Spring 5-9-2020

Exploring the Role of ccdc141 in Zebrafish Heart Development

Luci Strong
University of Mississippi

Exploring the role of *ccdc141* in zebrafish heart development

By

Luci Strong

A thesis submitted to the faculty of the University of Mississippi in partial fulfillment of the requirements of the Sally McDonnell Barksdale Honors College.

Oxford
May 2020

Approved by:

_____
Advisor: Dr. Joshua Bloomekatz

_____
Reader: Dr. Sara Liljegren

_____
Reader: Dr. Gregg Roman

For Granny. For LCTIII.

ACKNOWLEDGEMENTS

I would like to offer a special thanks to Dr. Joshua Bloomekatz for his continual guidance and encouragement throughout my time in the lab. His advice and constructive feedback have helped me shape and execute this project. Without this valuable support, this project could not have been completed.

I would like to thank the American Heart Association for providing funding for this project. The efforts put toward research concerning heart disease and stroke are important and essential to understanding these deadly diseases.

I would like to thank the Bloomekatz lab (Dr. Tess McCann, Rabina Shrestha, Joseph Natalizio, Jaret Lieberth, Savanna Tillman, Devin Chen, Alea Jones, Brianna McCollum, and Matthew Knerr) and the Sally McDonnell Barksdale Honors College for this opportunity. I have learned invaluable information and worked with exceptional people through both entities. I am forever grateful for their assistance with this milestone.

I am forever indebted to the fish for making all of this possible.

I would like to express my deepest gratitude to my mother and father for their unending support of my collegiate career and career endeavors. They have offered me lifelong wisdom and unceasing love that have helped me through this journey. For that, I am forever grateful.

ABSTRACT

Luci Caitlin Marie Strong: Role of *ccdc141* in cardiac fusion
(Under the direction of Dr. Joshua Bloomekatz)

Zebrafish are a good model organism in which to identify genes important for heart development because most human genes have a zebrafish ortholog and genes in zebrafish are relatively simple to modify. *ccdc141* is a gene with unknown function, which we originally found to be highly expressed in cardiac cells in mouse embryos. In this project, we are seeking to determine the function of *ccdc141* using zebrafish. To perform this experiment, we are targeting *ccdc141* for mutation using CRISPR/Cas9. In this project we have analyzed *ccdc141* to determine regions of conservation, designed CRISPR guides to target those regions (exon 2 or 3), injected ribonucleoprotein (RNP) complexes containing the generated gRNA and Cas9 into zebrafish embryos, and designed and executed assays to assess the efficacy of those injections. Once a DNA cut is made using gRNA/Cas9, the cell repairs those cuts during which a mutation may be introduced. The mutation introduced may disrupt regular gene function, allowing for a chance to view how the organism functions without a functional *ccdc141* gene. The expectation is that *ccdc141* is necessary for heart function, including the early stages of heart tube formation, namely cardiac fusion and cardiac cell movement. Although we were unable to determine whether a phenotype was created, we did determine that we successfully created a double-strand break in *ccdc141*.

TABLE OF CONTENTS

LIST OF FIGURES

Definitions and Abbreviations List

1. ccdc141 (coiled-coil domain 141): gene of interest; gene containing protein domains called coiled-coil and lies on chromosome 9 in zebrafish
2. DNA (deoxyribonucleic acid): self-replicating genetic material in all organisms contained within the cell
3. mRNA (ribonucleic acid): a messenger molecule carrying instruction for DNA replication and protein synthesis
4. PVUII: restriction enzyme
5. NCOI: restriction enzyme
6. PCR (polymerase chain reaction): a laboratory technique that makes many copies of a segment of DNA
7. sgRNA (single guide RNA): single-stranded RNA molecule that contains a crRNA with targeting sequence and tracrRNA sequence
8. cDNA (complementary DNA): DNA synthesized from single-stranded RNA molecules using an enzyme called reverse transcriptase
9. Crispr/Cas9 ribonucleoprotein complex: RNA-protein complex that is targeted to a specific DNA sequence which is cut by the Cas9 nuclease

Introduction

Heart development consists of many complex and multifaceted events. These important events include cardiac cell differentiation, the formation of an early heart tube, and the development of heart chambers (Miquerol, 2013). These events are controlled by different genes turning on and off, known as gene expression, at various points during development. Each event and set of expressed genes eventually lead to a fully functioning heart. Although there are differences between the hearts of different vertebrate species- for example human hearts have four chambers while zebrafish have two- the early stages of heart development in zebrafish, including bilateral heart fusion, migration of cardiac cells, and cardiac looping, are fundamentally similar between zebrafish and humans. Thus, zebrafish heart development can be studied to determine the mechanisms and processes that drive all vertebrate heart development including humans (Bakkers, 2011). This ability to study and analyze zebrafish at critical points in heart development make them an ideal model organism.

The first step in zebrafish heart formation is cardiac fusion. The majority of mutations that have been found to cause cardiac fusion phenotypes are located in genes that function in non-cardiac cells or the extracellular matrix. However, genes with specific roles in cardiac cells such as *pdgfra* are also involved in cardiac fusion showing that these cells likely have a role in their own movement (Bloomekatz, unpublished data). This further suggests that those genes in cardiac cells should be identified and studied. One possible gene is *ccdc141*. This gene was shown to be highly expressed in cardiac cells and thus warranted more investigation. This experiment sets out to

discover more about the role of *ccdc141* in zebrafish heart development and the impact on the organism without the gene.

I. Zebrafish as a model organism

A model organism is defined as a controllable system that is used to study a more significant theme in biology. Model organisms have been used in single-cell regeneration research, epigenetics, and cytoplasm organization studies (Russell, 2017). In the case of zebrafish, they have been used to study cellular processes, gene expression, and even human diseases. Conclusions from zebrafish studies have helped to increase knowledge of organ function and cellular mechanisms as well as provide more efficient and healthier conditions for the system being modelled.

One topic of interest is cardiac development. Understanding the molecular, cellular, and genetic components for this monumental event is important to understanding the diseases that affect the heart. Zebrafish are a good choice for learning more about these components because the embryos do not solely rely on a functional cardiovascular system for their development. While embryos can lack blood circulation due to a defect, oxygen is still able to enter the embryo and reach the tissues via passive diffusion, allowing the embryo to survive past the initial phase of embryonic development (Bakkers, 2011). This is especially helpful when wanting to analyze what specifically is happening to the heart as the embryo is developing without immediate mortality. Additionally, zebrafish have low maintenance needs, quickly mature, and have a similar genetic makeup to humans.

The key components for zebrafish care are tank system maintenance, feeding, breeding, and raising of larvae. Zebrafish should be kept in a circulating system that constantly provides filtered water. This filtration helps to remove excess food or fish waste. The tanks are cleaned regularly to remove any debris or algae inside the tank. Zebrafish can be fed dry food or live food. The amount given should remain relatively constant, unless modified due to that specific line of fish.

Zebrafish begin breeding upon onset of light. For successful breeding, the male and female pair should be of approximate equal size. They are placed in a breeding tank in the afternoon or evening and are allowed to mate when the light turns on again the next morning. In order to maintain peak embryo health, embryos are collected soon after breeding, and the parental pair are returned to their original tanks or a screening tank. With these protocols established, the care of zebrafish is easier and cheaper than other model organisms such as rodents (Avdesh et al., 2012). A one-day breeding process allows for many zebrafish to be hatched in a short amount of time. Thousands of fish can all be kept in the same room, conserving time and money. Zebrafish also mature quickly from embryo to adult, aiding in their cost effectiveness.

