The University of Maine
DigitalCommons@UMaine

Honors College

Spring 5-2016

# Diversity of Small RNA Expression During Zebrafish Caudal Fin Regeneration

Jefferson Adams
*University of Maine*

Follow this and additional works at: https://digitalcommons.library.umaine.edu/honors

 Part of the Biology Commons

## Recommended Citation

Adams, Jefferson, "Diversity of Small RNA Expression During Zebrafish Caudal Fin Regeneration" (2016). *Honors College*. 365.
https://digitalcommons.library.umaine.edu/honors/365

This Honors Thesis is brought to you for free and open access by DigitalCommons@UMaine. It has been accepted for inclusion in Honors College by an authorized administrator of DigitalCommons@UMaine. For more information, please contact um.library.technical.services@maine.edu.

# DIVERSITY OF SMALL RNA EXPRESSION DURING ZEBRAFISH CAUDAL FIN REGENERATION

By

Jefferson Adams

A Thesis Submitted in Partial Fulfillment

of the Requirements for a Degree with Honors

(Biology)

The Honors College

University of Maine

May 2016

Advisory Committee:

Keith W. Hutchison, Professor Emeritus of Biochemistry & Molecular Biology, Advisor
Benjamin King, Senior Staff Scientist, MDI Biological Laboratory, Advisor
Sally D. Molloy, Assistant Professor, Honors College and Department of Molecular and Biomedical Sciences.
Clarissa Henry, Associate Professor of Biological Sciences
Karen Pelletreau, Lecturer and Research Associate; School of Biology and Ecology

**Abstract**

Over the years microRNA have been shown to play a role in the regulation of genes involved in regeneration of zebrafish (*Danio rerio)* tissues. However, recent research suggest that there may be other types of small RNA that play a regulatory role in these regenerative processes. For the most part these other small RNA (sRNA), such as transfer RNA (tRNA) derived fragments, ribosomal RNA (rRNA) derived fragments, and small nucleolar RNA, are disregarded. Here I analyzed the expression pattern of these sRNA during the regeneration of the zebrafish caudal fin. High-throughput sequencing was used to characterize the expression of small RNAs during zebrafish caudal fin regeneration at zero and four days post amputation using biological triplicate samples. Sequence data for each sample were processed to generate a set of 100,170 unique sequence tags that were present in all samples. The abundance of each tag in each sample was recorded as the number of reads that exactly matched the tag. Tags were annotated to various sRNA classes through sequence alignment to small RNA and genome databases in a hierarchical fashion. The observed abundance and the differential expression of the various types of sRNA were analyzed. This analyses showed that microRNA and tRNA derived fragments were the most abundant sRNA types. These included miR-21, tRNA412-LysCTT, and a novel small RNA, Tag_41941 5'-GGGATGAGTCCCTATCACGGCACAGGAGTGTGACTTT-3'. The differential expression of these tRNA derived fragments and previous research into gene silencing by sRNA suggest that these other sRNA have the potential to play a regulatory role in zebrafish caudal fin regeneration.

**Table of Contents**

**List of Figures and Tables**

## Introduction

Many animals including lower vertebrates are capable of regenerating most tissues. Humans, however, have a very limited ability to regenerate damaged tissues. Some exceptions include part of the liver and blood. Our impaired ability to repair and regenerate injured or diseased tissue such as after a heart attack or amputation drastically impacts human health. However, if a mechanism was found to promote tissue regeneration following severe damage or amputation then a better quality of life could be obtained for these patients. After cardiac events patients would be able to have their heart tissue completely restored versus just developing scar tissue (Jaźwińska A, Sallin P). Perhaps a mechanism for preventing diseases caused by old age could be found as well. Because human beings cannot regenerate most of our damaged organs or tissues, the mechanisms for controlling this process have to be studied in organisms that can naturally regenerate body tissues such as heart tissue, appendages, or even the spinal cord (Thatcher EJ et al.). Zebrafish (*Danio rerio*) make a great model organism as they have this remarkable regenerative ability and because of their role as a major model organism they already have a nearly completely sequenced genome (Poss KD et al.).

Studies of zebrafish caudal fin regeneration have shown that within the first 48 hours after amputation epithelial cells migrate to the amputation site to cover the wound (Chen CH, et al.). Following this stage the blastema, a mass of proliferating cells primarily made up of dedifferentiated mesenchymal cells, begins to form (Jaźwińska A, Sallin P, Poss KD et al. and Poleo G et al.). The blastema differentiates into the needed tissue types at four days post amputation (Poss KD et al.). This temporally controlled

process of regeneration allows for the investigation of differential expression of genes to be examined from a state immediately following amputation to the initial outgrowth of the new fin tissue.

Research that has already been done looking at control mechanisms for this regenerative process has found that microRNA (miRNA) play a regulatory role during zebrafish fin regeneration (Thatcher EJ et al., Yin VP et al.). MiRNA are short sequences of non-coding RNA that target messenger RNA (mRNA) in a sequence specific manner (He L. and Hannon G. J.) Examples of these miRNA include miR-203 and miR-133, which were both found to play a regulatory role in the regeneration pathway (Thatcher EJ et al., Yin VP et al.). These miRNA are derived from either transcribed sections of DNA that contain multiple miRNA or these miRNA arise from an intron that was excised from a mRNA strand (Bartel DP). The enzymes Dicer and Drosha cleave the stem-loop structure of the miRNA to form a short double stranded molecule approximately 22 nucleotides in length (Carthew RW, Sontheimer EJ). Following this the dsRNA strands separate and one of the miRNA strands binds to the protein argonaute, which is part of the RNA-induced silencing complex (RISC) while the other miRNA strand is degraded by the cell (Carthew RW, Sontheimer EJ and Ender C and Meister G). The miRNA sequence allows RISC to target a very specific complementary sequence of mRNA so as to initiate the degradation of the mRNA or block translation from occurring (Ender C, Meister G). If the miRNA aligns completely to the mRNA then degradation will be triggered. If there is not complete complementary alignment then translation will be blocked and degradation of the mRNA will occur more gradually (Ender C, Meister G).

