CHARACTERIZATION OF CAS9-MEDIATED MICRORNA KNOCKOUT

A Thesis

by

ANDREW MORIN

Submitted to the Office of Graduate and Professional Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

MASTERS OF SCIENCE

Chair of Committee,	Beiyan Zhou
Committee Members,	Stephen Safe
	Michael Golding
Head of Department,	John Stallone

December 2015

Major Subject: Biomedical Sciences

ABSTRACT

MicroRNAs are short regulatory RNAs that primarily operate at the post-transcriptional level, acting as part of the RISC complex to destabilize mRNA prior to, or during, translation. MicroRNAs are subject to several levels of processing by Drosha and Dicer prior to achievement of the active mature form, this processing being highly dependent on the secondary structure of the microRNA transcript. As such, microRNA biogenesis is highly sensitive to mutations. In this study, it was attempted to optimize Cas9-mediated genome editing techniques for microRNA knockout for two hematopoietic microRNAs, miR-142 and miR-223.

While it has been previously demonstrated that Cas9 is capable of inhibiting microRNA biogenesis, the results have been characterized by variable and low-level efficiencies currently inherent to the system. This study attempted to rectify this shortcoming by bombarding the small pre-miR transcript with a combination of multiple independently targeted Cas9 complexes, using lentiviral delivery methods, measuring for evidence of impaired microRNA biogenesis by real-time quantitative PCR (RT-qPCR). The results suggest that there may be specific regions of the microRNA site in which mutations are more deleterious to function, but that further study is required. Additionally, it was determined that the use of multiple sgRNAs is not experimentally useful, proving unable to and in some cases, an impediment to successful microRNA knockout.

DEDICATION

This work is dedicated to the many science fiction authors who have inspired me to a career in the biological sciences, and the march to free humanity from the chains of merely natural selection. *Scientia omnia vincit*.

ACKNOWLEDGEMENTS

I would like to thank my committee chair, Dr. Zhou, and my committee members, Dr. Safe, and Dr. Golding, for their guidance and support throughout the course of this research. Thanks also go to the members of the Zhou research group, past and present, for their help and companionship through this period.

TABLE OF CONTENTS

		Page
ABSTRACT		ii
DEDICATION		iii
ACKNOWLEDGEMENTS		iv
TABLE OF CONTENTS		v
LIST OF FIGURES		vi
LIST OF TABLES		vii
1.	INTRODUCTION	1
2.	METHODS	5
	2.1 Construct Design	5
	2.2 pX458 and pSICHECK2 Transfection	5
	2.3 Lentiviral Transduction	6
	2.4 RT-qPCR miRNA Detection and Data Analysis	7
3.	RESULTS	8
	3.1 *In Vitro* Plasmid Disruption	8
	3.2 DNA Damage Assessment	11
	3.3 miRNA Loss of Function Assessment	15
4.	CONCLUSIONS	22
REFERENCES		25

LIST OF FIGURES

FIGURE		Page
1	Design of sgRNAs and pSICHECK system	9
2	Cas9-induced Renilla luciferase down-regulation	11
3	HEK293T transfection and genomic DNA mutation assessment	14
4	EML miR-142-3p expression for one sgRNA	17
5	RAW264.7 miR-223-3p expression for one sgRNA	18
6	Effects of adding multiple anti-miR-142 sgRNAs to EML culture	20
7	Effects of adding multiple anti-miR-223 sgRNAs to RAW264.7 culture	21

LIST OF TABLES

TABLE — Page

1 sgRNA sequences 10

1. INTRODUCTION

MicroRNAs are a class of short (20-22 bases) non-coding regulatory RNAs believed to act primarily at the post-transcriptional level, functioning as the target-specifying element of the RNA Induced Silencing Complex (RISC) [1]. MiRNA biogenesis is a multi-step process, heavily reliant on the secondary structure of the maturing RNA for proper recognition and processing [2, 3]. The primary miRNA (pri-miR) is transcribed as an oligonucleotide of variable length containing one or more precursor miRNAs (pre-miR) as bulged hairpin structures [4]. Each pre-miR is cleaved from the pri-miR at a highly defined point 11 bp from the 5' base of the stem, and the 65-100 base pre-miR is exported from the nucleus, where it is recognized and processed by Dicer [5]. Dicer activity leaves a 20-22 bp dsRNA molecule which is loaded into the RISC complex with one strand being the retained guide strand, and the other being the ejected (and presumably degraded) passenger strand [6]. The RISC-associated guide miRNA then assumes its main function as a post-transcriptional regulatory element [7], binding to and destabilizing mRNAs that contain 3' UTR sites partially complementary to the RISC guide RNA [8, 9]. This partial complementarity is characterized by a 6-8 bp 5' "seed" region, a short non-complementary bulge, and a second region of complementarity stretching the remainder of the miRNA [10, 11].

MiRNA function is reliant on the primary sequence, both for accurate targeting of the RISC complex, and the biogenesis of the miRNA [12]. Both Drosha and Dicer recognize their respective miRNA substrates by the primary sequence-dependent

secondary structures [11, 13]. Additionally, each pre-miRNA contains the guide and passenger strand sections, disruption of which could lead to aberrant selection and loading into the RISC complex, or mistargeting of the RISC complex by mutations within the guide sequence [10]. Such mutations may render the RISC unable to bind the normal mRNA target UTR sites, direct it to other mRNAs, or convert the miRNA into a siRNA (which acts to cleave,rather than destabilize an mRNA) by enabling complete complementarity between the guide strand and the mRNA.

Cas9 is the nucleolytic component of the type II CRISPR (Clustered Regularly Interspaced Short Palindromic Repeat) system [14], and composed of the Cas9 nuclease, a variable-sequence 18-22 base ssRNA guide oligo (sgRNA) that defines the target dsDNA, and a 70-100 base ssRNA tracrRNA, which acts as a scaffold for the nuclease [15, 16]. This system has rapidly become the primary means of editing DNA, primarily due to the ease with which the complex can be re-targeted, requiring only the design of a new sgRNA [17]. The set of sequences that can be targeted by the Cas9 system is determined by the protospacer-adjacent motif (PAM), a short sequence on the DNA recognized by the protein component, and varying by species [18, 19]. spCas9, the most common complex used, recognizes the 'NGG' PAM, granting an extremely permissive spread in sgRNA design [20]. Both PAM recognition and base pairing between the sgRNA and DNA are required for Cas9 nucleolytic activity [21], generating a blunt double-stranded break in the target DNA molecule. Binding does not necessitate nuclease activity [22], and the complex exhibits high affinity for its target [23], a feature that can be exploited by using nuclease-deficient variants of Cas9 as transcriptional

inhibitors [24], or as the targeting component for other effectors fused to the complex [25]. Once a double-stranded break has been induced, it may be repaired by one of several cellular DNA damage pathways, the two primary methods being homology-directed repair (HDR) [26] and non-homologous end-joining (NHEJ) [27]. While HDR yields a faithful repair, it requires the availability of a homologous allele, and as such is only active during the G2 and S phases [28]. NHEJ is the primary repair pathway, active throughout all cell cycles [29, 30]. However, NHEJ is error-prone, typically generating small insertions or deletions (in/dels) at the repair site [31, 32].