The quick maturation of zebrafish makes them a prime target for research. Their growth happens in developmental stages, each with their defining features like body length. Larval zebrafish measure around 3.4-12 mm. Juvenile zebrafish are defined as measuring anywhere from 12-18 mm. Adult zebrafish are still fairly small at 18 mm (Singleman & Holtzman, 2012). While these lengths from snout to the base of the tail fin have a range, the considerable growth that occurs in about two to three months makes zebrafish an attractive model organism. While rates of growth vary from fish to fish due

to variations in feeding or tank quality, zebrafish are constantly growing (Singleman & Holtzman, 2014). This continuous growth is directly correlated to heart growth size. To compensate for the rapid and lifelong growth, there is a need for a larger cardiovascular output (Singleman & Holtzman, 2012). This need causes a developing change to the morphology of the heart that is observable and affects the maturation of zebrafish past the embryonic developmental stages (Singleman & Holtzman, 2012).

Zebrafish offer the benefit of being similar to human, in terms of genes. About 70% of human genes have a zebrafish ortholog (Santoriello & Zon, 2012). With this fact, different human diseases can be researched using zebrafish as the model organism. Researchers are able to mutate the genes in zebrafish corresponding to the ones linked to human disease and view how the mutation affects the zebrafish. Experiments can include forward genetic screens (Santoriello & Zon, 2012) or transgenics, containing one or more DNA sequences introduced from other species. Mutations causing human diseases such as hemochromatosis, melanomas, Duchenne muscular dystrophy, and human dilated cardiomyopathies all have orthologous genes in zebrafish (Santoriello & Zon, 2012). This burgeoning genetic and developmental research offers hope for the future of these crippling diseases. Additionally, novel genes, which are highly expressed in cardiac tissues with no previously known function, offer information that could shed light on a specific part of cardiac development. Different genetic tests must be performed on zebrafish to understand the role and purpose of these novel genes.

I. *ccdc141*

Using a single-cell RNA sequence (RNAseq) dataset, we have identified a novel gene called *ccdc141*. This dataset contains single-cell transcriptomes from approximately 10,000 mice cells, consisting of the earliest known cardiac progenitors and their cardiac lineage (Bloomekatz, unpublished grant). Approximately 200 genes were found to be differentially expressed in myocardial cells (Bloomekatz, unpublished data). One of the highest expressed genes among those 200 genes is *ccdc141* (**Figure 1)**.

A few studies of *ccdc141* have occurred in mice and humans. In these studies, *ccdc141* is associated with changes in the nervous system, due to impaired radial migration and the migration of GnRH neurons (Hutchins et al., 2016). Along with associations with cell movement, genome-wide association studies revealed an association between 20 genes at 11 loci and heart rate in humans. One of those genes was *ccdc141*. To analyze genes responsible for heart rate regulation, fruit flies and zebrafish were used as model organisms to test the downregulation of these genes to determine the effects on heart rate. However, *ttn*, which is located adjacent to *ccdc141* on the same chromosome and is a well-known cardiac gene, was presumed to be affected since it is already known to cause dilated and hypertrophic cardiomyopathy, which can lead to heart failure. This made it difficult to have conclusive results (Den Hoed et al., 2013). Our experiment builds upon the results of these association studies and investigates *ccdc141* further using a genomic editing tool called CRISPR/Cas9.

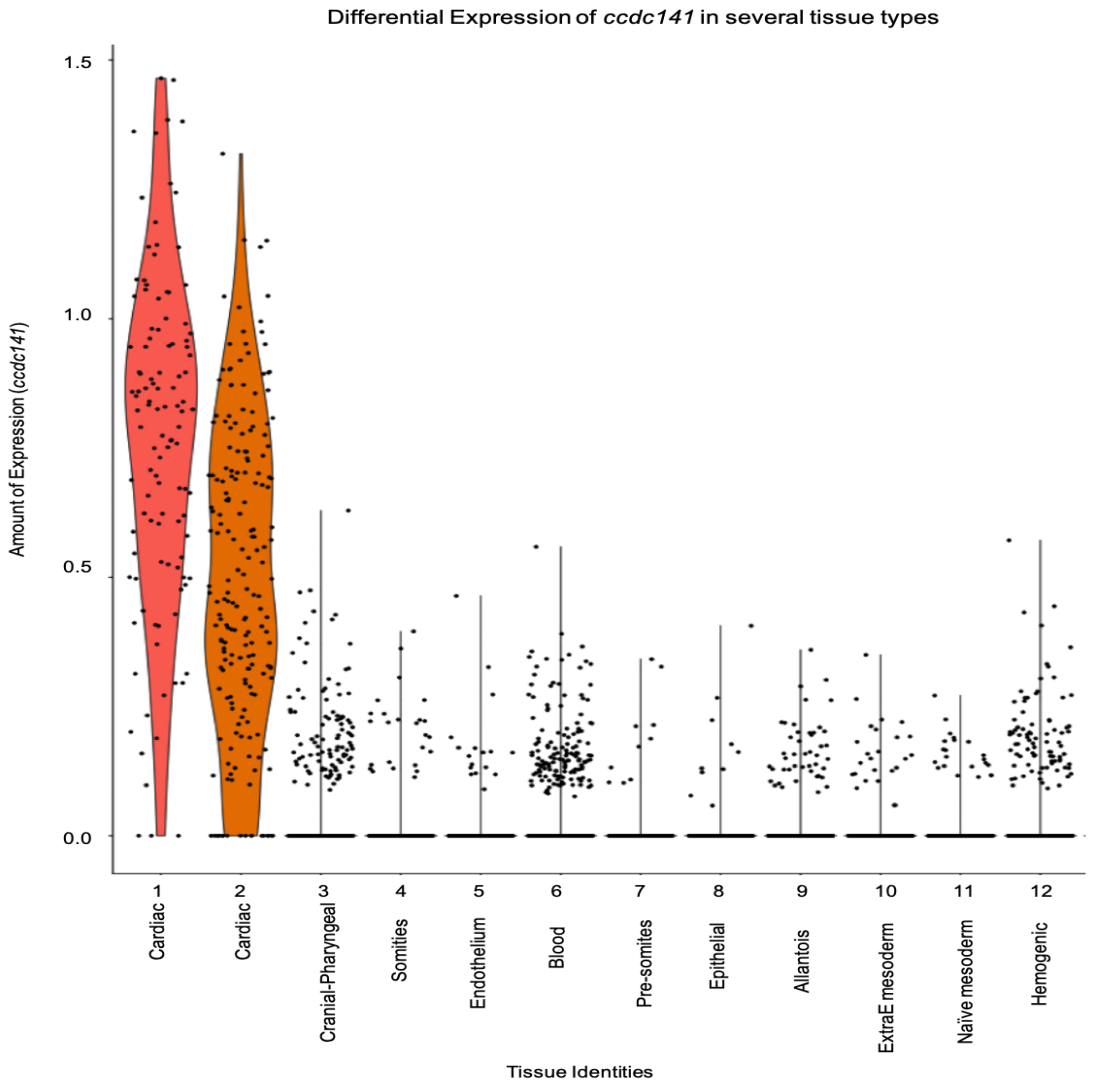

Figure 1. Distribution of *ccdc141* expression in different cell types. Each dot represents the relative amount of *ccdc141* expressed in a single cell. Y-axis represents log-fold change compared to average expression across all cells (Adapted from Bloomekatz, unpublished data).

III. CRISPR/Cas9 Function

This experiment seeks to gain more knowledge about *ccdc141* using a technology called CRISPR/Cas9. CRISPR stands for Clustered Regularly Interspaced Palindromic Repeats, and Cas9 stands for CRISPR Associated Protein 9 (Cas9). The system is derived from an adaptive immunity system in bacteria and archaea that integrates pieces of viral DNA from previous infections and uses the transcripts of those pieces as guide RNAs to cleave subsequent viral infection (Koonin & Makarova, 2013). There are multiple types of these CRISPR-Cas systems, and they are classified into three separate types: I, II, and III. Each type uses a specific protein or protein family to function. Type I functions via proteins from the RAMP superfamily, which largely encompasses Cas5 and Cas6 families. Type II uses Cas9, a large singular protein, that uses CRISPR RNA (crRNA) to cleave the target DNA. Type III mainly utilize RAMP proteins and Cas6 (Makarova et al., 2011). For this experiment, the type II system is used.