There are multiple models for how RISC goes about blocking translation in the case where there is incomplete alignment. The first is co-translational protein degradation, which involves RISC recruiting enzymes to break down the protein strand while it is being produced by the ribosome. The second is ribosomal drop off, where RISC causes the ribosome to dissociate from the mRNA strand before translation is complete. There is a protein domain in the argonaute complex that binds to the 5' cap of the mRNA strand thus creating competition between argonaute and the eukaryotic translational initiation factor, eIF4E, which is the protein responsible for the binding of the mRNA cap so that translation can occur. In the last mechanism argonaute is able to complex with elF6 and the large ribosomal subunit thus preventing the small ribosomal subunit from associating with the mRNA strand and the large subunit. (Carthew RW, Sontheimer EJ and Eulalio A et al.).

We are currently aware of miRNA and their regulatory role in regeneration. However, there are other sRNA besides miRNA that may also play a role in regulating these regenerative pathways. The more common ones are transfer RNA (tRNA) derived fragments, ribosomal RNA (rRNA) derived fragments, and small nucleolar RNA (snoRNA). There are also partial mitochondrial rRNA (Mt-rRNA) and partial mitochondrial tRNA (Mt-tRNA). The tRNA derived fragments in particular were found to have an association with argonaute, suggesting that these sRNA have similar gene-silencing functions as miRNA (Chen, et al. 2016, Shigematsu M, Yohei K, Sharma U et al.). While having an association with argonaute is not enough evidence to conclude that these tRNA derived fragments have this gene-silencing quality, it is definitely an

important initial step of the pathway (Thomson DW et al). There is, in fact, growing evidence that sRNA do have a regulatory function in biological systems and this opens the door to new possible regulatory pathways that we can use to better understand how our protein coding genes are being regulated. The four different mechanisms possible for how miRNA are capable of blocking translation are a prime examples. I would speculate that other small RNA (sRNA), besides miRNA, could also be using one of these mechanisms or perhaps a new one that has not been considered yet to silence particular protein coding genes. The term sRNA is used to describe a group of all short non-coding RNA sequences, miRNA are the most researched type of sRNA but there are other types of sRNA as well including transfer RNA (tRNA) derived fragments, ribosomal RNA (rRNA) derived fragments, and small nucleolar RNA (snoRNA) to name a few (Carthew RW, Sontheimer EJ).

In previous work, RNA samples were collected from the caudal fin amputation site of six biological replicate zebrafish. Then using reverse transcriptase these RNA sequences were converted into DNA. Using high-throughput DNA sequencing, it is then possible to characterize the expression of sRNA in tissue samples in a non-biased fashion (King BL, Yin VP). In my research, I used this dataset of short DNA sequences found in regenerating caudal fin tissues of six biological replicate zebrafish. These short sequences were 15-48 base pairs (bp) in length after adapter clipping. These data profiled sRNA expression from immediately following amputation (0 dpa) to regenerating tissues (4 dpa) after the blastema was formed. These 15-48 bp reads were then collapsed into unique sequence tags. Hereafter the term “tag(s)” will refer to the short DNA sequences

that are expressed in response to a stimulus, in this case the amputation of the caudal fin. These tags represent all sRNA expressed in these tissues within this size range. They represent miRNAs and many other known and novel sRNAs. My research focuses in on these tags.

## Literature Review

### *Zebrafish Regeneration*

The ability to regenerate damaged tissues or organs allows an organism to hold off morbidity much better than repair mechanisms that lead to scarring (Poss KD et al.). Humans, only have the ability to regenerate particular tissues or organs such as blood or part of the liver. Other vertebrates including zebrafish do have the ability to regenerate damaged tissues or organs such as appendages or part of the heart (Poss KD et al.). Studying zebrafish fin regeneration provides an understanding of genes and mechanisms involved in the regenerative process (Poleo G et al.). When researching zebrafish fin regeneration it has been found that the first stage of regeneration is for epithelial cells to cover the wound shortly after amputation (Chen CH, et al.). These epithelial cells in the zebrafish caudal fin migrated towards the wound site from as far as 750μm away in the first 48 hours after amputation occurred (Poleo G et al.). After a 48-hour period the cell migration was much more localized around the amputation site. These epithelial cells that migrate to the wound site form a structure called the wound epithelium (Jaźwińska A, Sallin P). Following this event is the formation of a blastema, which is a mass of proliferating cells that develops along the injury site, and is crucial to the regenerative process (Jaźwińska A, Sallin P). Fibroblast growth factors (Fgfs) were also found to play

a role in the development of the blastema (Poleo G et al.). The mesenchymal cells that make up the blastema are not as differentiated as the mature cells in the zebrafish fin (Poss KD et al.). However, how this dedifferentiation comes about is not certain. Dedifferentiation of the cells near the wound site and/or the migration of previously dormant stem cells could be possible sources of the less differentiated cells (Poss KD et al.). It is also possible the mesenchymal cells near the amputation site dedifferentiate to form the blastema (Poleo G et al.). Furthermore, communication between the epithelial and mesenchymal cells near the amputation site is crucial for the regenerative process and msx genes, which are responsible for maintaining the organization of the cells during regenerative outgrowth of the caudal fin, play a major role in the regeneration of the tissue (Poleo G et al., Song K et al.). This differentiation of the blastema into the needed tissue types begins at four days post amputation in zebrafish (Poss KD et al.).