It has already been demonstrated that a single sgRNA/Cas9 complex can lower the expression of a miRNA[33], and that even single-base alterations at the 5' Drosha processing site can completely abrogate miRNA biogenesis [34]. However, it is not known whether this is a property active during Drosha recognition, applicable to all processing steps, or whether in/dels in other locations may be equally efficacious. Additionally, the Cas9 complex shares the drawbacks of more traditional DNA editing systems (TALENs and zinc-finger nucleases), exhibiting a variable lack of both specificity and nuclease activity [35, 36]. Cas9 systems require high expression for acceptable levels of editing [23], which in turn leads to increases in off-target editing [25]. Furthermore, there is significant variability between different sgRNA sequences [17], not ascribable to chromatin or methylation status [37]. While chemically modified sg/tracrRNA oligos possessing 5' and 3' protective add-ons have been recently show to greatly improve Cas9 activity and consistency [38], these options are not available for many of the current delivery methods (virions, protein/DNA nanocages, naked DNA

delivery, etc.) [39-41], which require transcription by the target cell. Given the attractiveness of Cas9-based systems for genome editing, the highly variable and overall low efficiency of Cas9 is an obstacle in the path of gene therapy research, particularly in such areas as hematopoietic stem cell editing, in which high efficiency is paramount for effective transfusions [42, 43]. In the context of protein-coding genes, these difficulties may be overcome through the use of several sgRNAs, spaced over the available kilobases of gene [44]. However, the short stretch of a pre-miR sequence does not permit luxurious spacing with a multiplex approach. The interaction of multiple co-occurring Cas9 complexes in close proximity has not been specifically investigated, except for indirect observations gleaned from the use of a dual FokI-dCas9 fusion system (capable of reducing off-target editing at the cost of efficiency) [17, 45]. However, in that case, it is not known if the reduced efficiency of the two-sgRNA system is a result of complex proximity, or the altered protein, or lowered overall expression.

Accordingly, the intent of this study was to investigate the use of Cas9 to induce loss of function in two hematopoietic but functionally unrelated miRNAs, miR-142 [46] and miR-223 [47], with two specific aims. First, to determine if it is either necessary or beneficial to confine Cas9-generated in/dels to particular regions of the pre-miRNA secondary structure. The alternative hypothesis would be that any mutation is capable of causing loss of function, either by disruption of biogenesis or by altering the binding sequence of the assembled RISC. Second, to determine whether the low efficiencies of individual sgRNAs can act additively when multiple sgRNAs are co-delivered, or if the short sequence length proves an impediment to such brute force approaches.

2. METHODS

2.1 Construct Design

pSpCas9(BB)-2A-GFP (pX458, Addgene plasmid #48138) and lentiCRISPRv2 (Addgene plasmid #52961) were gifts from Feng Zhang, through Addgene. pSICHECK was obtained from Promega. All sgRNA oligos were designed using CRISPR Design Tool (http://crispr.mit.edu), and ordered from IDT. As both spCas9 constructs express the tracrRNA component as a single oligo with the inserted sgRNA oligo, design of the tracrRNA was not necessary. All seventeen guides were cloned into both the pX458 and lentiCRISPRv2 plasmids using Mix & Go 5α chemically competent cells (Zymo), and insertion verified by end-point PCR, restriction enzyme digestion, and sequencing. Pre-miR-142b (comprising the true pre-miR-142 and the extended Drosha recognition region) was PCR amplified from murine (C57BL/6) gDNA, and cloned into the MCS of pSICHECK2 using the same method as above. All PCR primers used throughout this study were designed using NCBI PrimerBLAST.

2.2 pX458 and pSICHECK2 Transfection

HEK293T cells were grown in IMDM (Hyclone) supplemented with 10% fetal bovine serum, at 37°C and 5% CO_2, and grown to ~70% confluency in 6-well plates. The pX458 contructs and pSICHECK constructs were transfected at a 1.5/1 molar ratio in 96-well plates using X-tremeGENE HP DNA transfection reagent (Roche) according to the manufacturer's protocols. The transfection reagent was removed after six hours, and the luciferase data obtained at forty-eight hours using the Dual-Glo Luciferase

Assay System (Promega), according to the manufacturer's protocols. To collect gDNA for analysis of spCas9 nuclease activity, the pX458 constructs were transfected into HEK293T cells in the same manner as before, and the DNA collected using the GenElute Mammalian Genomic DNA Miniprep kit (Sigma-Aldrich). The ~300 bp surrounding the pre-miR-142 region was PCR amplified, and the product purified by agarose gel extraction. Sequencing data was analyzed with the web-based TIDE tool (http://tide.nki.nl).

2.3 Lentiviral Transduction

Lentiviral transduction virions were generated using HEK293T cells, transfected in the same manner as before (quantities adjusted to 2.5 μg total DNA/well in 6-well plates), using the seventeen lentiCRISPRv2 constructs (ten targeting miR-142, seven targeting miR-223), the psPAX2 packaging plasmid (a gift from Didler Trono, Addgene plasmid #12260), and a VSV-G expression construct, at an equimolar ratio. The transfection reagent was removed after six hours, and the virion-containing media collected at forty-eight and seventy-two hours, having been stored at 4°C. The two collection fractions were combined, brought to 0.4M NaCl and 8.5% w/v PEG6000, and rotated for ninety minutes at 4°C. The pegylated virions were then centrifuged at 10,000 rcf for fifteen minutes, and the viral pellet re-suspended at a 30X concentration. The concentrated virus was then transduced into EML and RAW 264.7 cells, as appropriate.

EML cells were grown in IMDM (Hyclone) supplemented with 20% fetal bovine serum and 15% BHK/MKL-conditioned media, at 37°C and 5% CO_2. RAW264.7 cells were grown in IMDM (Hyclone) supplemented with 10% fetal bovine serum, at 37°C

and 5% CO_2. For both cell types, 50,000 cells were transduced at an approximate MOI of ten, in the presence of 6 μg/mL polybrene, and centrifuged at 800 rcf for 75 minutes at 30°C immediately after introduction of the virion-containing media. The media was replaced after twelve hours, and total RNA collected after forty-eight hours using the RiboZol RNA Extraction Reagent (Amresco) and the Direct-Zol RNA Miniprep Kit (Zymo). MOI was measured by titration of the viral concentrate into HEK293T cells, and selecting for infection events by 2 μg/mL puromycin.

2.4 RT-qPCR miRNA Detection and Data Analysis

From the total RNA extraction, cDNA was generated using the Universal cDNA Synthesis Kit (Exiqon), and subjected to RT-qPCR using the iTaq Universal Sybr Green One-Step Kit (Bio-Rad) and the appropriate Exiqon LNA miRNA primers. Output was collected on Bio-Rad CFX Manager, and analyzed with Prism 6.0a by one-way ANOVA or unpaired t-tests, as appropriate. Normalization was conducted using primers against miR-103a, and the Exiqon UniSp6 spike-in control. Significance values reported are for multiple comparison t-tests vs. the experimental control values.

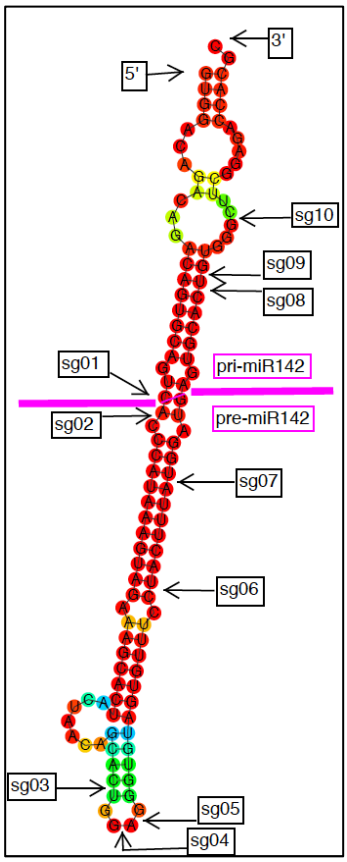
3. RESULTS

3.1 In Vitro Plasmid Disruption

In order to determine the variability of efficiency for the available suite of sgRNAs, 10 sgRNAs for mmu-pre-miR-142 were designed and cloned into the pX458 transfection plasmid, expressing spCas9-eGFP and a fused sgRNA/tracrRNA construct driven by the U6 promoter (demonstrated to be indistinguishable in function from a wild-type two-oligonucleotide system) (Figure 1). Additionally, the mmu-pre-miR-142 sequence was PCR amplified from C57BL/6 mice, and inserted into the multiple cloning site of the pSICHECK2 plasmid, downstream of a renilla luciferase reporter, and upstream of a synthetic poly-A tail. As diagramed in Figure 1A, the actual miRNA sequence used was larger than the strictly defined pre-miRNA, extending into the Drosha binding and recognition region of the pri-miRNA, but will be referred to as the pre-miRNA-142 for convenience. Since Cas9 can inhibit transcription by tightly binding substrate DNA, it was hypothesized that the relative effectiveness of the Cas9 sgRNAs could be determined by observing decreased levels of luciferase expression, via either disruption of transcription, or degradation of the plasmid by exposure to exonucleases. The pX458-sgRNA plasmids was co-transfected with the pSICHECK-pre-miR142 plasmid into HEK293T cells, and luciferase expression assayed after 48 hours. Uniformly, the presence of Cas9 and an sgRNA reduced luciferase expression to between 40% to 63% of the no-sgRNA control (Figure 2A). While this limited variability was not remarkable, given that the effects did not rely on the specific location

of binding, it demonstrated that there was a degree of heterogeneity in binding/cleaving efficiency. Interestingly, the addition of the pre-miR-142 sequence to the pSICHECK2 3' UTR consistently increased luciferase expression by an average of 50.6% in the absence of a targeting sgRNA (Figure 2B).