As previously stated, the type II CRISPR-Cas system works largely with the Cas9 protein shown in **Figure 2**. The Cas9 protein is made up of at least two nuclease (enzyme that cleaves nucleotide chains) domains: the RuvC-like nuclease domain and the HNH nuclease domain (Makarova et al., 2011). The HNH nuclease domain is contained in many restriction enzymes and has endonuclease activity (Jakubauskas et al., 2007). This could explain the system's ability to cleave target DNA. For this cleavage to occur, the transcript of crRNA must be processed. This is done by endoribonucleases that act in a group or singularly (Makarova et al., 2011). An

alternative called *trans*-encoded small RNA (tracrRNA) can be used to act as a guide for processing the transcript of the crRNA (Deltcheva et al., 2011).

The tracrRNA-processed crRNA and Cas9 protein will both be injected into the zebrafish embryo. The first part is the single guide RNA (sgRNA), or a fusion of crRNA and a tracrRNA with a targeting guide sequence that aids in DNA cleavage by Cas9 (Jinek et al., 2012, Hsu, 2014). This guide sequence matches a sequence in the DNA that is 20 bp long followed by the protospacer adjacent motif (PAM) sequence, NGG (N: any nucleotide) (Kotani et al., 2015). Using a specific method, the sgRNA will be made using oligonucleotides (oligos), free of a plasmid (Burger et al. 2014). The second part is the Cas9 protein, which is specialized for zebrafish by using a specific amount of the protein in solution, that has the ability to cut DNA. The sgRNA acts as a guide for the Cas9 protein, guiding it to the specific place of DNA that matches part of the sgRNA. Once there, the Cas9 protein cuts the DNA. Double-standed breaks are repaired using either homology-directed repair (HDR) pathway or non-homologous end joining (NHEJ) (the joining of the two broken ends of DNA) (Chang et al., Hwang et al., 2013a; Hwang et al., 2013b, Jao et al., 2013).

NHEJ is inherently deleterious, inserting or subtracting nucleotides (indel) during the process. These indels can leave the gene nonfunctional. Without all of the proper nucleotides in the DNA sequence, the protein eventually encoded from the mutated gene would be incomplete or missing multiple amino acids, ridding the protein of its original function altogether. Additionally, indels can shift the reading frame of a protein dramatically, changing the amino acids and often leading to the insertion of an early stop codon. This technique is very common and allows for efficient mutagenesis. The

goal is to create an optimal rate of mutagenesis via germ-line transmission that is not too high (causing double-strand break, chromosomal rearrangements, and eventually apoptosis) or too low (causing inefficient mutagenesis and more embryos for screening) (Bloomekatz, unpublished data). With efficient use of the CRISPR-Cas9 system, *ccdc141* can be knocked out, allowing for the analysis of the gene's possible function.

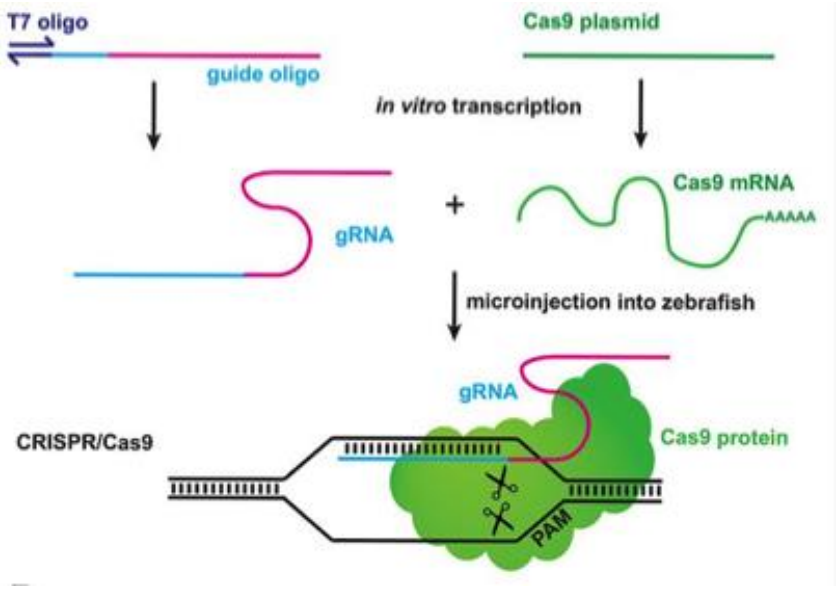

Figure 2. Use of CRISPR/Cas9 to create double strand breaks. Using the CRISPR/Cas9 system to create a double strand break involves two components: a guide RNA and the Cas9 protein, which makes a cut in the double-stranded DNA. Repair of the double strand break often creates a mutation via the insertion or deletion of DNA (Adapted from Hruscha, 2013).

Analysis of this knockout of *ccdc141* could reveal a molecular mechanism behind cardiac fusion and heart development in zebrafish. The hypothesis of this experiment is *ccdc141* is essential for cardiac fusion and plays a role in the movement of the cardiac cells. Using the technique outlined above, we will observe the differences between mutant and wildtype zebrafish. We will be looking for confirmed mutagenesis via gel electrophoresis and possible phenotypic changes. The results from this experiment will help to provide more information about *ccdc141* in zebrafish and the role it plays in zebrafish heart development.

Methods

I. Pre-injection

A target sequence within the gene *ccdc141* was chosen to be mutated using different bioinformatic techniques. The first program used was Clustal, an online software that allows for multiple genetic sequences from different species to be aligned. These sequences are arranged to view regions of similarity between the sequences. This similarity is known as conservation. Various species' amino acid sequences were copied and pasted into the sequence aligner. From there, the software aligned the sequences in order of common ancestry and marked the areas where the species had similar amino acid stretches.

Next, CHOPCHOP, a program used to design oligonucleotides for making sgRNAs and for validating mutagenesis, was used. The complementary DNA (cDNA) sequence for *ccdc141* was inputted into the program, and possible target sequences and left and right primers for each target selection were shown. Once a target sequence and primers were chosen, a restriction enzyme site could be chosen as well. Ideally, it should be near or within the target sequence to allow for accurate mutagenesis. The restriction enzyme site is destroyed if the Cas9 is properly targeted.

Last, the single-guide RNA (sgRNA) was made using a constant and custom oligonucleotide (oligo). The constant oligo was ordered and is the same for all guide RNA sequences. The custom oligo was made using a promoter region, a 20-base pair (bp)

spacer region that is specific to the target site, and an overlap region that matches the constant oligo. At 100 μM each, 1 μL of the constant and custom oligo were added together with purified water. The samples were placed into the polymerase chain reaction (PCR) machine to run on an annealing setting. Once annealed, T4 DNA polymerase was used to fill in the rest of the nucleotide bases of the template. The SP6 kit, a DNA-dependent RNA polymerase that synthesizes RNA sequences from short DNA templates containing the 18 base pair promoter region, was used to transcribe the template (Tom Stump). DNase was added to remove any final traces of DNA. After using a cleanup kit, the sgRNA template was complete. It was then analyzed using PCR and gel electrophoresis

II. Injections

Adult wildtype, AB, fish were used to start the mutagenesis. DNA samples from these adult fish were collected and sequenced, a process where the order of nucleotides of the DNA is determined. These sequences were used for our bioinformatic analyses (see above). Adult fish with the same sequence at the *ccdc141* region were bred, and the resulting embryos were used for this experiment. Injections of the embryos occurred at the one-cell stage at 24 hours post-fertilization (hpf). First, the sgRNA was mixed with the Cas9 protein, and then the Cas9 ribonucleoprotein (RNP) complex. The Cas9 protein itself was tagged with a green fluorescent protein (GFP) to verify injection. After 2-3 months, these embryos (now known as the F0 population) were considered adults and could be used for outcrosses.

