAAACCAGCCAGCCGGAAGGGCCGAGCGCAGAAGTGGTCC
TGCAACTTTATCCGCCTCCATCCAGTCTATTAATTGTTGCCGGGAAGCTAGAGTAAG
TAGTTCGCCAGTTAATAGTTTGCGCAACGTTGTTGCCATTGCTACAGGCATCGTGGT
GTCACGCTCGTCGTTTGGTATGGCTTCATTCAGCTCCGGTTCCCAACGATCAAGGCG
AGTTACATGATCCCCCATGTTGTGCAAAAAAGCGGTTAGCTCCTTCGGTCCTCCGAT
CGTTGTCAGAAGTAAGTTGGCCGCAGTGTTATCACTCATGGTTATGGCAGCACTGCA
TAATTCTCTTACTGTCATGCCATCCGTAAGATGCTTTTCTGTGACTGGTGAGTACTCA
ACCAAGTCATTCTGAGAATAGTGTATGCGGCGACCGAGTTGCTCTTGCCCGGCGTCA
ATACGGGATAATACCGCGCCACATAGCAGAACTTTAAAAGTGCTCATCATTGGAAA
ACGTTCTTCGGGGCGAAAACTCTCAAGGATCTTACCGCTGTTGAGATCCAGTTCGAT
GTAACCCACTCGTGCACCCAACTGATCTTCAGCATCTTTTACTTTCACCAGCGTTTCT
GGGTGAGCAAAAACAGGAAGGCAAAATGCCGCAAAAAAGGGAATAAGGGCGACAC
GGAAATGTTGAATACTCATACTCTTCCTTTTTCAATATTATTGAAGCATTTATCAGGG
TTATTGTCTCATGACCAAAATCCCTTAACGTGAGTTTTCGTTCCACTGAGCGTCAGAC
CCCGTAGAAAAGATCAAAGGATCTTCTTGAGATCCTTTTTTTCTGCGCGTAATCTGCT
GCTTGCAAACAAAAAAACCACCGCTACCAGCGGTGGTTTGTTTGCCGGATCAAGAG
CTACCAACTCTTTTTCCGAAGGTAACTGGCTTCAGCAGAGCGCAGATACCAAATACT
GTTCTTCTAGTGTAGCCGTAGTTAGGCCACCACTTCAAGAACTCTGTAGCACCGCCT
ACATACCTCGCTCTGCTAATCCTGTTACCAGTGGCTGCTGCCAGTGGCGATAAGTCG
TGTCTTACCGGGTTGGACTCAAGACGATAGTTACCGGATAAGGCGCAGCGGTCGGG
CTGAACGGGGGGTTCGTGCACACAGCCCAGCTTGGAGCGAACGACCTACACCGAAC
TGAGATACCTACAGCGTGAGCTATGAGAAAGCGCCACGCTTCCCGAAGGGAGAAAG
GCGGACAGGTATCCGGTAAGCGGCAGGGTCGGAACAGGAGAGCGCACGAGGGAGC
TTCCAGGGGGAAACGCCTGGTATCTTTATAGTCCTGTCGGGTTTCGCCACCTCTGACT
TGAGCGTCGATTTTTGTGATGCTCGTCAGGGGGGCGGAGCCTATGGAAAAACGCCA
GCAACGCGGCCTTTTTACGGTTCCTGGCCTTTTGCTGGCCTTTTGCTCACATGTTCTTT
CCTGCGTTATCCCCTGATTCTGTGGATAACCGTATTACCGCCTTTGAGTGAGCTGATA
CCGCTCGCCGCAGCCGAACGACCGAGCGCAGCGAGTCAGTGAGCGAGGAAGCGGA
AG

Appendix D

Integration by Gamma Irradiation

Potential parent strains for integration were created by microinjecting pBYU1 and pCR4_TOPO_ben-1 supercoiled plasmids at a ratio of 10:1 in 10mM Tris pH 7.4. Stable lines were identified by pharyngeal GFP expression and maintained by picking individual animals. A single line, with a transmission rate of about 35% was selected for the integration. Synchronized L4 animals were obtained by bleaching gravid animals in fresh 20% bleach 1 M NaOH solution for several minutes (~3-5 min) until worms began to dissolve. Eggs were washed 3 times to remove residual bleach and then hatched overnight in M9 with gentle rocking at 25 °C. Synchronized L1 larvae were spread onto new NGM plates and incubated at 20 °C for 48 hours until animals had reached L4 stage. Approximately 100 transgenic L4 animals were picked onto a clean NGM plate and irradiated with 3800 rads of gamma rays (~23 minutes). See equation below.