***Small RNA Gene Silencing Mechanisms***

In the past 10 years our understanding of the regulation of various protein encoding genes has expanded tremendously. Small noncoding genes have been shown to play a role in these regulatory pathways. In previous research sRNA were divided up into three types: microRNA (miRNA), small interfering RNA (siRNA), and piwi-interacting RNA (piRNA) (Carthew RW, Sontheimer EJ). Of these three types they primarily focused on miRNA and siRNA. The authors describe miRNA as the regulators of the organism's' own genes while siRNA are generally responsible for gene transcripts coming from viruses or other potential invaders. They go on to state that siRNA also have the ability to silence genes from the same area of the genome from which they were

transcribed while miRNA silence mRNA sequences from other locations in the genome (Carthew RW, Sontheimer EJ).

The siRNA and miRNA have been shown to have similar pathways leading to their maturation (Carthew RW, Sontheimer EJ). For the sake of simplicity and since the pathways of these two sRNA are so similar I will only be discussing the miRNA pathway. The double stranded RNA molecule that will become a mature miRNA is transcribed from either sections of noncoding genes in the DNA with multiple miRNA or as part of an intron that was excised during mRNA maturation (Bartel DP). The primary miRNA transcript folds into  a stem-loop structure that is cleaved by the enzymes Dicer and Drosha leaving a short dsRNA with 3' overhangs of roughly 2 nucleotides (Bartel DP). Drosha and an associating cofactor make the first cut in the development of the mature miRNA (Carthew RW, Sontheimer EJ). Dicer then interacts with the 3' overhang and makes the second cut, which is two helical turns away from the 3' end or about 22 nucleotides. Once the dsRNA has been cleaved it will bind to one or more proteins in the argonaute superfamily. During this process the dsRNA unwinds and the strands separate leaving only a single RNA strand to bind with argonaute and make up the RNA-Induced Silencing Complex (RISC) meanwhile the other strand is degraded by the cell (Carthew RW, Sontheimer EJ).

The activity of the miRNA is dependent on its relationship to the target mRNA. When there is complete complementary binding of the miRNA to its mRNA target immediate degradation of the mRNA transcript will be triggered. Incomplete base pairing or the formation of a central bulge in the miRNA mRNA duplex will block translation

(Carthew RW, Sontheimer EJ and Eulalio A et al.). In the latter case there are four proposed mechanisms for how inhibition of translation occurs. The process starts with miRNA binding to the target mRNA, most often to the 3' UTR . This creates a problem because translation is initiated at the 5' cap of the mRNA strand (Carthew and. Sontheimer. 2009). For the following proposed mechanisms the miRNA acts as the sequence sensitive component responsible for identifying the target mRNA strand. The first of the proposed translation inhibition mechanisms states that the mRNA transcript it prevented from binding to the ribosome due to competition for the 5' cap. Typically, the 5' cap binds to eukaryotic translational initiation factor 4E (eIF4E) to allow translation to occur, however, there is a protein domain in argonaute that is very similar to the active site of eIF4E. Thus creating competition between the two active sites for the 5' cap. If the 5' cap binds to argonaute instead of eIF4E then translation will not be able to occur (Carthew RW, Sontheimer EJ and Eulalio A et al. and Kiriakidou M et al.). An additional mechanism suggests that argonaute also has the potential to bind to eukaryotic translational initiation factor 6 (elF6), which will prevent the large ribosomal subunit from associating with the small subunit, thus preventing translation from occurring (Eulalio A et al. (Chendrimada et al., 2007)). A third mechanism proposed for translation inhibition is a premature termination of translation due to the ribosome dissociating from the mRNA (Carthew RW, Sontheimer EJ and Eulalio A et al.). Lastly, RISC recruits enzymes to degrade the protein as it is being constructed by the ribosome. This mechanism was termed co-translational protein degradation (Carthew RW, Sontheimer EJ and Eulalio A et al.). It seems that there are several possible pathways in which

miRNA could cause gene silencing. With multiple pathways proposed that lead to gene silencing I speculate that other sRNA types could also take part in one of these pathways in which gene silencing could be achieved.

***Research With miRNA Involved In Fin And Heart Regeneration***

Previous research found that a miRNA pathway, particularly miR-203, is required to regulate fin regeneration in zebrafish (Thatcher EJ et al.). When miR-203 was expressed Lef1, which is a transcription factor for the Wnt signaling pathway, was repressed and regeneration did not occur. When miR-203 was repressed Lef1 was over expressed and overgrowth of the fin occurred. During regeneration miR-203 was found to be down regulated and Lef1 to be up regulated. A miRNA pathway is needed for successful zebrafish fin regeneration (Thatcher EJ et al., Yin VP et al.). However, these authors focus their research on different miRNA. While one of them focusses on miR-203, the other found that miR-133 also plays an important role in the regulation of fin regeneration (Thatcher EJ et al., Yin VP et al.). This prior research found that miR-133 had high expression levels in normal fins but was down regulated during fin regeneration (Yin VP et al.). The same authors also conducted another experiment looking at zebrafish heart regeneration and found that again miR-133 plays a major regulatory role in the regeneration process(Yin VP et al.). They found that when miR-133 levels were increased the amount of heart regeneration was greatly reduced. Meanwhile, when miR-133 levels were decreased heart regeneration was increased. These experiments highlight an important notion that, as in most biological processes, more than one mechanism is working toward the same goal, in this case regeneration. With that in mind it is possible

that other types of sRNA could also be involved in this regulatory process at varying levels. By categorizing different sRNA and building a foundation we can begin to look at different candidate sRNA that may also play a regulatory role in regeneration.