A



B

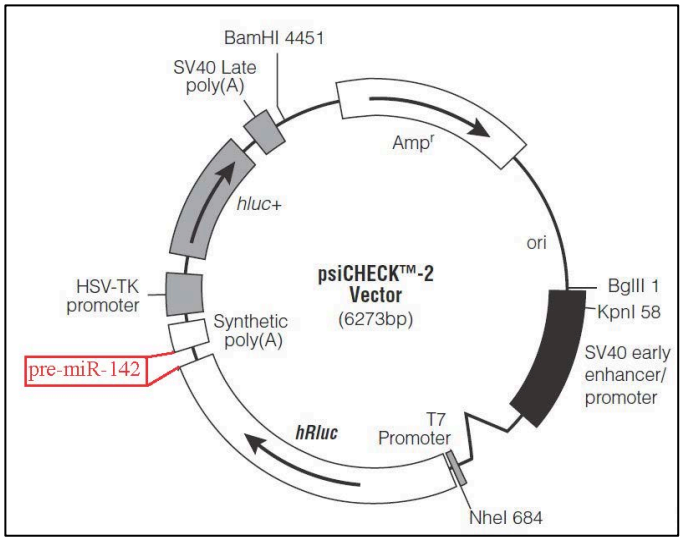


Figure 1: Design of sgRNAs and pSICHECK system. (A) Positioning of Cas9 cut sites across the secondary structure of the pre-miR-142 and Drosha recognition region. Arrows indicate the specific sites of potential mismatch in the event of double-stranded break mutation. (B) Design of the pSICHECK2-pre-miR-142 construct, with the mmu-pre-miR-142 inserted between the hRluc reporter and the synthetic poly-A tail. Base figure provided by Promega.

sgRNA	Sequence	Cut position
miR-142 sg01	AGTAGTGCTTTCTACTTTAT	30
miR-142 sg02	TAGTAGTGCTTTCTACTTTA	32
miR-142 sg03	GAAAGCACTACTAACAGCAC	53
miR-142 sg04	AGCACTACTAACAGCACTGG	56
miR-142 sg05	GCACTACTAACAGCACTGGA	57
miR-142 sg06	AGTGCACTCATCCATAAAGT	79
miR-142 sg07	GTGTAGTGTTTCCTACTTTA	79
miR-142 sg08	CTTTATGGATGAGTGCACTG	94
miR-142 sg09	TTTATGGATGAGTGCACTGT	95
miR-142 sg10	GATGAGTGCACTGTGGGCTT	101
miR-223 sg01	ACGGAGCGTGACACTGCAGA	10
miR-223 sg02	GTGTATTTGACAAGCTGAGT	45
miR-223 sg03	GCTGAGTTGGACACTCTGTG	58
miR-223 sg04	ATAGGCATGAGCCACACTTG	90
miR-223 sg05	TGATAGGCATGAGCCACACT	92
miR-223 sg06	TTTGTCAAATACCCCAAGTG	92
miR-223 sg07	GATAGGCATGAGCCACACTT	91

Table 1: sgRNA sequences. Complete set of 20-base sgRNA sequences as generated by CRISPR Design Tool, with the nucleolytic site location on the relevant pre-miR, counting 5' to 3'. In the cases in which two sgRNAs are marked as possessing different sequences, but the same cleave point, this is an indicator of the two sgRNAs binding opposite strands of the target DNA, and happening to overlap exactly on the site of nuclease activity.

A B

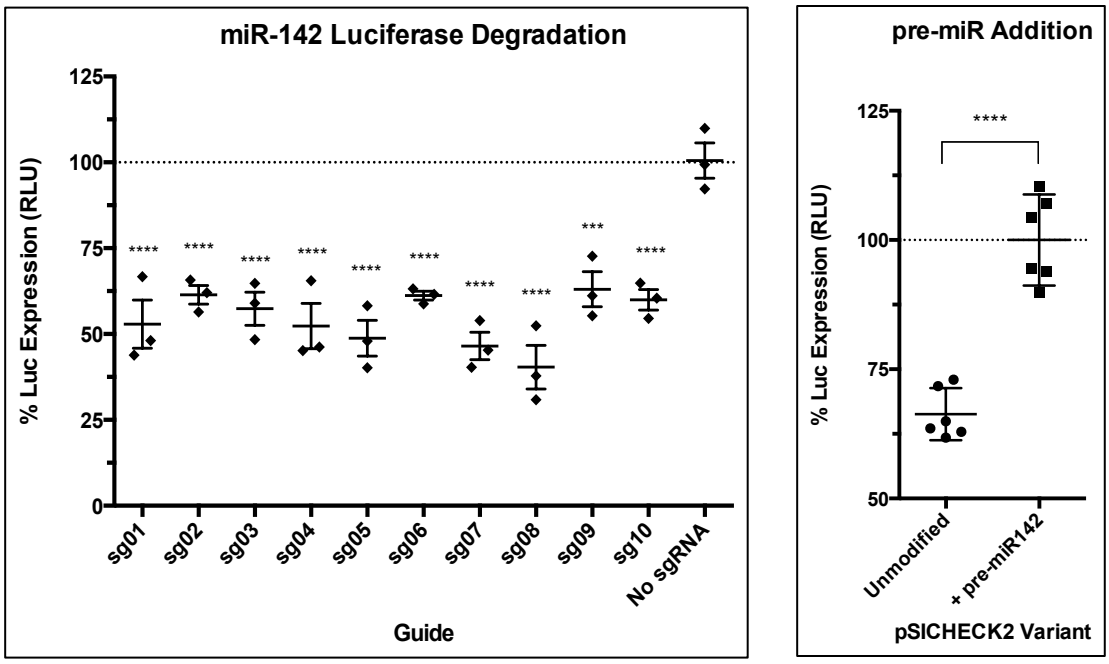


Figure 2: Cas9-induced Renilla luciferase down-regulation. (A) Relative Renilla luciferase expression in HEK293T cells 48 hours post-transfection (pX458 with a single sgRNA, or no-sgRNA control, and pre-miR-142-pSICHECK2), normalized against firefly luciferase expression. Error bars displayed indicate mean and SEM for three replicates. (B) Note the reduced expression of the unaltered pSICHECK2 vs. the pre-miR-142-pSICHECK2. ★★★ = $p \leq 0.001$; ★★★★ = $p \leq 0.0001$, by one-way ANOVA and multiple comparison vs. control.