III. Outcross

The F0 population was then outcrossed, bred with a wildtype (non-mutant) fish. One F0 male was bred with a female wildtype (AB, TL, or AB x TL) of equal size. One F0 female was bred with a male wildtype (AB, TL, or AB x TL) of equal size. The embryos produced from this mating event were collected. These fish are known as the F1 population. The F0 pairs were then placed into a screening tank with a unique code until analysis of the embryos was completed.

IV. DNA extraction

About 10 embryos from each clutch were selected and placed into a 1.5 μL centrifuge tube. DNA was extracted by adding 50 μL of lysis buffer, incubating for 10 minutes at 98°C, adding 10 μL of proteinase K, and incubating at 55°C overnight. The following day, the extracted DNA was then incubated at 98°C for 10 minutes and cooled on ice for 10 minutes. Once cooled, the DNA was diluted using purified water. Next, the DNA pieces were amplified using PCR. Then, the restriction enzyme site was digested overnight. The next day, the DNA samples for each clutch were analyzed using gel electrophoresis.

Results

This experiment sought to identify the function of the gene *ccdc141* on cardiac fusion in zebrafish. Our preliminary data revealed that *ccdc141* is expressed in the myocardium during cardiac fusion (Bloomekatz, unpublished data). However, this does not fully answer the question of the role of *ccdc141* in heart development. To answer this question, our goal was to introduce a mutation that would disrupt the sequence of *ccdc141*, making *ccdc141* inoperative. Then, we could analyze zebrafish heart development in the absence of *ccdc141* function.

I. Choosing a sequence to target in ccdc141

This first step in the experiment used the software called Clustal to choose a region of *ccdc141* to target. It was important to choose a target sequence that is conserved indicating an essential function, is located near the N-terminus, and is near a restriction enzyme site, a place in or near the target sequence that can be digested by a restriction endonuclease, to verify that a cut was made by the Cas9 protein. Clustal is a software that aligns input sequences based on similarities in the amino acid sequence. Once this is

finished, Clustal marks where the sequences are conserved using punctuation marks **(Figure 3)**. To start, the different species were chosen, and their sequences were pasted into Clustal.

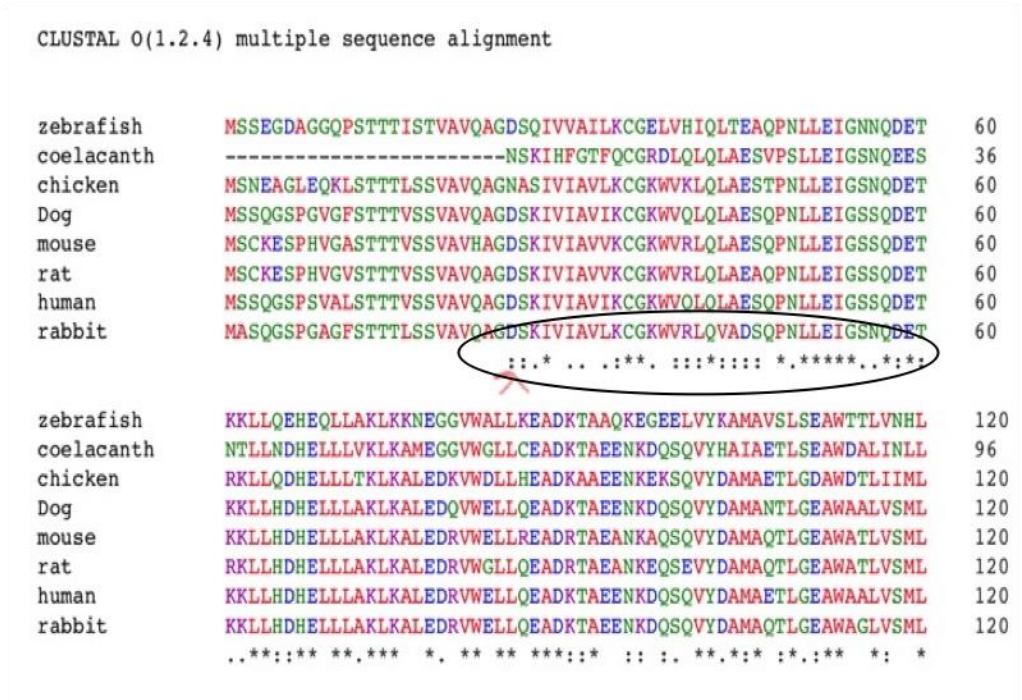

Figure 3. CLUSTAL Results. Amino acid sequences of *ccdc141* orthologs from different animals were aligned using the software CLUSTAL in order to identify conserved regions. Clustal is a number of computer programs in bioinformatics for multiple sequence alignment. An asterisk denotes a single, fully conserved amino acid. A colon signifies conservation between groups of amino acids with strongly similar properties. A period represents conservation between groups of amino acids with weakly similar properties. The different colors of each letter represent the properties of the amino acid. Red means the amino acid is small and hydrophobic, or water repellent. Blue means the amino acid is acidic, having a low pH. Magenta means the amino acid is basic, having a high pH. Green means the amino acid side chain contains special groups such as hydroxyl, sulfhydryl, or amine. Grey means the amino acid is rare and does not fit into the major categories. The numbers to the right of the sequences indicate where in the sequence these amino acids are. These conserved regions are important because it suggests these regions are essential to the protein's function.

When choosing the different species, some were more closely related, sharing a closer common ancestor to zebrafish than others. For example, a coelacanth, a large lobed fish, is more closely related to the zebrafish than a dog. Choosing a wide variety of species was crucial because it shows that *ccdc141* is highly conserved across phylogeny. The high level of conservation suggests that this gene is important for most species of animals. It highlights regions within the protein that have been conserved, despite large evolutionary divergence between the species.

Asterisks, periods and colons at the bottom of the input sequences indicate how conserved a stretch of amino acids is between the different species. An asterisk denotes a single, fully conserved amino acid. A colon signifies conservation between groups of amino acids with strongly similar properties. A period represents conservation between groups of amino acids with weakly similar properties. The different colors of each letter represent the properties of the amino acid. Red means the amino acid is small and hydrophobic, or water repellent. Blue means the amino acid is acidic, having a low pKa. Magenta means the amino acid is basic, having a high pKb. Green means the amino acid side chain contains special groups such as hydroxyl, sulfhydryl, or amine. Grey means the amino acid is rare and does not fit into the major categories. Using the reference amino acid sequences from Ensembl (a genome browser for different species) for alignment, it could be deduced that the most conserved regions fell within exons 2 and 3. These exons were chosen as the target sequences.

These findings identified regions for targeting and excluded other inadequate targets. It also offers the chance for replication of the experiment in other model organisms, since the region is highly conserved in the various species.

II. Choosing the primers and restriction enzyme for verifying targeted mutagenesis

We used the website, CHOPCHOP (http://chopchop.cbu.uib.no/) to choose a target sequence within exon 2 and exon 3 of *ccdc141* and to identify primers and a restriction enzyme for the verification of mutagenesis. CHOPCHOP identifies 20 base pair sequences immediately upstream of a PAM site (NGG) within the specified region and ranks them based on GC content and self-complementary (indicators of annealing), the number of possible off-targets when there is 0, 1, 2, and 3 mismatches and predicted efficiency of cutting based on machine-learning algorithms (Shen et al., 2014). Once a target sequence is chosen, CHOPCHOP identifies primers that can amplify the region surrounding the cut site, as well as restriction enzyme sites near the cut site.

The left and right primer sequences displayed in the first row of **Figure 4** (Pair 1) were chosen for exon two because they offered the largest product size. Choosing the primers that yield the largest product size offers a better chance to identify a mutation after it is created.