***Other sRNA***

The research reported on in this thesis is focused on determining if there are sRNAs such as tRNA derived fragments, rRNA derived fragments, and snoRNA that may play a regulatory role in the regeneration of the zebrafish caudal fin. Using the miRNA maturation pathway and the proposed translation inhibition pathways as templates researchers have started to explore whether or not these other sRNA have the gene-silencing ability of miRNA. Due to the amount of research already conducted there is more known about tRNA derived fragments than about the other kinds of sRNA. While the research conducted with tRNA derived fragments is very recent, it focuses on specific examples versus looking at the expression of these sRNA as a whole . This lack of information about snoRNA and rRNA derived fragments in itself supports the need for more research into these other types of sRNA and their potential role in gene regulation.

Typically snoRNA play a major precursory role in the development of non-coding RNA such as rRNA and small nuclear RNA (Ender C et al., Matera et al. 2007). Unlike miRNA, snoRNA can be processed without binding to Drosha; however, they do require binding to Dicer to be cleaved appropriately. Mature snoRNA then complex with argonaute to function as the sequence sensitive component of RISC to potentially block translation (Carthew RW, Sontheimer EJ and Ender C et al.).

Several papers mention rRNA derived fragments when discussing possible sRNA involved in regulation of genes involved in the regeneration process. However, very few of these articles discuss rRNA derived fragments in depth or elaborate on specific rRNA derived fragments and possible roles they play in the regulation of genes (Thomson DW et al). My research included rRNA derived fragments so as to develop an initial understanding of the expression pattern of these sRNA that might be incorporated into the regeneration process.

Recent research was done examining the gene regulatory effects of tRNA derived fragments also called transfer sRNA (tsRNA). In particular this research was studying the regulatory roles of tsRNA in mice embryos as a result of a modified diet of the male in the F0 generation (Chen Q et al., Sharma U et al.). These tsRNA ranged from 30-34 nucleotides in length, additionally, the researchers found that tsRNA arose predominantly from the 5' end of mature tRNA and seemed to be incorporated into a paternal mouse's sperm that then fertilized the oocyte(Chen Q et al.). When the paternal mouse was fed a high fat diet the F1 generation developed glucose intolerance and insulin resistance. When the F1 oocytes were injected with purified tsRNA the F1 generation developed glucose intolerance but did not develop insulin resistance. The authors speculate that DNA methylation or histone arrangement could be part of this trigger for insulin resistance in the F1 generation since insulin resistance did not develop when only the tsRNA were injected into the oocyte (Chen Q et al.). Additionally, while these tsRNA are derived from the 5' end of a tRNA transcript, there were tRNA derived fragments recently found that were derived from just beyond the 3' end of the tRNA transcript

(Pekarsky Y et al.). These tRNA derived fragments that were studied were found to have a gene regulating function of T-cell leukemia/lymphoma 1 (TCL1) which, when unregulated leads to Chronic Lymphocytic Leukemia (Perkarsky Y et al.). This novel tRNA derived fragment location raises the notion that there is still a great deal that we do not know about these tsRNA and that more research is needed to better understand how these tsRNA are produced and what their role is in biological systems. An increased expression level of tRNA derived fragments could be caused by an increased tRNA turnover rate during regeneration. However, previous research suggest that these tRNA derived fragments can play a regulatory role in other biological pathways (Perkarsky Y et al.). Therefore, increased expression levels of tRNA derived fragments should not be discounted and need to be investigated further.

Previous research has shown that an association with argonaute is important for identifying if a sRNA type has gene silencing qualities (Chen Q et al., Thomson DW et al., Sharma U et al.). These studies also show that snoRNA, rRNA, and tRNA all have an association with argonaute. However, it should be noted that this association alone is not enough to make the claim that a particular sRNA has a gene silencing function without more supporting evidence (Thomson DW et al.). Further research found that after conducting another experiment the snoRNA, tRNA, and rRNA were unable to suppress the reporter gene even at very high concentrations of these sRNA (Thomson DW et al.). This lead to the conclusion that it is unlikely these sRNA have gene-silencing roles (Thomson DW et al.). However, a recently published paper suggests that these tRNA derived fragments do indeed play a regulatory role as a result of binding to argonaute and

that they even have a similar maturation pathway as miRNA including being cleaved by Dicer (Shigematsu M, Yohei K).

The literature reveals that we know a good deal about miRNA and their regulatory role in multiple biological pathways, including regeneration. We do not yet know much about the other types of sRNA or what sort of roles these other sRNA might have in a developmental system. My research has been focused on categorizing the various sRNA types and identifying new potential sRNA to be tested in the future. However, understanding the proposed mechanisms of how these other sRNA, such as tRNA derived fragments, snoRNA, and rRNA derived fragments, might function similarly to miRNA is important in perhaps extending those mechanisms to more sRNA types in the future.

My hypothesis is that these sRNA, other than miRNA, play a regulatory role in regeneration of the caudal fin of zebrafish. To address this issue I have taken a much broader look at the sRNA that are present in zebrafish following the amputation of the caudal fin. In doing so I hope to help develop a better understanding of the various kinds of sRNA present under these conditions and speculate how some of these sRNA could affect the regeneration process. To observe the sRNA that are being expressed during this regenerative process as a result of the formation of the blastema, which is a proliferating mass of dedifferentiated cells that will give rise to the new fin outgrowth and matures at four days post amputation, my research focuses on the differential expression of small RNA (sRNA), specifically sRNA other than miRNA, from zero days post amputation to four days post amputation.