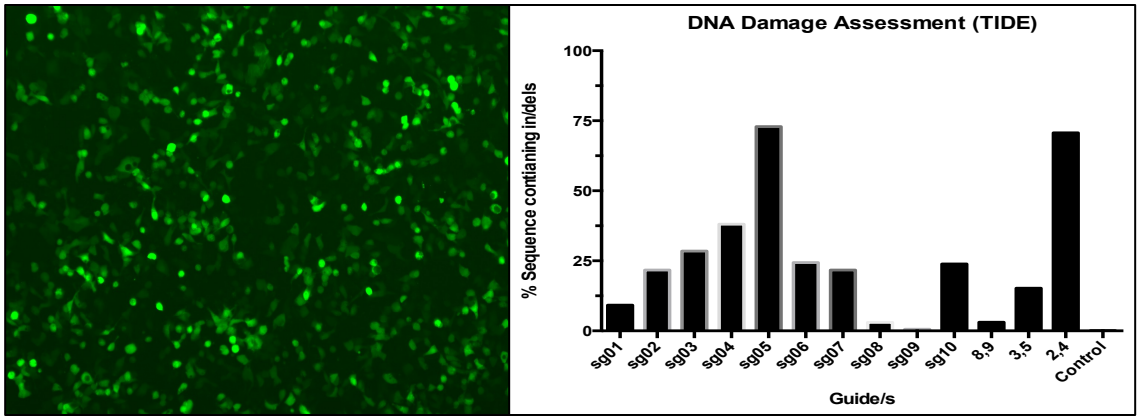
3.2 DNA Damage Assessment

While assessment of Cas9 activity in down-regulating luciferase expression provided reassurance that the system was functional, it did not provide data on whether this was a result of nuclease activity, or merely bound but inactive complexes interfering with transcription. It was also a concern that Cas9 activity against a plasmid may not reflect activity against genomic DNA. Accordingly, the ten miR-142 guides were tested

by transfecting HEK293T cells with the sgRNA-Cas9 constructs, individually and in pairs, verifying consistent transfection efficiency by eGFP expression, and collecting genomic DNA after 48 hours. PCR primers were designed against the gDNA using NCBI PrimerBLAST to flank the pre-miR-142 sequence by approximately 300 bp up- and downstream, and used to both amplify the ~700 bp sequence from the gDNA, and sequence the amplified material. The resulting chromatographs were then analyzed via TIDE (Tracking of In/dels by DEcomposition), a DNA heterogeneity assessment tool designed and made available by the Bas van Steensel research group (Netherlands Cancer Institute) (Figure 3). Three items of note were observed. There was a lack of consistency between the luciferase experimental data and the frequency of actual gDNA mutation, the abrogation of mutation induced by a single base mismatch between the target and guide, and the differing effects of combining multiple guides. Although all constructs generated similar levels of luciferase inhibition, there was a wide difference between efficiencies when confronted by gDNA, ranging from 9.2% to 72.9%. A broad pattern was evident as to particular regions being more or less favorable for mutation, with the terminal loop being the most frequently mutated, and decreasing towards the Drosha cleavage point and beyond. This pointed towards the explanations that either anti-plasmid efficiency is drastically different from anti-gDNA efficiency, or else the dominant effect upon the pSICHECK2 system was steric transcription blockage. However, whether or not the pattern possesses any significance was still unknown, since HEK293T cells do not express miR-142, and thus gDNA lesion frequency could not be correlated with variation in expression of the mature miRNA within this system. Second,

single base mismatches between the sgRNA and the target (due to sequence inconsistencies between the human and murine miR-142) were sufficient to lower the nuclease activity by one to two orders of magnitude, as evidenced by sg08 (single mismatch 6 bases from PAM, 0.6% efficiency), and sg09 (single mismatch 5 bp from PAM, 3.0% efficiency). This result is entirely in accordance with the current understanding of Cas9 target binding and cleaving mechanics, which stipulate that mismatches far from the PAM may have little effect, but that there is a critical seed region. This region is generally defined as the 5-12 bases proximal to the PAM (the exact boundary is still a matter scientific dispute), which is essential for proper binding activity. Additionally, when co-transfected, neither mismatched complex was capable of rescuing an efficient degree of nucleolytic activity. Two other pairs of complexes were co-transfected, sg02/sg04 (separated by 25 bp) and sg03/sg05 (separated by 4 bp). Notably, the first pair (sg02/sg04) appeared to act additively, generating a combined in/del rate of 70.7%, compared to 21.7% and 38.1% respectively, when delivered individually. However, the spatially constricted (overlapping binding regions) sg03/sg05 pair generated a mutation level of 15.7%, less than the level of either individual sgRNA (28.5% and 72.9%, respectively), indicative of the two complexes engaging in mutual inhibition.

A B



C

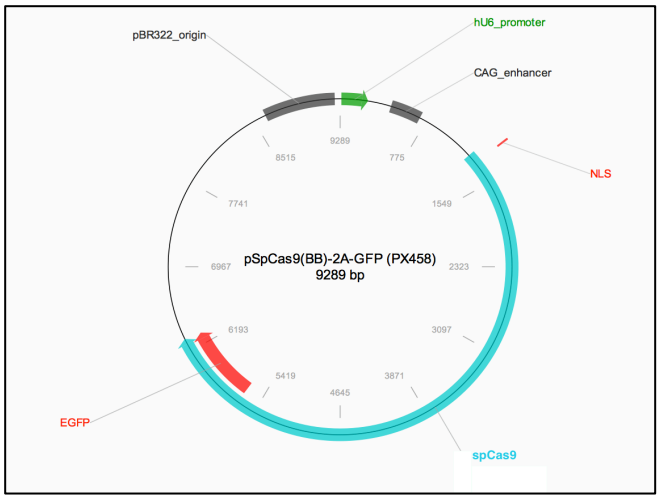


Figure 3: HEK293T transfection and genomic DNA mutation assessment. (A) Representative image of eGFP expression for a HEK293T ~90% confluent well, transfected with sgRNA-pX458. (B) Mutation frequency for each sgRNA-pX458 and selected co-transfected pairs, as measured by TIDE by analysis of chromatograph base call certainty up and downstream (sequencing-wise) of a given predicted cut site. As the sgRNAs are labeled 5' to 3' by location on the pre-miR structure, a pattern can be observed, as mutation frequency builds from the Drosha cleavage site (sg01 and 02), comes to an apex at the terminal loop (sg05), and then drops again moving back down the stem. One the right side, the spatially overlapping 03/05 pair appear to inhibit each other, whereas the well separated 02/04 pair appear to function additively. (C) Plasmid map of the spCas9 expression sgRNA-pX458 construct.

3.3 miRNA Loss of Function Assessment

While HEK293T cells provide an environment for assessing the capabilities of the suite of sgRNAs to generate gDNA lesions, the lack of endogenous expression of either miR-142 or miR-223 render it impossible to verify that the damage has any effect on expression of mature, functional miRNA. Accordingly, for miR-142 expression, I turned to the EML cell line (a hematopoietic progenitor model), and for miR-223, to the RAW264.7 cell line (a macrophage model). Seven sgRNAs were designed for pre-miR-223 as described for pre-miR-142. However, since both lines were not amenable to standard transfection methods, I used a replication-incompetent VSV-G lentivirus, carrying spCas9 and the sg/tracrRNA hybrid as in the previous experiments, and substituted puromycin resistance for eGFP. While it would have been desirable to use a non-integrating vector, integrase-deficient retroviruses are known for poorly expressing their payload (which is inimical to Cas9 activity). Adenoviruses are poor transducers of the hematopoietic stem cells intended to be represented by the EML line, and the size of the spCas9 and U6/sgRNA cassettes is prohibitive of smaller vectors. Thus, the lentiviral constructs were generated, used to form infectious virions in HEK293T cells according to standard viral generation protocols [48], and spin-transduced into the miRNA-expressing target cells (all miR-142 sgRNAs into EML, all miR-223 sgRNAs into RAW264.7). After 48 hours (approximately 2.5 cell doublings, permitting sufficient time for the integrated spCas9 to afflict the miRNA locus), total RNA was collected and assayed for the mature miRNA by reverse-transcriptase quantitative PCR (RT-qPCR) (Figure 4, 5). sgRNAs were delivered singly, and also in sets of two and three. In

general, there was minimal enhancement from the use of more than one sgRNA per population, multiple sgRNAs acting either to inhibit the total loss of function effect in the case of miR-142-3p, or displaying no significant effect in the case of miR-223-3p. In fact, the combination of two poorly performing sgRNAs (sg03 and sg04) against miR-223 appeared to increase miR-223-3p expression over that of the no-sgRNA control, and the pairing of the inefficient sg03 with more efficient sg07 completely rescued miR-223-3p function. This was unexpected given that the two are separated by 32 bases, bind separate strands and are thus oriented away from each other. However, for both miRNAs, transduction of a single Cas9-sgRNA was sufficient to strongly and specifically reduce miRNA expression; no sgRNA construct significantly affecting the expression of the unrelated miR-103a (data not shown). Additionally, sgRNAs against miR-142 were unable to affect expression of miR-223, and vice versa. This was unsurprising, given that the two miRNAs are unrelated and located on different chromosomes (chr11 for miR-142, and chrX for miR-223).