Radioactive Decay Equation

$$A_t = A_o\, e^{-0.693t/T}$$

A_t= activity at a given time
A_o= initial activity R/min
t= elapsed time 17.526 yrs
T= Cs 137 half-life 30.07 yrs
R= roentgens
* 1 R ~ 0.95 rads on soft tissue

Irradiated L4 animals were picked to seeded NGM plates (5-10 animals per plate). Adult animals were maintained at 25 °C moved twice a day for 4 days. 400 F_1 animals were singled to individual plates from egg plates corresponding to days 2-4 post irradiation. 150 from day 2, 300 from day 3, and 150 from day 4. These plates were allowed to grow up and the subsequent population screened for higher incidence of GFP expression in the progeny. 2-3 animals were singled from plates exhibiting higher extrachromosomal transmission rates than the parent strain, allowed to grow up, and checked for homozygosity. This produced 17 integrated lines for an integration rate of 2.8%.

sgRNA synthesis
Synthesis of sgRNAs was done by following the manufacture's protocol for MEGAshortscript T7™ kit from ThermoFisher. Templates for use with the MEGAshortscript T7™ kit were generated by PCR of plasmid px330 (from Feng Zhang lab) with reverse primer oSMJ009 and a unique forward primer. Forward primers 5' > 3' had 21bp T7 promoter, CRISPR targeting sequence generated using CRISPRdirect[30], and 21bp for px330 priming. See Alder Lab Protocol Appendix D. sgRNA was extracted with phenol/chloroform and resuspended in RNAse-free water. RNA concentrations were measured using a NanoDrop Lite™ from ThermoFisher.

Testing in vitro CRISPR/Cas9 activity
Activity of CRISPR/Cas9 was confirmed by allowing sgRNA to complex with Cas9 at room temperature and then incubate with an appropriate template at 37 °C for 1 hour. Digests were visualized on 1.5% agarose gel.

CRISPR/Cas9 injection mixes
Concentrated Cas9 will precipitate quickly if the ionic strength of the solution becomes to weak. When preparing injection mixes, sgRNAs in RNase-free water should be added last. We maintained a 2:1 ratio by mass of Cas9 (obtained from Alder Lab) to sgRNA sg006 for CRISPR/Cas9 integration experiments. Final concentrations in the injection mixes were as follows: 0.8 ug/uL Cas9, 4mM Tris pH 8, 6mM HEPES pH 7.4, 200mM KCl, 2mM $MgCl_2$, 5% glycerol, a cumulative 0.4 ug/uL of sgRNA. Injection mixes were prepared fresh immediately prior to use.

Microinjection and recovery
CRISPR/Cas9 injection mixes are more viscous than standard injection mixes prepared in 10mM Tris pH 7. Consequently, all capillaries must be thoroughly cleaned to remove any debris before using them for needle pulling. (This is also highly recommended for standard injections as well since it greatly reduces the incidence of needle clogs.) Slightly larger young adults (~8 eggs present) recover better. 1-5 worms were mounted and then microinjected in both syncytial gonad arms and filled until injection fluid turned/reached the zone of transition. Injected worms were recovered directly on the agarose pad in 6 uL of Recovery Buffer for 10 minutes. After 10 minutes, 2 uL of M9 was added every 5 minutes until 30 minutes had elapsed. Recovered worms were then picked onto a seeded plate with 20 uL of M9. Recovering worms were incubated overnight at 16 °C and then moved to new seed plates.

Capillary cleaning and needle pulling
Filamented capillaries (Narishige GD-1), were washed by vortexing 15-20 in a 50 mL conical tube with 30 mL of cleaning solution. Cleaning solution was 3% dishsoap and 17% ethanol in Mili-Q water. 1 mL of fresh cleaning solution was passed through each capillary by cutting a p1000 pipet tip to accommodate the capillary, attaching it to a 10 mL pipet, and using a pipet aid. Capillaries were then rinsed with new Mili-Q water first by vortexing and then passing 2 mL of new Mili-Q water through them. Capillaries were dried in a loosely capped 50 mL conical tube at 37 °C for 24 hrs.

Needles were pulled using a Narishige PC-10. Needles were inserted halfway and secured. Needles were pulled in one stage with a heater setting of 55.0 and all weights attached. Both resulting needles were taken and stored in a covered container for later use.

Testing Cas9 activity in vitro (Alder Lab Protocol)

10X Cas9 buffer

200mM HEPES
1M NaCl
50mM MgCl2
1mM EDTA
pH 6.5

PCR target

Order primers that will amplify ~500 bps flanking target site. Perform ~4 50uL PCRs and column purify PCR product and nanodrop to determine concentration

sgRNA

Generate template for in vitro transcription. Use px330 (from Feng Zhang lab).
Forward Primer – T7 promoter, guide sequence (example is for GFP), binding to pX330
TGTAATACGACTCACTATAGGGGGCGAGGAGCTGTTCACCGgttttagagctagaaatag
c
Reverse Primer –
AAAAAAGCACCGACTCGGTGCCACT

PCR Condition
Phusion HF buffer
98- 30 sec
98-30 sec | 34X
72- 20 sec |
72-1 min
Column purify PCR product and use as template for in vitro transcription using MEGAshortscript T7 kit from Ambion (life tech now I think). Typically yield is 1.5 ug/uL RNA.

1) Thaw Cas9 protein and sgRNA on ice.
2) Prepare 4 reactions in 1X Cas9 buffer
    a. 800ng PCR
    b. 800ng PCR + 1uL sgRNA
    c. 800ng PCR + 1uL Cas9
    d. 800ng PCR + 1uL sgRNA + 1uL Cas9 (sgRNA and Cas9 are premixed and incubated at RT for 5minute prior to adding)
3) Incubate at 37C for 1 hour then run on gel

Note: Our experience is that Cas9 solubility is highly dependent on the total ionic strength of the buffer. If you have concentrated Cas9 to >10mg/mL, add KCL to sgRNA prior to mixing, otherwise the Cas9 will immediately precipitate when you mix them. The final KCL concentration should be around 500mM.