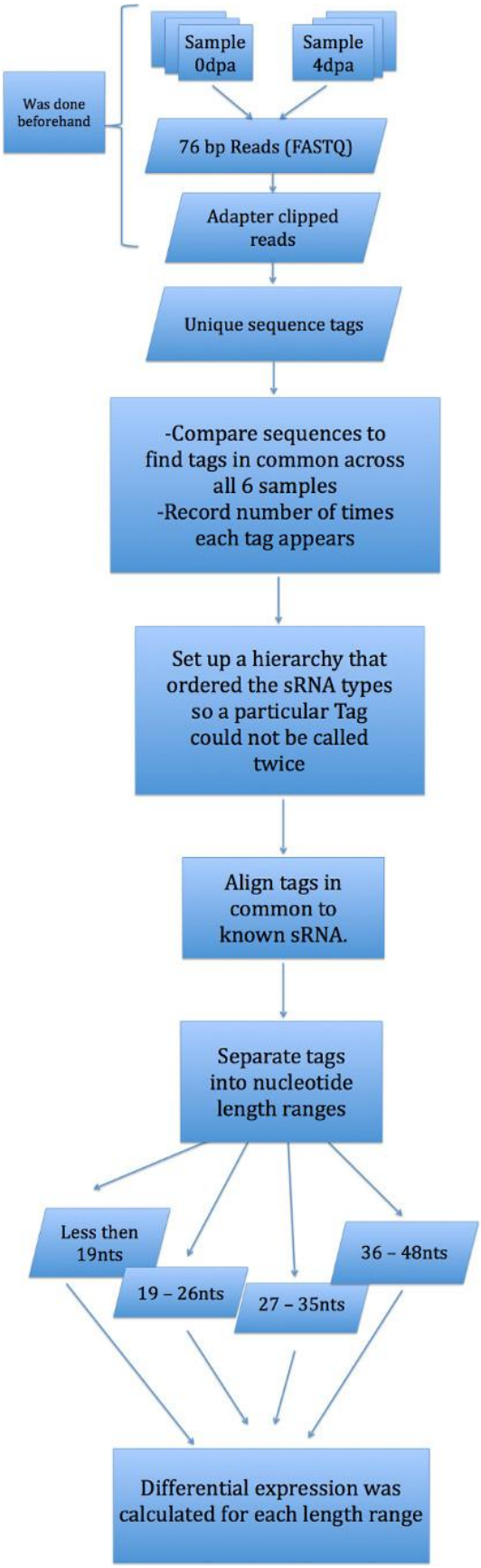


**Figure 1:** Methods Flowchart. Outlining the research that was done. Regular rectangles denote processing steps whereas slanted rectangles are for data and files.

## Materials and Methods

### *Data Collection*

Illumina high-throughput sRNA sequence data were obtained from six biological replicate zebrafish samples following the amputation of the caudal fin (King BL, Yin VP). To observe the expression changes of these sRNA, RNA was collected from the amputation site in the zebrafish caudal fin. Three of the samples were from zero days post amputation (0 dpa) and three of the samples were taken four days post amputation (4 dpa). Total RNA from each sample were used to generate barcoded small RNA sequencing libraries for each sample following manufacturer's protocols (Illumina, San Diego, CA). Sequence library construction involves producing complementary DNA (cDNA) from the RNA and ligating two different adapters to the 5' and 3' ends of the DNA

strands so as to provide a known sequence of nucleotides that could then be used as primer binding sites and oligo binding sites when the sample was in the flowcell (Mardis ER). The sRNA sequencing libraries were characterized using 76 base pair (bp) reads. This was done to put a limit on the number of DNA sequence that were received in the output to make the data more manageable. The adapter sequences were removed (adapter clipping) leaving only the DNA that represents the small RNA expressed in the samples.

***Data Analysis***

The six samples were compared to each other using the Python script LEN_Final_Version_0dpa+4dpa.pyw. This script identifies the tags that were common to all six samples and records the number of times that each tag appeared in each sample (read count)(Fig. 1)(Appendix A). The read count number was used in identifying the expression level of a particular nucleotide sequence. The read counts were summed for the three 0 dpa samples and for the three 4 dpa samples. This information was used to examine the expression levels of the tags and to ultimately calculate the differential expression of each tag.

***Data Organization***

To prevent a tag from being called by more than one sRNA class the Python script, Tab_organized_data.pyw, was developed to order the tags that aligned to a canonical sRNA sequence into a hierarchy (Fig. 1)(Appendix B). This hierarchy was structured as such: miRNA, tRNA, snoRNA, Mitochondrial tRNA (Mt-tRNA), Mitochondrial rRNA (Mt-rRNA), rRNA, and then a second rRNA database. The miRNA were given priority because they are better understood and if a tag mapped to a miRNA I

could be more assured that it was correct. The remaining sRNA types were placed following similar logic as to where they seemed most appropriate and the degree of confidence that could be placed in the sRNA type assignment. If a tag did not align to any of these databases with a high degree of accuracy it was called as an "unknown sRNA".

Tags were aligned using BLAST (Geer, LY et al.) to various sRNA databases: MirGeneDB, NCBI's BLAST, Ensembl, Silva, and The Genomic tRNA Database (Fig. 1). The collective output from these alignments was compiled into a tab deliminated file. The alignments for these tags were restricted using the same Python script, Tab_organized_data.pyw, to only allow alignment scores of 18 or higher and an alignment percent of at least 95 to be called. This script also separated the tags by length into groups: less than 19 nucleotides (nt), 19 nt to 26 nt, 27 nt to 35 nt, and 36 nt to 48 nt. The number of sRNA of each type for each length range was determined and the data presented using pie charts as a graphical representation ofthe percent of each length range that a particular sRNA occupied. The expression level of each tag in each sample was analyzed to determine which were significantly differentially expressed between 0 and 4 dpa using the R/edgeR package (Robinson MD et al, 2010) in the R Statistical Computing Environment (http://www.r-project.org). Tags were considered to be differentially expressed if the false discovery rate(FDR)-adjusted p-values were less than 0.05. These data were also used to generate an MA plot, which compares the average level of expression of the tags to the log fold change (logFC) the tag. All of the tags that had a logFC with a FDR of less than 0.00001 were considered differentially expressed.