Pair	Left primer coordinates	Left primer	Left primer Tm	Left primer off-targets	Right primer coordinates	Right primer	Right primer Tm	Right primer off-targets	Pair off-targets	Product size
1	chr9:43137676-43137698	GATCACCTTAAGTTTGGCAAGG	60.0	0	chr9:43137925-43137947	AGTGGAGAGAACCTGCTTTCAG	60.1	0	0	271
2	chr9:43137677-43137699	ATCACCTTAAGTTTGGCAAGGA	60.0	0	chr9:43137925-43137947	AGTGGAGAGAACCTGCTTTCAG	60.1	0	0	270
3	chr9:43137676-43137698	GATCACCTTAAGTTTGGCAAGG	60.0	0	chr9:43137924-43137946	GTGGAGAGAACCTGCTTTCAGT	59.9	0	0	270
4	chr9:43137677-43137699	ATCACCTTAAGTTTGGCAAGGA	60.0	0	chr9:43137924-43137946	GTGGAGAGAACCTGCTTTCAGT	59.9	0	0	269
5	chr9:43137677-43137699	ATCACCTTAAGTTTGGCAAGGA	60.0	0	chr9:43137852-43137874	GCATGTACAGTAGCACAGCCAT	60.2	0	0	197

- Primers selected

Figure 4. Possible Primers near Exon 2 of *ccdc141* recommended by CHOPCHOP for assessing whether a mutation has been created after injection of gRNA and Cas9. This result offers the possible primers and their locations in relation to the target site from the software CHOPCHOP. This software gives the number of off-targets for primer annealing and the product size. The largest product size was preferred. Therefore, the first set of primers was chosen with the least off-targets for primer annealing and largest product size at 271 bp.

A target sequence is chosen near a restriction enzyme site, so that it can be used to confirm that the intended mutagenesis by sgRNA and Cas9 was successful. The restriction enzyme will digest amplified non-mutagenized DNA. However, mutagenesis by CRISPR/Cas9 will destroy the restriction enzyme site and thus the restriction enzyme will not digest amplified mutagenized DNA. Possible restriction enzyme sites can be seen in **Figure 5**. Unfortunately, there was not a restriction enzyme in the cut site. So, we considered one within 5 bps of the target. Being directly within the target sequence would ensure an accurate cut. Also, the accessibility of the restriction enzyme played a role, as some would not be possible to purchase. Once all factors about the restriction enzyme were considered, a choice could be made based on the results given from CHOPCHOP.

The restriction enzyme site closest to the target sequence of exon two was PVUII (**Figure 5**). To ensure that the restriction enzyme would work correctly for zebrafish, a test was done. Using wildtype adult zebrafish embryos, PCR was conducted using the primer sequences in Figure 4, and then PVUII was added to the PCR product. To determine whether the PVUII cut the amplified DNA, we separated the DNA fragments using gel electrophoresis on the PCR product that had experienced the restriction enzyme digest. The digest was deemed successful by the elimination of one band around 270 bps and the appearance of two distinctive bands around 120 and 170 bps (**Figure 6**). This illustrates that the restriction enzyme digested the amplified DNA, cutting it in half (lane 3, marked embryo 1 PVUII) compared to undigested PCR samples (lane 2, marked embryo 1).

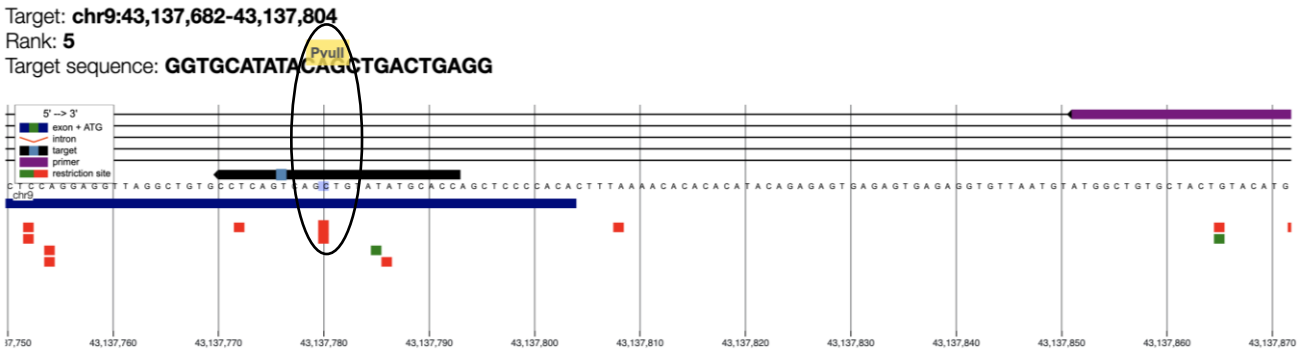

Figure 5. Possible Restriction Enzymes to use to assess CRISPR/Cas9 cutting. Using CHOPCHOP, all restriction enzyme sites near the target sequence are shown. The circled restriction enzyme is PVUII. It is 5 bp away from the target sequence. The restriction enzyme in the same location as PVUII is MspA1I. The next closest restriction enzyme (5 bps before the target sequence) is DdeI (not shown) (From http://chopchop.cbu.uib.no/).

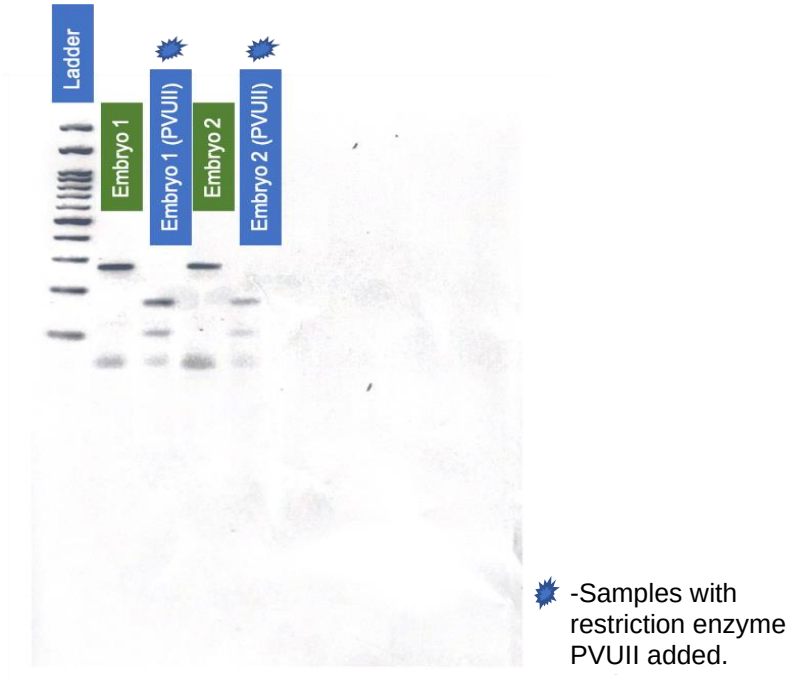

Figure 6. Testing PVUII Digest of PCR Product. The gel electrophoresis image shows the 100-bp ladder and four samples. Using the primers for exon 2, the restriction enzyme site for PVUII was digested in two different embryos. In the samples with PVUII, there are two distinct bands, indicating that the original band was completely digested.

III. Designing the sgRNA for the target sequence with Assay

The sgRNA is one of the two components of the Crispr/Cas9 ribonucleoprotein complex. With the guide RNA, the Cas9 protein will know where to make the cut in the target sequence. To generate the sgRNA, a gene-specific oligo was created, consisting of an RNA polymerase promoter sequence, the 20 bp target sequence immediately upstream of the PAM site, and an overlapping region with the constant oligo, which contains the tracrRNA sequence. These oligos were mixed together with dNTPs and DNA polymerase to create a double-stranded sequence that will contain the hybrid crRNA and tracrRNA sequences. This double-stranded sequence is purified and mixed with RNA polymerase and rNTPs to create an sgRNA via in vitro transcription.