No statistically significant correlation was found between miRNA loss of function in EML cells (Figure 4) and the HEK293T DNA damage data (Figure 3B). It was noted that both of the sgRNAs containing a mismatch to human miR-142 performed well when targeting the murine EML gDNA, one in fact generating the best loss of function effect (sg08, with miR-142-3p expression reduced to 3.6% on average of the control level). Of the ten sgRNAs tested for miR-142, only one (sg06, targeting the 3p miRNA itself) was incapable of dropping miRNA expression below 50%. This could be interpreted as an artifact of the RT-qPCR detection method (picking up signal from

miRNAs not sufficiently mutated to prohibit amplification), or it may indicate that sites not directly interfacing with Drosha/Dicer machinery are poor choices if complete

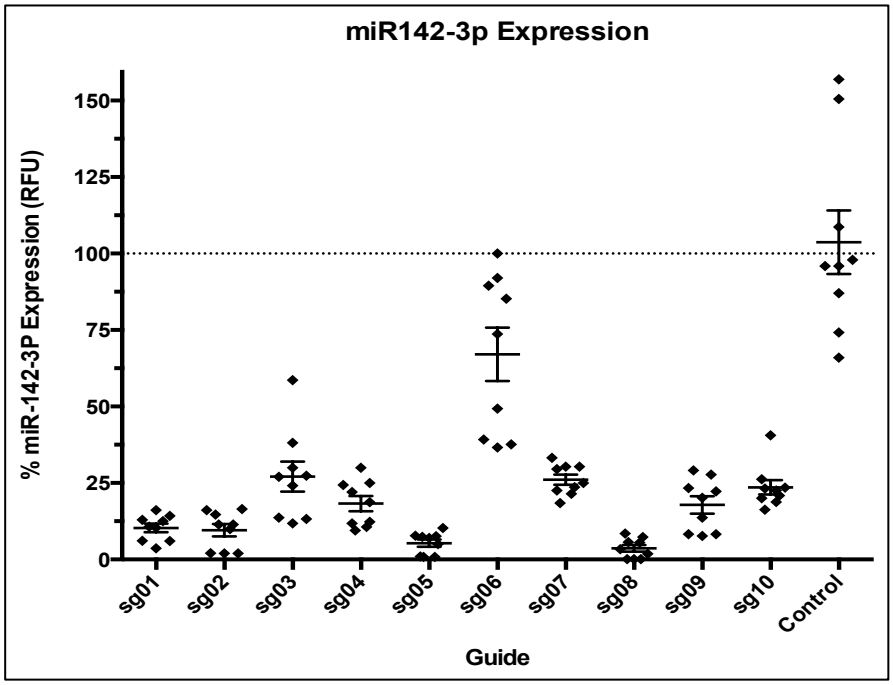


Figure 4: EML miR-142-3p expression for one sgRNA. Fig 4. Expression of miR-142-3p in EML cells as measured by RT-qPCR 48 hours post-transduction. Error bars display mean and SEM, expression normalized to define relative expression as a percentage compared to the no-sgRNA control. All expression results for this figure are significant at $p \leq 0.0001$ by one-way ANOVA and multiple comparison vs. control. Normalization was conducted using miR-103a and the Exiqon UniSp6 spike-in control.

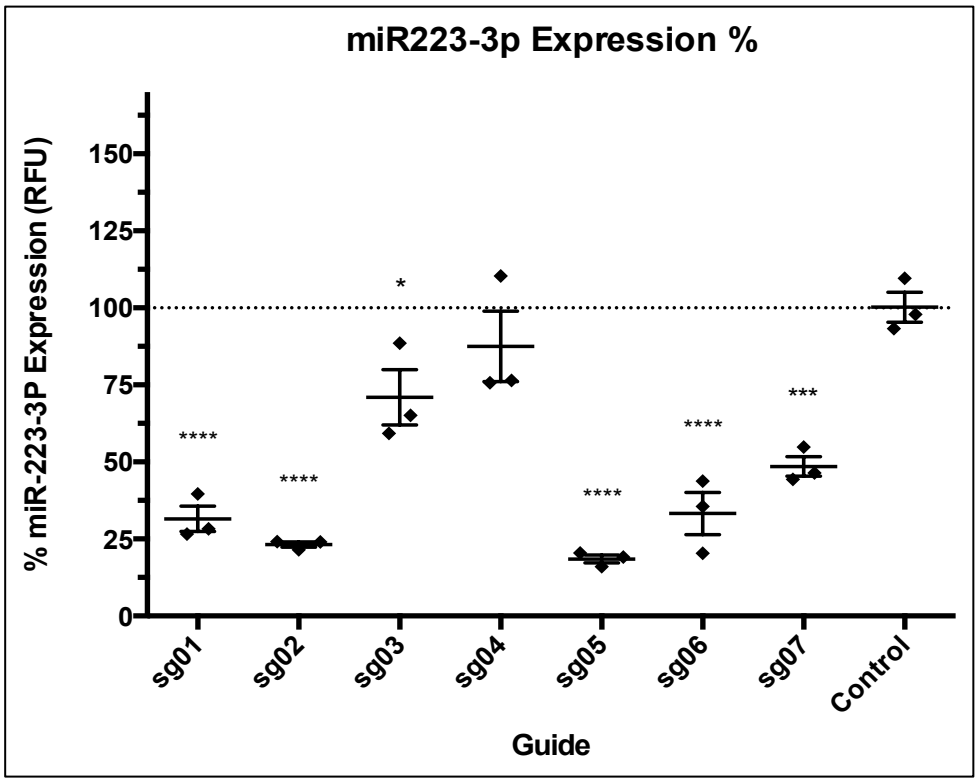
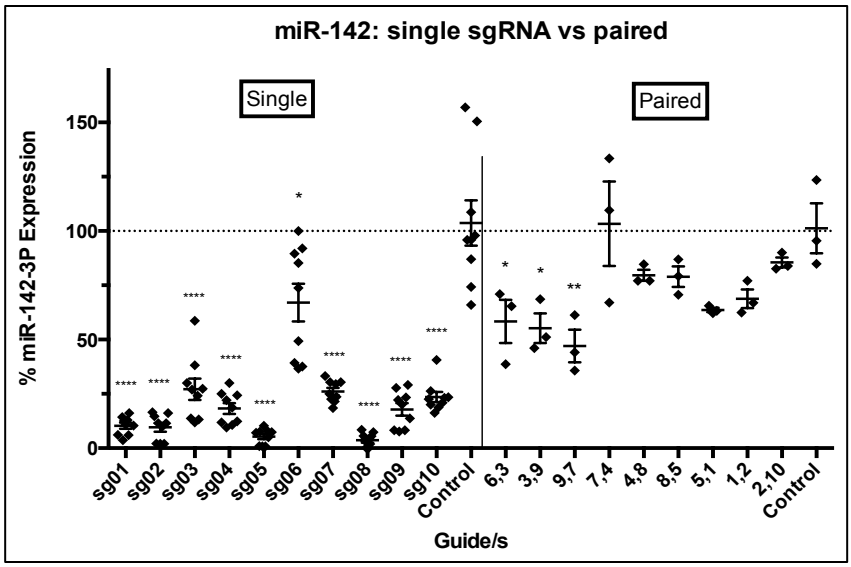