**Results**

A comparison of the results from the six zebrafish samples showed 100,170 unique sequence tags commonly expressed in all six samples. The abundance of each sRNA at a particular tag length was variable (Fig. 2). As part of a proof of concept experiment the most abundant tags were queried at ten nucleotide intervals. The tag lengths chosen were: 15, 25, 35, and 45 nucleotides. Since 22 nt was the most abundant tag length in these data it was also added to the tags queried (Fig. 2, Table 1). There were 3 tags that were 48 nucleotides in length though they are difficult to see in Fig. 2 due to scaling. Tag lengths 15-18 nts were all unknown sRNA due to the requirement of an alignment score of at least 18 for a tag to be aligned to a canonical sRNA sequence. The most abundant tag at this length was miR-22a and it was also the most down regulated gene (Table 1). The most abundant tag at tag length 23 nt was miR-21 and it was also the most up regulated gene. Tag length 34 nt was tRNA-Lys-AAG which, had the highest over all read count and tag length 48 nt which, was a rRNA, was the longest tag in these data (Table 1).

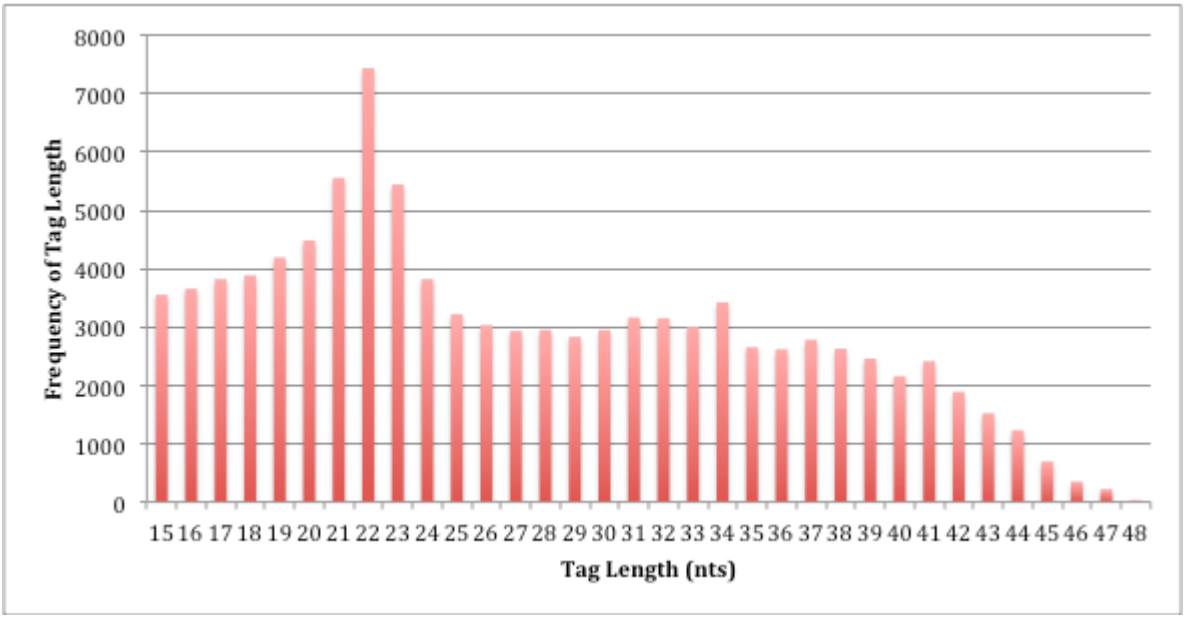


**Figure 2: Frequency Distribution of Tag Lengths.** Tags were sorted based on tag length without regard to the tag sequence.

**Table 1: Identity of The Most Abundant Tag For Each Length Class.**

| Tag | Number of Nucleotides | Symbol | Notes |
|---|---|---|---|
| TTTGCGGCCCCGGGTTCCTCCCGGGGCCACGCCTGTCTGGGCGTCGCT | 48 | rRNA | Longest Tag |
| GTACTTGGATGGGAGACCGCCTGGGAATACCAGGTGCTGTAAGCT | 45 | rRNA | |
| TACGACCTCAGATCAGACGAGACAACCCGCTGAAT | 35 | rRNA | |
| GCCCGGCTAGCTCAGTCGGTAGAGCATGAGACTC | 34 | tRNA-Lys-AAG | Highest Read Count |
| AACTCTTAGCGGTGGATCACTCGGC | 25 | rRNA | |
| TAGCTTATCAGACTGGTGTTGGC | 23 | miR-21 | Most up regulated gene |
| AAGCTGCCAGCTGAAGAACTGT | 22 | miR-22a | Most down regulated gene |
| TGTCTGAGCGTCGCT | 15 | N/A | Too short to identify |

Panel A in each figure shows the abundance of each sRNA type for all 100,170 tags, whereas panel B in each figure represents only the tags that were differentially expressed with a FDR < 0.00001 (Fig. 3-6). Of the 396 tags that were differentially expressed with an FDR below 0.00001 only 21 of them were down regulated (Fig. 8). The Mt-rRNA and Mt-tRNA each make up a very small portion of the sRNA population (Fig. 3A). Neither of these sRNA types had tags that were differentially expressed (Fig. 3B). Tags that aligned to miRNA and tRNA derived fragments account for the majority of the sRNA that were differentially expressed. Of the 8,456 total tags that were miRNA, 507 of these were isomiRs of miR-21. Of the 396 tags that were differentially expressed 197 were miRNA (Fig. 3B). IsomiRs of miR-21 account for 160 of the 197 miRNA (Fig. 3B+4B). Between the tag lengths 27-35 nt, 86% of the differentially expressed tags aligned to tRNA derived fragments (Fig. 5B). Of the differentially expressed tags between 35-48 nt 63% were unable to be aligned to any of the canonical sRNA sequences that were used (Fig. 6B). In this tag length range are also the 2 snoRNA that were differentially expressed (Fig. 6B). These 2 snoRNA were identified as SNORD79, and while they are the only 2 snoRNA that were differentially expressed, they are not very highly expressed compared to the other sRNA(Fig. 7).