Analysis of the process of creating the gRNA was done using gel electrophoresis (**Figure 7B**). The faint band around 200 bp could be due to left-over DNA fragments, hence why the band is light and not of interest. After PCR, a fragment of the RNA had been clearly defined around 120 bp, the target size for the annealed custom and constant oligos. With confirmation that the oligos were annealed and the gRNA could be created, the experiment could proceed, and mutagenesis could begin.

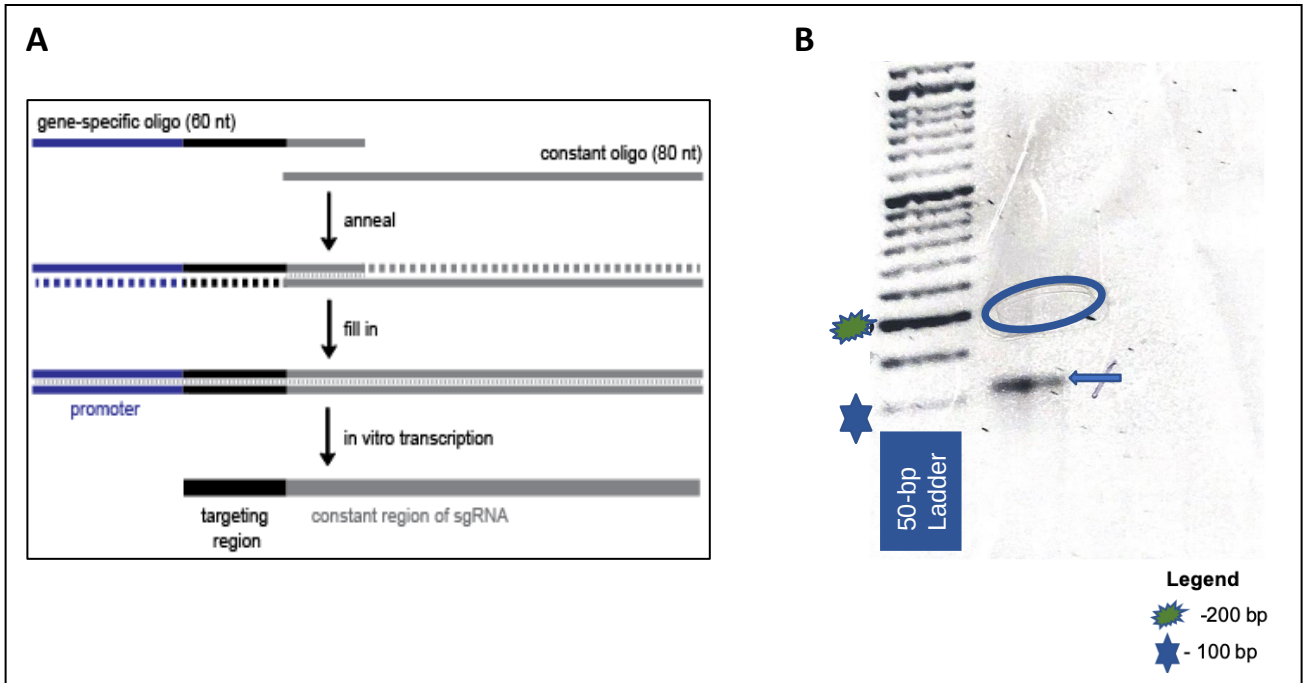

Figure 7. Process for making a gRNA. (A) The above figure describes the process for making the gRNA. A custom oligo, specific for our target, is annealed to a constant oligo containing a tracr sequence. These oligos are annealed together and then used to create a double-stranded piece of DNA using DNA polymerase. This piece of double-stranded DNA contains an RNA polymerase promoter, guide (targeting DNA), and tracr DNA. RNA polymerase is then added to create gRNA which is used to guide the RNP complex to the target site (Adapted from Burger, 2014). **(B)** show a gel electrophoresis image of the successful production of the annealed custom and constant oligos. The ladder to the left is 50-bp, and the darker band around 100 bp (denoted with an arrow) is the extracted double-stranded piece of DNA.

IV. Performing injections for mutagenesis

To confirm whether the gRNA and Cas9 RNP complex would work, they were injected into the wildtype fish embryos. Pooled embryos from each injection were used for the analysis. The image from the gel electrophoresis (**Figure 8**) shows a distinctive difference between the F0 population and the wildtype population. It is evident by the size differences of the bands that the injected gRNA and Cas9 RNP destroyed the restriction enzyme site. Thus, the restriction enzyme did not work in embryos injected with the gRNA/Cas9 RNP complex. The shift in the bands in **Figure 8** indicates that the original band around 270 bps has been cut in two around 120 bps and 170 bps by the restriction enzyme in un-injected wildtype embryos (labeled as – for negative control) while these bands were not observed in samples in the gRNA/Cas9 RNP complex had been injected and which the restriction enzyme was added. From this data, it cannot be said definitively what was the nature of the mutation. To delve deeper into this piece of the experiment, outcrosses and screenings would be performed to gather more information.

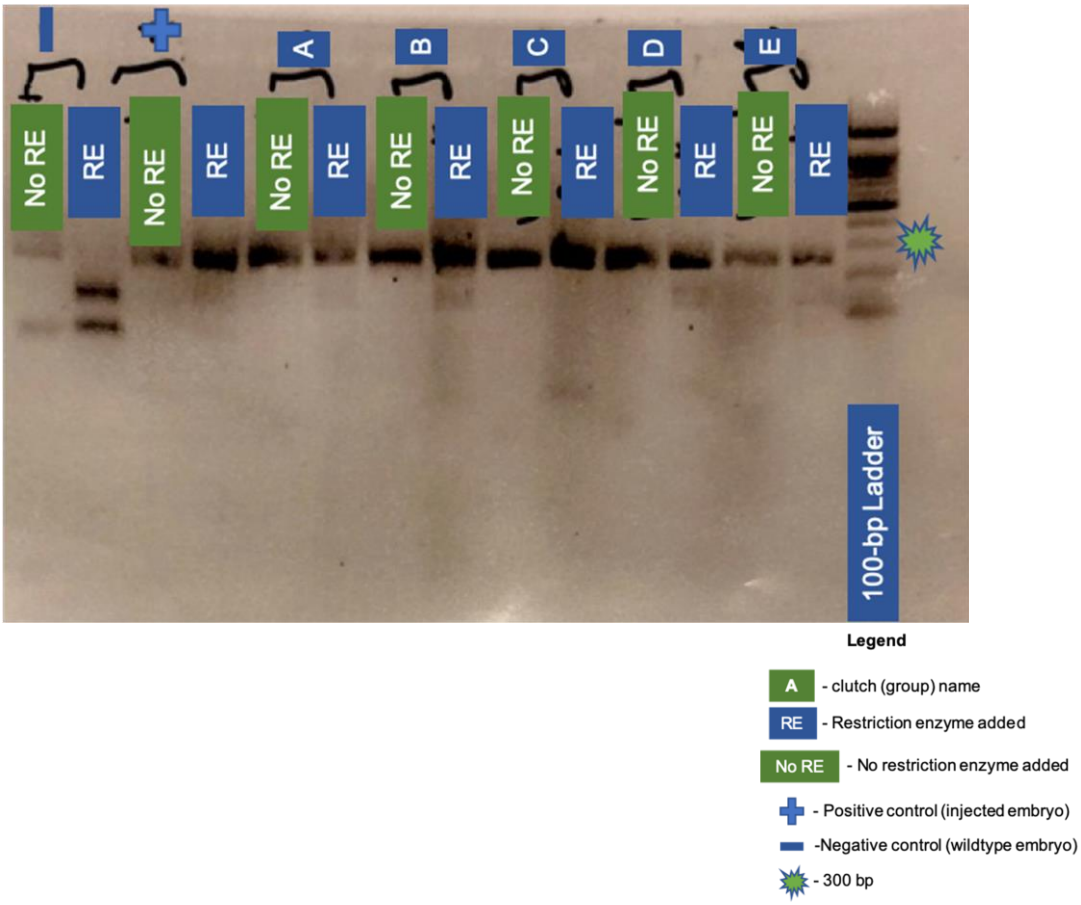

Figure 8. Assessment of gRNA cutting reveals successful mutation. The gel electrophoresis image shows the successful destruction of the restriction enzyme site by gRNA/Cas9 injection. For the controls, there was an injected (+) and uninjected (-) sample. All samples are embryos of a founder and wildtype outcross. The lanes labeled RE had the restriction enzyme added. The lanes labeled No RE only contained the pooled injected embryo DNA. The split from one band to two for the injected samples prove that the restriction enzyme site was destroyed, resulting in two bands for each cut section of the site.