Figure 5: RAW264.7 miR-223-3p expression for one sgRNA. Expression of miR-223-3p in RAW cells as measured by RT-qPCR 48 hours post-transduction. Error bars display mean and SEM, expression normalized to define relative expression as a percentage compared to the no-sgRNA control. ⋆ = $p \leq 0.05$; ⋆⋆⋆ = $p \leq 0.001$; ⋆⋆⋆⋆ = $p \leq 0.0001$, by one-way ANOVA and multiple comparison vs. control. Normalization was conducted using miR-103a and the Exiqon UniSp6 spike-in control.

knockout is desired (as opposed to the generation of mistargeted or targetless miRNAs). However, sg07, which targets the exact same phosphodiester bonds but from the opposite orientation, reduces expression to 26.06% of the control, on par with the effects generated by most of the other sgRNAs. The seven sgRNAs against miR-223 gave results comparable to the sgRNAs against miR-142; five lowered expression to less than 50% of the control (Figure 7). Interestingly, the two poor performers (sg03, sg04) are located in the same rough location as the best performer (sg05), clustered on the edge of the miR-223-3p element and trailing to the Drosha cut site. All data was obtained form whole populations, without selection for transduced cells.

For miR-142, in no case did the addition of a second or third sgRNA improve the loss of function effect (Figure 6). In fact, multiple sgRNAs not only inhibited each other, but in many cases, they acted to completely rescue miRNA expression, and/or increase it over that of the control population. This generally was not the case for the battery arrayed against miR-223, which did not exhibit such counter-intuitive mechanics, with the exception of sg03 and sg04, which increased miRNA expression over the baseline, and one case of functional rescue when combining sg03 and sg07 (Figure 7). However, the addition of more than one sgRNA per population did not significantly enhance the loss of function effect, and I concluded from these studies on miR-142 and miR-223, that it is not beneficial to stack multiple Cas9 complexes without sufficient space. A caveat to this conclusion is that all sgRNA combinations were administered at the same overall MOI (10 MOI, from viral preparations concentrated to approximately 1E6 TU/mL, tittered via puromycin resistance [49])

A



B

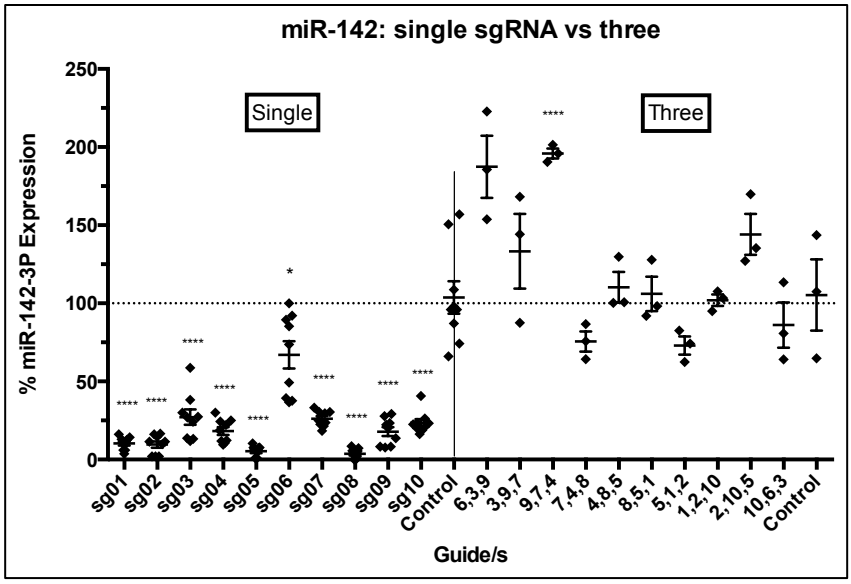
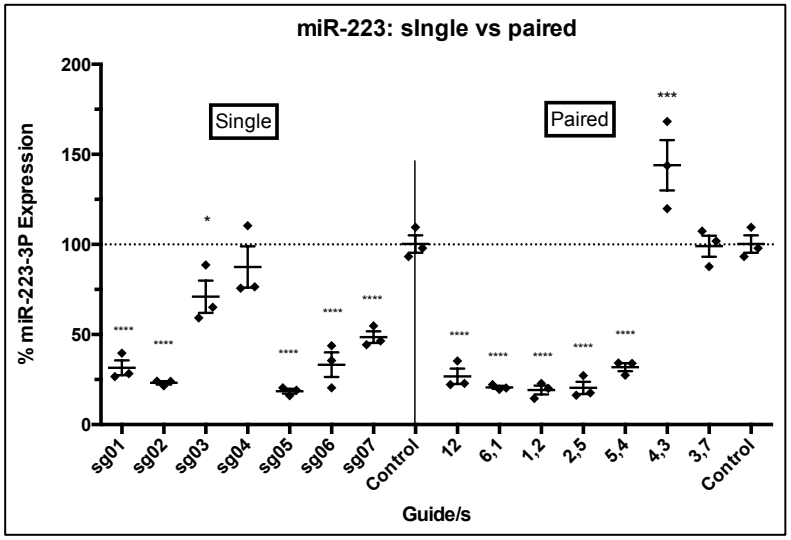


Figure 6: Effects of adding multiple anti-miR-142 sgRNAs to EML culture. Expression of miR-142-3p in EML cells 48 hours post-transduction. Single sgRNA-transduced populations are compared to paired (A), or three (B) sgRNAs. Error bars display mean and SEM, expression normalized to define relative expression as a percentage compared to the no-sgRNA control. ⋆ = $p \leq 0.05$; ⋆⋆ = $p \leq 0.01$; ⋆⋆⋆⋆ = $p \leq 0.0001$, by one-way ANOVA and multiple comparison vs. control. Normalization was conducted using miR-103a and the Exiqon UniSp6 spike-in control.

A



B

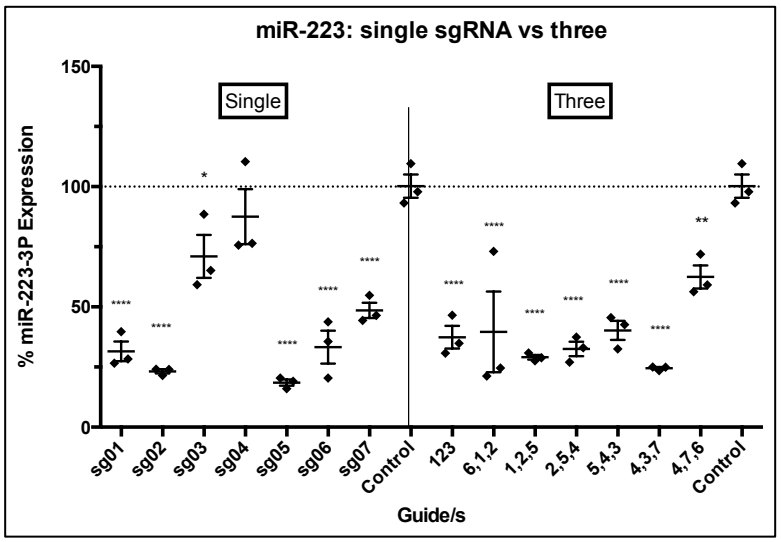


Figure 7: Effects of adding multiple anti-miR-223 sgRNAs to RAW264.7 culture. Expression of miR-223-3p in RAW cells 48 hours post-transduction. Single sgRNA-transduced populations are compared to paired (A), or three (B) sgRNAs. Error bars display mean and SEM, expression normalized to define relative expression as a percentage compared to the no-sgRNA control. ⋆ = $p \leq 0.05$; ⋆⋆ = $p \leq 0.01$; ⋆⋆⋆ = $p \leq 0.001$; ⋆⋆⋆⋆ = $p \leq 0.0001$, by one-way ANOVA and multiple comparison vs. control. Normalization was conducted using miR-103a and the Exiqon UniSp6 spike-in control.