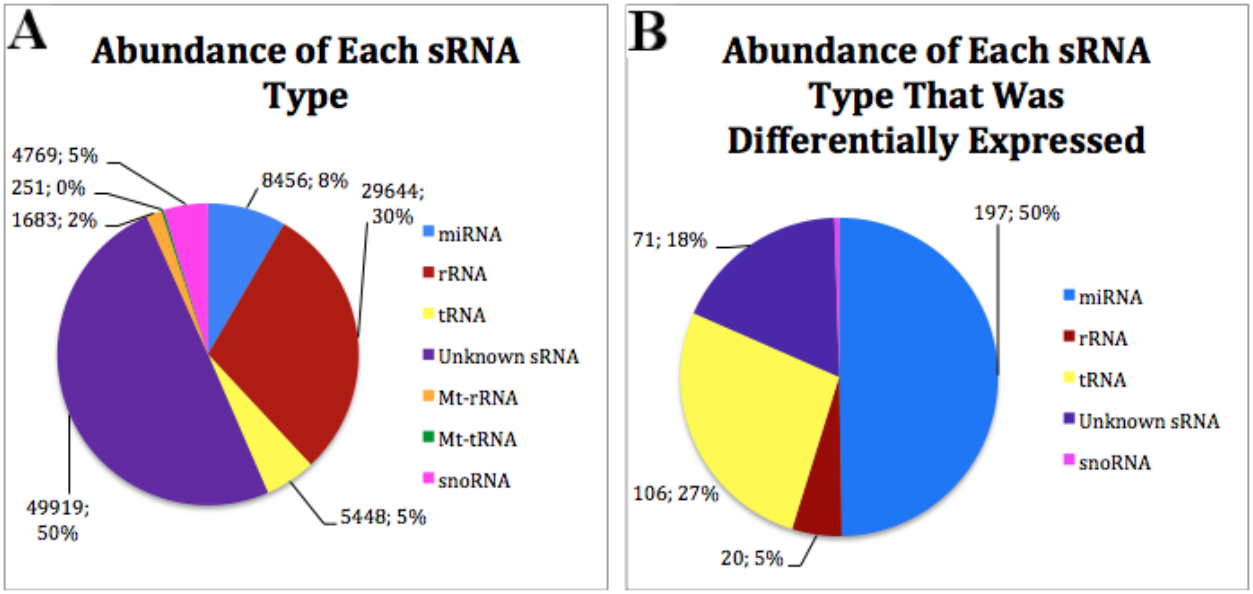


**Figure 3**: **Diversity of Tags Expressed During Caudal Fin Amputation Grouped by sRNA Functional Class.** 3A) Type and relative abundance of all 100170 tags; 3B) Type and relative abundance of the 396 differentially expressed tags.

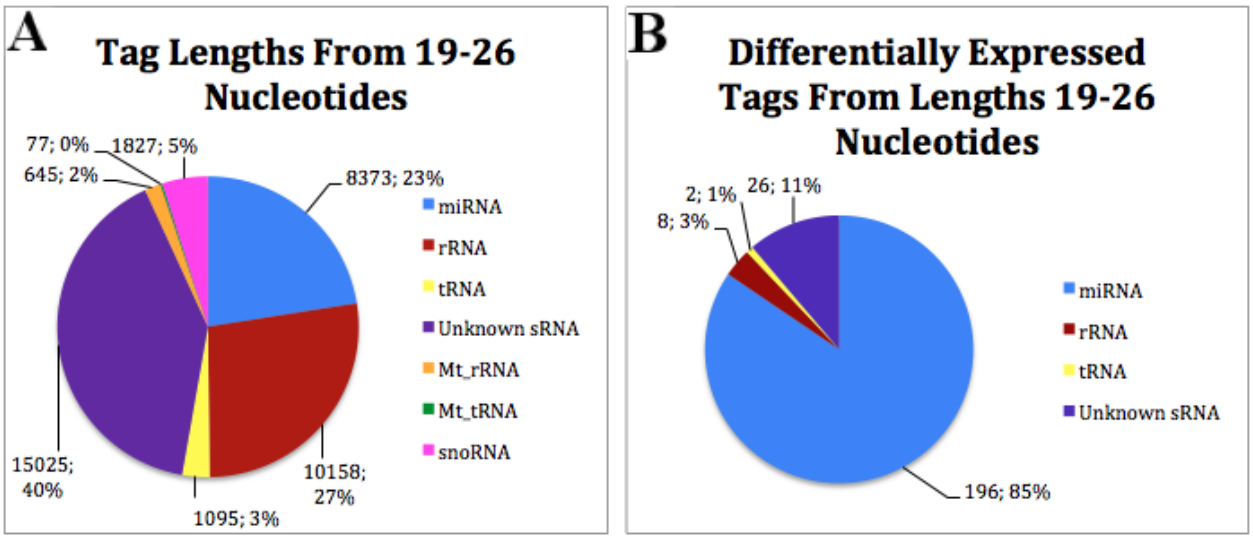


**Figure 4: Diversity of Tags Between 19-26 Nucleotides in Length Expressed During Caudal Fin Amputation Grouped by sRNA Functional Class**. 4A) Type and relative abundance of all tags 19-26 nt; 4B) Type and relative abundance of differentially expressed tags of 19-26 nt