V. Outcross and Screening of mutant embryos

The founder population of fish were then mated with wildtype fish in an outcross. The pairs were then placed into a screening tank as shown in **Figure 9**. With successful mutants confirmed, it was important to determine which founders possessed the strongest germline contribution of the mutation and which could definitively pass the mutation to their offspring. F0 animals were bred with wildtype animals of the opposite gender (AB, TL, or AB x TL), and the resulting embryos were analyzed individually, noting which founder parented that particular embryo. We are currently in the process of analyzing these embryos. For F0 animals with a large germline contribution of the mutation, we expect a large undigested band (top) in the samples in which the restriction enzyme is added. For F0 animals with little to no germline contribution of the mutation we expect to see an almost completely digested sample.

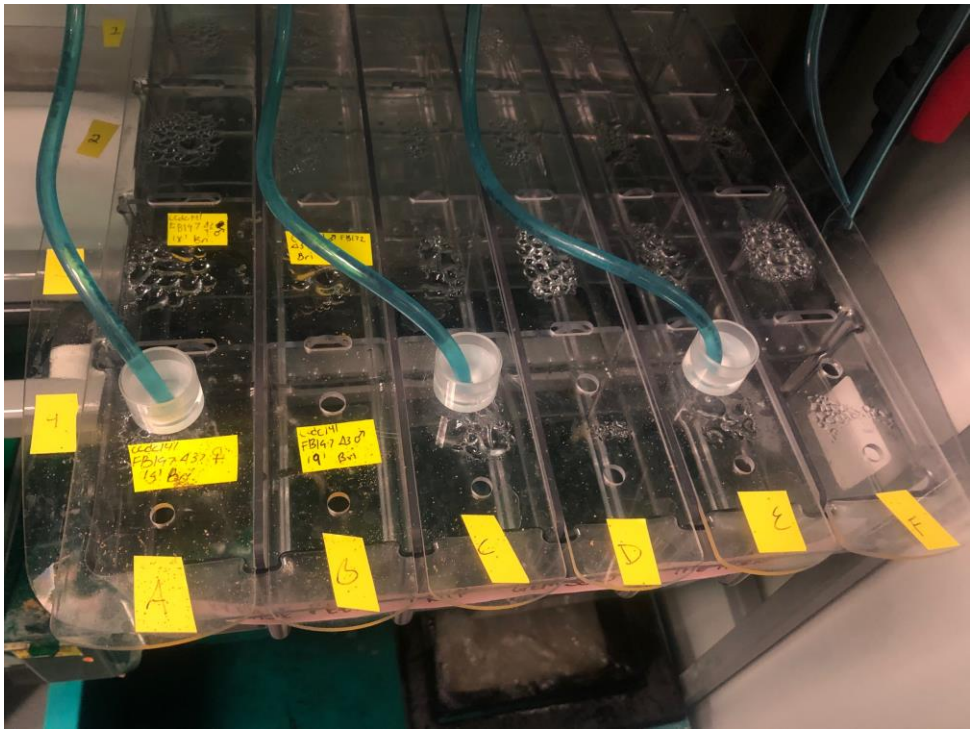

Figure 9. Screening tank holding possible mutant fish
. Great care was taken to note which pair (F0 and wildtype fish of opposite sex) was in each box. Once in the cube, a special code was given depending on where the pair was in the tank. For example, the pair with Founder 3C was found in column C in the third row. This way we could match the individual fish to their sample assessed by PCR and restriction enzyme digestion.

Discussion

Our preliminary data suggests that *ccdc141* is important for zebrafish heart development due to the high expression in the cardiac cells of mouse embryos. However, the role of *ccdc141* is unknown, and this experiment sought to discover its function. We sought to do this by disrupting the normal function of *ccdc141* and seeing is that could affect cardiac fusion. Towards the end, we initially sought to determine which regions of *ccdc141* were conserved and thus likely essential for its function. By aligning amino acid sequences of a variety of different organisms, we identified exons 2 and 3 as being highly conserved using the CLUSTAL algorithm. We then used CHOPCHOP to identify several possible target sites for CRISPR/Cas9 mediated mutagenesis. CHOPCHOP also identified primers and restriction enzyme sites for verification of mutagenesis by our sgRNA/Cas9 RNP complex. The left and right primers with the largest product sizes and least number of off-targets were chosen. This provides a better product with little error when amplified. Next, the development of the sgRNA was important to using the Crispr/Cas9 protein that is necessary to cause the cut in the DNA at the target site. While there were trace amounts of DNA in the first analysis of the gRNA, it proved to work efficiently with the primers and Crispr/Cas9 complex to make accurate cuts in the target site of *ccdc141*.

While none of the current results answer specifically at what role *ccdc141* plays in cardiac fusion and cardiac cell movement, they help to set the foundation for the future directions of this project. Important future experiments include establishing a stable F1 lineage that is heterozygous from the founder population. Once stabilized, this line could give rise to the F2 lineage. This F2 lineage, once three months or older, could be intercrossed to analyze homozygous mutants and determine the specific role of *ccdc141.* Additionally, another direction could include improvement to the CRISPR-Cas9 system. A study done by Wu et al. researches a more optimized and efficient way to use guide RNA to produce a gene knockout in a gene that is expressed in zebrafish hearts as well. Their technique used a “preassembled four-guide Cas9 RNP,” allowing for the rapid destruction of the gene function. They found that using four-guide sets, four guide RNAs for one gene, to target one gene allowed for better gene disruption and higher penetrance of the mutation. This technique created a line of null phenotypes, meaning that there was a knock out of the target gene and could be passed through the germline (Wu et al., 2018). While this technique is not identical to the one done in this experiment, it involves many of the same processes, such as injecting guide RNAs and the RNP complex into zebrafish embryos within the first 2-3 days of fertilization. Some of the defects they observed included “prominent atrial enlargement (AE), impaired atrioventricular separation, and decreased heart rates compared with a scrambled control guide-injected embryos” (Wu et al., 2018). These same defects could be possible phenotypes that are observed as results from this experiment when further analysis is done.

Another future direction is using next-generation sequencing (NGS)- focusing on the exons of *ccdc141* only (Bloomekatz, unpublished grant). Next-generation sequencing involves high-throughput, allowing lots of sequences to be analyzed at once. NGS could be used to reveal the specific location, efficiency of mutagenesis, and type of mutation caused by the knockout. The mutants confirmed with gel electrophoresis and future lines such as heterozygotes are prime options to use for samples to send for WGS. This future direction allows us to investigate the DNA sequence itself further and how that affects the function of *ccdc141*.

While this experiment has pushed towards discovery, more work is still needed. This experiment has revealed strong mosaicism in the founder population for the induced mutation in *ccdc141*. This strong propagation of the induced mutation is crucial to this experiment because the future generations of zebrafish could reveal more about *ccdc141*. The work in this experiment is necessary for understanding the role of *ccdc141* in zebrafish heart development as well as perfecting the technique used to create the mutation.