4. CONCLUSIONS

In this study, I attempted to determine the ability of the Cas9 system in general to affect miRNA function and biogenesis, specifically using miR-142 and miR-223, and to also investigate the possibility of overcoming the inherent inefficiencies of Cas9 by employing several independently targeted complexes against the same general sequence. This study included assays on the ability of spCas9 to down-regulate a luciferase reporter, to induce gDNA lesions in HEK293T cells, and to knock out miRNA function in EML and RAW264.7 cell lines. The primary discovery was the broad utility of Cas9 in miRNA knockout, the generation of mutation along any portion of the pre-miRNA or Drosha binding area being sufficient to effectively and specifically disrupt miRNA function. miRNA knockout can be accomplished by the introduction of a single-sgRNA Cas9 complex, without needing to target specific points on the pre-miRNA or to employ two complexes in the hope of deleting the sequence in between. For the seventeen sgRNAs used in this study, three failed to reduce miRNA expression <50%, and the sgRNAs gave variable results; however as long as several sgRNAs are screened, the process of developing new knockout models should be successful. With this comes the caveat that these sgRNAs must be screened in the final system, since neither the luciferase reporter assay nor the estimation of gDNA lesion efficiency proved to be an accurate predictor of miRNA knockdown. It was discovered that the addition of pre-miR-142 is capable of significantly increasing the expression levels of the hRluc component of the pSICHECK system. The addition results in an approximate tripling of

the length of the 3' UTR, a large portion being double stranded, which could possibly increase translational efficiency, simply generate a more stable mRNA isoform.

Interestingly, the addition of more than one Cas9 complex to a single system proved to be at best neutral, and more often detrimental to the desired effect. This might be a result of steric hindrance, however this interfering effect was observed even with complexes separated by 69 bases and binding opposite strands. In some cases, this interference resulted in miRNA expression being not only completely rescued, but somewhat increased, which is a counterintuitive finding. Given that all sgRNA combinations were administered at the same overall MOI as the single sgRNA systems (thus diluting the effective expression levels of each individual sgRNA with increasing numbers), one possible explanation is that no individual complex achieved a sufficient concentration to effectively edit its targeted site. One issue left unaddressed by this study was to determine which stage of processing specifically is inhibited by each targeting site, and whether or not accumulation of the substrate prior to the point of inhibition could be detected. Additionally, miRNAs are capable of generating both 3p and 5p products, dependent on which strand is selected as the guide and the passenger. For both of the miRNAs selected for this study, the 3p form is the dominant and only detectible form in the chosen cell type, and was thus the measure of miRNA expression. However, it would be interesting to determine whether the location of Cas9-indiced mutations can affect strand selection, in the event that the miRNA is processed up to the RISC-loading phase. Additionally, all miR expression data was obtained form whole populations, without selection for the transduced subpopulation. Given the high MOI used throughout

the course of the experiment (validated by titration), it can be assumed that the majority of the population was affected for each case. However, the effects observed may have been somewhat muted by the presence of a non-trivial number of non-transduced cells, and the generation of stable sgRNA-spCas9 lines for each sgRNA may have provided a more precise measurement of the true actions of the spCas9 complex.

In conclusion, I have demonstrated that the spCas9 complex is fully sufficient for effective miRNA knockout, using only a single sgRNA guide oligo, directed to any point along the pre-miRNA in question, with only tentative evidence to suggest certain locations as having any superior effect. I have further shown that there is no shortcut or viable proxy for measurement of this knockout effect, and that there is no evidence to suggest that attempts to overcome low efficiency levels by multiplexed saturation will be successful.

REFERENCES

1. Ha, M. and V.N. Kim, *Regulation of microRNA biogenesis.* Nat Rev Mol Cell Biol, 2014. **15**(8): p. 509-524.

2. Starega-Roslan, J., E. Koscianska, P. Kozlowski and W. Krzyzosiak, *The role of the precursor structure in the biogenesis of microRNA.* Cellular and Molecular Life Sciences, 2011. **68**(17): p. 2859-2871.

3. Tsutsumi, A., T. Kawamata, N. Izumi, H. Seitz and Y. Tomari, *Recognition of the pre-miRNA structure by Drosophila Dicer-1.* Nat Struct Mol Biol, 2011. **18**(10): p. 1153-1158.

4. Lee, Y., M. Kim, J. Han, K.H. Yeom, S. Lee, et al., *MicroRNA genes are transcribed by RNA polymerase II.* The EMBO Journal, 2004. **23**(20): p. 4051-4060.

5. Cullen, B.R., *Transcription and Processing of Human microRNA Precursors.* Molecular Cell, 2004. **16**(6): p. 861-865.

6. Davis, B. and A. Hata, *Regulation of MicroRNA Biogenesis: A miRiad of mechanisms.* Cell Communication and Signaling, 2009. **7**(1): p. 18.

7. Lewis, B.P., C.B. Burge and D.P. Bartel, *Conserved Seed Pairing, Often Flanked by Adenosines, Indicates that Thousands of Human Genes are MicroRNA Targets.* Cell, 2005. **120**(1): p. 15-20.

8. Griffiths-Jones, S., R.J. Grocock, S. van Dongen, A. Bateman and A.J. Enright, *miRBase: microRNA sequences, targets and gene nomenclature.* Nucleic Acids Research, 2006. **34**(suppl 1): p. D140-D144.

9. Gurtan, A.M., V. Lu, A. Bhutkar and P.A. Sharp, *In vivo structure–function analysis of human Dicer reveals directional processing of precursor miRNAs.* RNA, 2012. **18**(6): p. 1116-1122.

10. Hogg, D.R. and Lorna W. Harries, *Human genetic variation and its effect on miRNA biogenesis, activity and function.* Biochemical Society Transactions, 2014. **42**(4): p. 1184-1189.

11. Long, D., R. Lee, P. Williams, C.Y. Chan, V. Ambros, et al., *Potent effect of target structure on microRNA function.* Nat Struct Mol Biol, 2007. **14**(4): p. 287-294.

12. Kumar, M.S., J. Lu, K.L. Mercer, T.R. Golub and T. Jacks, *Impaired microRNA processing enhances cellular transformation and tumorigenesis.* Nat Genet, 2007. **39**(5): p. 673-677.

13. Schwab, R., C. Speth, S. Laubinger and O. Voinnet, *Enhanced microRNA accumulation through stemloop−adjacent introns.* EMBO reports, 2013. **14**(7): p. 615-621.

14. Jinek, M., K. Chylinski, I. Fonfara, M. Hauer, J.A. Doudna, et al., *A Programmable Dual-RNA–Guided DNA Endonuclease in Adaptive Bacterial Immunity.* Science, 2012. **337**(6096): p. 816-821.

15. Ran, F.A., P.D. Hsu, J. Wright, V. Agarwala, D.A. Scott, et al., *Genome engineering using the CRISPR-Cas9 system.* Nat. Protocols, 2013. **8**(11): p. 2281-2308.

16. Wang, H., H. Yang, Chikdu S. Shivalila, Meelad M. Dawlaty, Albert W. Cheng, et al., *One-Step Generation of Mice Carrying Mutations in Multiple Genes by CRISPR/Cas-Mediated Genome Engineering.* Cell, 2013. **153**(4): p. 910-918.

17. Hsu, Patrick D., Eric S. Lander and F. Zhang, *Development and Applications of CRISPR-Cas9 for Genome Engineering.* Cell, 2014. **157**(6): p. 1262-1278.

18. Anders, C., O. Niewoehner, A. Duerst and M. Jinek, *Structural basis of PAM-dependent target DNA recognition by the Cas9 endonuclease.* Nature, 2014. **513**(7519): p. 569-573.

19. Xu, K., C. Ren, Z. Liu, T. Zhang, T. Zhang, et al., *Efficient genome engineering in eukaryotes using Cas9 from Streptococcus thermophilus.* Cellular and Molecular Life Sciences, 2015. **72**(2): p. 383-399.

20. Hsu, P.D., D.A. Scott, J.A. Weinstein, F.A. Ran, S. Konermann, et al., *DNA targeting specificity of RNA-guided Cas9 nucleases.* Nat Biotech, 2013. **31**(9): p. 827-832.

21. Pattanayak, V., S. Lin, J.P. Guilinger, E. Ma, J.A. Doudna, et al., *High-throughput profiling of off-target DNA cleavage reveals RNA-programmed Cas9 nuclease specificity.* Nat Biotech, 2013. **31**(9): p. 839-843.

22. Mali, P., J. Aach, P.B. Stranges, K.M. Esvelt, M. Moosburner, et al., *CAS9 transcriptional activators for target specificity screening and paired nickases for cooperative genome engineering.* Nat Biotech, 2013. **31**(9): p. 833-838.

23. Sternberg, S.H., S. Redding, M. Jinek, E.C. Greene and J.A. Doudna, *DNA interrogation by the CRISPR RNA-guided endonuclease Cas9.* Nature, 2014. **507**(7490): p. 62-67.

24. Larson, M.H., L.A. Gilbert, X. Wang, W.A. Lim, J.S. Weissman, et al., *CRISPR interference (CRISPRi) for sequence-specific control of gene expression.* Nat. Protocols, 2013. **8**(11): p. 2180-2196.

25. Sanjana, N.E., O. Shalem and F. Zhang, *Improved vectors and genome-wide libraries for CRISPR screening.* Nat Meth, 2014. **11**(8): p. 783-784.

26. Zhu, S., X. Rong Z Fau - Lu, Y. Lu X Fau - Xu, X. Xu Y Fau - Fu and X. Fu, *Gene Targeting Through Homologous Recombination in Monkey Embryonic Stem Cells Using CRISPR/Cas9 System.* (1557-8534 (Electronic)).

27. Lieber, M.R., Y. Ma, U. Pannicke and K. Schwarz, *Mechanism and regulation of human non-homologous DNA end-joining.* Nat Rev Mol Cell Biol, 2003. **4**(9): p. 712-720.

28. Takata, M., M.S. Sasaki, E. Sonoda, C. Morrison, M. Hashimoto, et al., *Homologous recombination and non−homologous end−joining pathways of DNA double−strand break repair have overlapping roles in the maintenance of chromosomal integrity in vertebrate cells.* The EMBO Journal, 1998. **17**(18): p. 5497-5508.

29. Weterings, E. and D.J. Chen, *The endless tale of non-homologous end-joining.* Cell Res, 2008. **18**(1): p. 114-124.

30. Mladenov, E. and G. Iliakis, *Induction and repair of DNA double strand breaks: The increasing spectrum of non-homologous end joining pathways.* Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis, 2011. **711**(1–2): p. 61-72.

31. Stark, J.M., A.J. Pierce, J. Oh, A. Pastink and M. Jasin, *Genetic Steps of Mammalian Homologous Repair with Distinct Mutagenic Consequences.* Molecular and Cellular Biology, 2004. **24**(21): p. 9305-9316.

32. Burma, S., B.P.C. Chen and D.J. Chen, *Role of non-homologous end joining (NHEJ) in maintaining genomic integrity.* DNA Repair, 2006. **5**(9–10): p. 1042-1048.

33. Zhao, Y., Z. Dai, Y. Liang, M. Yin, K. Ma, et al., *Sequence-specific inhibition of microRNA via CRISPR/CRISPRi system.* Scientific Reports, 2014. **4**: p. 3943.

34. Jiang, Q., X. Meng, L. Meng, N. Chang, J. Xiong, et al., *Small indels induced by CRISPR/Cas9 in the 5' region of microRNA lead to its depletion and Drosha processing retardance.* RNA Biology, 2014. **11**(10): p. 1243-1249.

35. Kuscu, C., S. Arslan, R. Singh, J. Thorpe and M. Adli, *Genome-wide analysis reveals characteristics of off-target sites bound by the Cas9 endonuclease.* Nat Biotech, 2014. **32**(7): p. 677-683.

36. Sakuma, T., A. Nishikawa, S. Kume, K. Chayama and T. Yamamoto, *Multiplex genome engineering in human cells using all-in-one CRISPR/Cas9 vector system.* Scientific Reports, 2014. **4**: p. 5400.

37. Yaung, S.J., K.M. Esvelt and G.M. Church, *CRISPR/Cas9-Mediated Phage Resistance Is Not Impeded by the DNA Modifications of Phage T4.* PLoS ONE, 2014. **9**(6): p. e98811.

38. Hendel, A., R.O. Bak, J.T. Clark, A.B. Kennedy, D.E. Ryan, et al., *Chemically modified guide RNAs enhance CRISPR-Cas genome editing in human primary cells.* Nat Biotech, 2015. **33**(9): p. 985-989.

39. Wang, D., H. Mou, S. Li, Y. Li, S. Hough, et al., *Adenovirus-Mediated Somatic Genome Editing of Pten by CRISPR/Cas9 in Mouse Liver in Spite of Cas9-Specific Immune Responses.* (1557-7422 (Electronic)).

40. Kim, S., D. Kim, S.W. Cho, J. Kim and J.-S. Kim, *Highly efficient RNA-guided genome editing in human cells via delivery of purified Cas9 ribonucleoproteins.* Genome Research, 2014. **24**(6): p. 1012-1019.

41. Maggio, I., M. Holkers, J. Liu, J.M. Janssen, X. Chen, et al., *Adenoviral vector delivery of RNA-guided CRISPR/Cas9 nuclease complexes induces targeted mutagenesis in a diverse array of human cells.* Scientific Reports, 2014. **4**: p. 5105.

42. Szilvassy, S.J., T.E. Meyerrose, P.L. Ragland and B. Grimes, *Differential homing and engraftment properties of hematopoietic progenitor cells from murine bone marrow, mobilized peripheral blood, and fetal liver.* Blood, 2001. **98**(7): p. 2108-2115.

43. Duran-Struuck, R. and R.C. Dysko, *Principles of Bone Marrow Transplantation (BMT): Providing Optimal Veterinary and Husbandry Care to Irradiated Mice in BMT Studies.* Journal of the American Association for Laboratory Animal Science : JAALAS, 2009. **48**(1): p. 11-22.

44. Xie, K., B. Minkenberg and Y. Yang, *Boosting CRISPR/Cas9 multiplex editing capability with the endogenous tRNA-processing system.* Proceedings of the National Academy of Sciences of the United States of America, 2015. **112**(11): p. 3570-3575.

45. Guilinger, J.P., D.B. Thompson and D.R. Liu, *Fusion of catalytically inactive Cas9 to FokI nuclease improves the specificity of genome modification.* Nat Biotech, 2014. **32**(6): p. 577-582.

46. Wang, F., X.-S. Wang, G.-H. Yang, P.-F. Zhai, Z. Xiao, et al., *miR-29a and miR-142-3p downregulation and diagnostic implication in human acute myeloid leukemia.* Molecular Biology Reports, 2012. **39**(3): p. 2713-2722.

47. Zhuang, G., C. Meng, X. Guo, P.S. Cheruku, L. Shi, et al., *A Novel Regulator of Macrophage Activation: miR-223 in Obesity Associated Adipose Tissue Inflammation.* Circulation, 2012.

48. Kutner, R.H., X.-Y. Zhang and J. Reiser, *Production, concentration and titration of pseudotyped HIV-1-based lentiviral vectors.* Nat. Protocols, 2009. **4**(4): p. 495-505.

49. Ichim, C.V. and R.A. Wells, *Generation of high-titer viral preparations by concentration using successive rounds of ultracentrifugation.* Journal of Translational Medicine, 2011. **9**: p. 137-137.