Works Cited

Avdesh, A., Chen, M., Martin-Iverson, M. T., Mondal, A., Ong, D., Rainey-Smith, S., … Martins, R. N. (2012). Regular care and maintenance of a Zebrafish (Danio rerio) laboratory: An introduction. *Journal of Visualized Experiments*, (69). https://doi.org/10.3791/4196

Bakkers, J. (2011). Zebrafish as a model to study cardiac development and human cardiac disease. *Cardiovascular Research*, *91*(2), 279–288. https://doi.org/10.1093/cvr/cvr098

Bloomekatz, Unpublished grant

Bradford, Y. M., Toro, S., Ramachandran, S., Ruzicka, L., Howe, D. G., Eagle, A., … Westerfield, M. (2017). Zebrafish models of human disease: Gaining insight into human disease at ZFIN. *ILAR Journal*, *58*(1), 4–16. https://doi.org/10.1093/ilar/ilw040

Burger, A., Lindsay, H., Felker, A., Hess, C., Anders, C., Chiavacci, E., … Mosimann, C. (2016). Maximizing mutagenesis with solubilized CRISPR-Cas9 ribonucleoprotein complexes. *Development (Cambridge)*, *143*(11), 2025–2037. https://doi.org/10.1242/dev.134809

Den Hoed, M., Eijgelsheim, M., Esko, T., Brundel, B. J. J. M., Peal, D. S., Evans, D. M., … Loos, R. J. F. (2013). Identification of heart rate-associated loci and their effects on cardiac conduction and rhythm disorders. *Nature Genetics*, *45*(6), 621–631. https://doi.org/10.1038/ng.2610

Effects, S. A. (2008). Zebrafi sh as a Model for Studying Adult. *Physiology & Behavior*, 113–120.

Howe, D. G., Bradford, Y. M., Conlin, T., Eagle, A. E., Fashena, D., Frazer, K., … Westerfield, M. (2013). ZFIN, the Zebrafish Model Organism Database: Increased support for mutants and transgenics. *Nucleic Acids Research*, *41*(D1), 854–860. https://doi.org/10.1093/nar/gks938

Howe, D. G., Bradford, Y. M., Eagle, A., Fashena, D., Frazer, K., Kalita, P., … Westerfield, M. (2017). The Zebrafish Model Organism Database: New support for human disease models, mutation details, gene expression phenotypes and searching. *Nucleic Acids Research*, *45*(D1), D758–D768. https://doi.org/10.1093/nar/gkw1116

Hruscha, A., Krawitz, P., Rechenberg, A., Heinrich, V., Hecht, J., Haass, C., & Schmid, B. (2013). Efficient CRISPR/Cas9 genome editing with low off-target effects in zebrafish. *Development (Cambridge)*, *140*(24), 4982–4987. https://doi.org/10.1242/dev.099085

Hsu, P. (2014). *Dritter voläufiger bericht über die von der notgemeischaft der deutschen wissenschaft in Uruk unternommenen ausgrabungen*. *157*(6), 1262–1278. https://doi.org/10.1016/j.cell.2014.05.010.Development

Hutchins, B. I., Kotan, L. D., Taylor-Burds, C., Ozkan, Y., Cheng, P. J., Gurbuz, F., … Wray, S. (2016). CCDC141 mutation identified in anosmic hypogonadotropic hypogonadism (kallmann syndrome) alters GnRH neuronal migration. *Endocrinology*, *157*(5), 1956–1966. https://doi.org/10.1210/en.2015-1846

Jiang, F., & Doudna, J. A. (2017). *CRISPR – Cas9 Structures and Mechanisms*. 505–531.

Karvelis, T., Gasiunas, G., Miksys, A., Barrangou, R., Horvath, P., & Siksnys, V. (2013). *Rna-10-841*. (May), 841–851.

Koonin, E. V., & Makarova, K. S. (2013). CRISPR-Cas: Evolution of an RNA-based adaptive immunity system in prokaryotes. *RNA Biology*, *10*(5), 679–686. https://doi.org/10.4161/rna.24022

Ma, Y., Zhang, L., & Huang, X. (2014). Genome modification by CRISPR/Cas9. *FEBS Journal*, *281*(23), 5186–5193. https://doi.org/10.1111/febs.13110

Makarova, K. S., Haft, D. H., Barrangou, R., Brouns, S. J. J., Charpentier, E., Horvath, P., … Koonin, E. V. (2011). Evolution and classification of the CRISPR-Cas systems. *Nature Reviews Microbiology*, *9*(6), 467–477. https://doi.org/10.1038/nrmicro2577

Miquerol, L., & Kelly, R. G. (2013). Organogenesis of the vertebrate heart. *Wiley Interdisciplinary Reviews: Developmental Biology*, *2*(1), 17–29. https://doi.org/10.1002/wdev.68

Ota, S., Hisano, Y., Ikawa, Y., & Kawahara, A. (2014). Multiple genome modifications by the CRISPR/Cas9 system in zebrafish. *Genes to Cells*, *19*(7), 555–564. https://doi.org/10.1111/gtc.12154

Riera, M., Burguera, D., Garcia-Fernàndez, J., & Gonzàlez-Duarte, R. (2013). CERKL Knockdown Causes Retinal Degeneration in Zebrafish. *PLoS ONE*, *8*(5), 1–12. https://doi.org/10.1371/journal.pone.0064048

Russell, J. J., Theriot, J. A., Sood, P., Marshall, W. F., Landweber, L. F., Fritz-Laylin, L., … Brunet, A. (2017). Non-model model organisms. In *BMC Biology* (Vol. 15). https://doi.org/10.1186/s12915-017-0391-5

Sanders, G. E. (2012). Zebrafish housing, husbandry, health, and care: IACUC considerations. *ILAR Journal*, *53*(2), 205–207. https://doi.org/10.1093/ilar.53.2.205

Santoriello, C., & Zon, L. I. (2012). Science in medicine Hooked ! Modeling human disease in zebrafish. *Science in Medicine*, *122*(7), 2337–2343. https://doi.org/10.1172/JCI60434.combines

Shen, M.W., & Arbab, M. (2018). Predictable and precise template-free CRISPR editing of pathogenic variants. *Nature,* 563(7733):646-651. doi: 10.1038/s41586-018-0686-x

Singleman, C., & Holtzman, N. G. (2012). Analysis of postembryonic heart development and maturation in the zebrafish, Danio rerio. *Developmental Dynamics*, *241*(12), 1993–2004. https://doi.org/10.1002/dvdy.23882

Singleman, C., & Holtzman, N. G. (2014). Growth and maturation in the zebrafish, Danio Rerio: A staging tool for teaching and research. *Zebrafish*, *11*(4), 396–406. https://doi.org/10.1089/zeb.2014.0976

Tian, X., Gu, T., Patel, S., Bode, A. M., Lee, M.-H., & Dong, Z. (2019). CRISPR/Cas9 – An evolving biological tool kit for cancer biology and oncology. *Npj Precision Oncology*, *3*(1). https://doi.org/10.1038/s41698-019-0080-7

van de Vegte, Y. J., Tegegne, B. S., Verweij, N., Snieder, H., & van der Harst, P. (2019). Genetics and the heart rate response to exercise. *Cellular and Molecular Life Sciences*, *76*(12), 2391–2409. https://doi.org/10.1007/s00018-019-03079-4

Wu, R. S., Lam, I. I., Clay, H., Duong, D. N., Deo, R. C., & Coughlin, S. R. (2018). A Rapid Method for Directed Gene Knockout for Screening in G0 Zebrafish. *Developmental Cell*, *46*(1), 112-125.e4.

https://doi.org/10.1016/j.devcel.2018.06.003