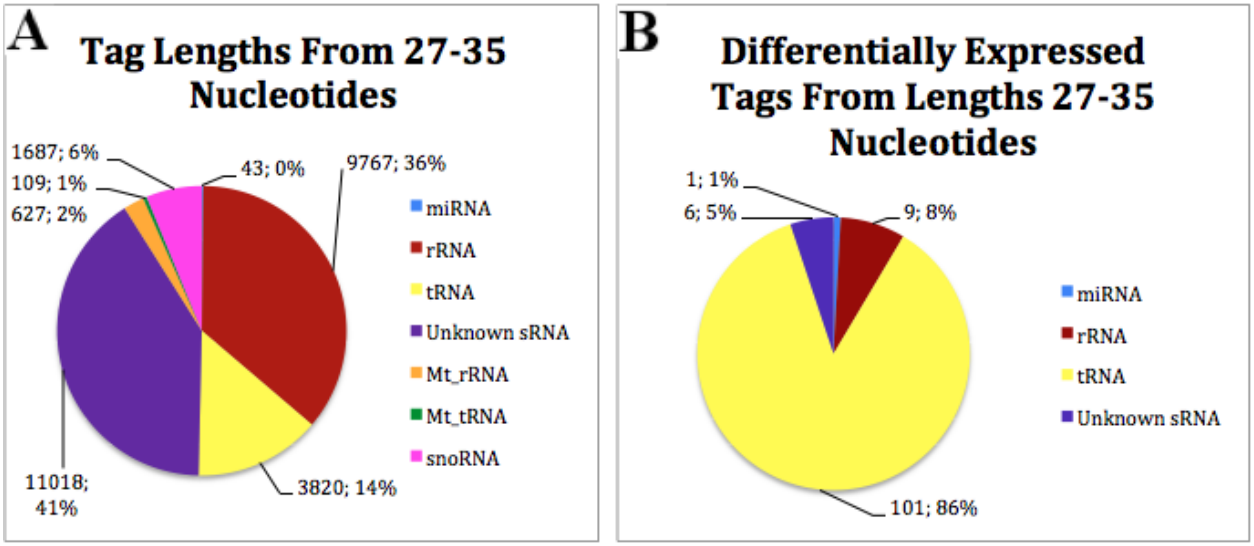


**Figure 5**: **Diversity of Tags Between 27-35 Nucleotides in Length Expressed During Caudal Fin Amputation Grouped by sRNA Functional Class.**
5A) Type and relative abundance of all tags 27-35 nt; 5B) Type and relative abundance of differentially expressed tags of 27-35 nt.

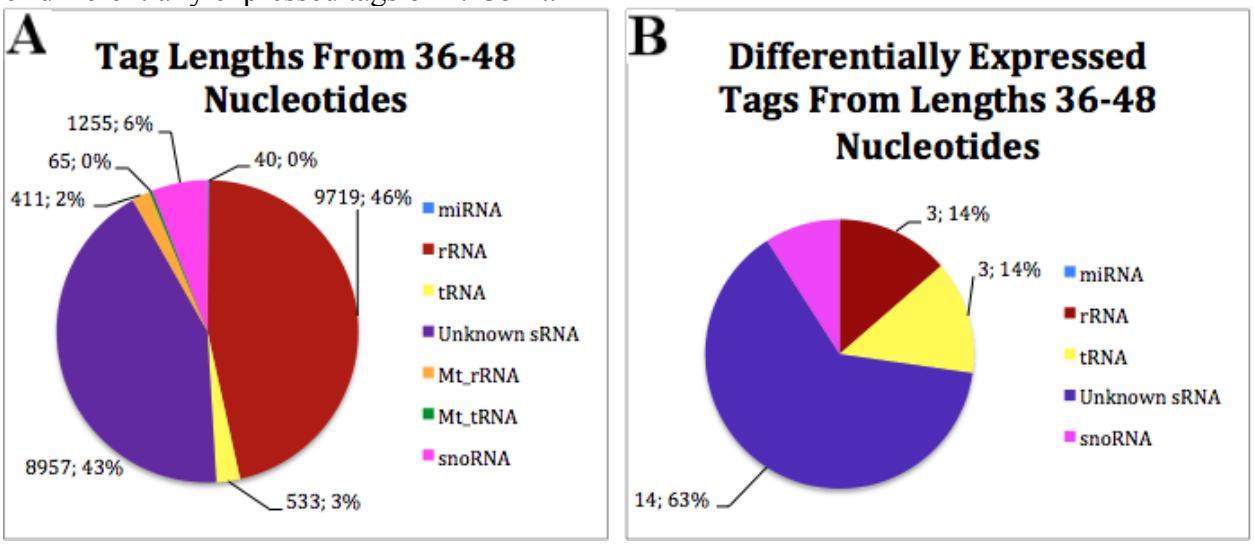


**Figure 6: Diversity of Tags Between 36-48 Nucleotides in Length Expressed During Caudal Fin Amputation Grouped by sRNA Functional Class.** 6A) Type and relative abundance of all tags 36-48 nt; 6B) Type and relative abundance of differentially expressed tags of 36-48 nt.

**Table 2: Ten Most Highly Expressed Transfer RNA Derived Fragments of Any Length.** LogCPM stands for Counts per Million and is the average expression of a particular tag in these data.

| Tag # | Sequence | Symbol | LogCPM |
|---|---|---|---|
| Tag_60878 | GCCCGGCTAGCTCAGTCGGTAGAGCATGAGAC | Danio_rerio_chr8.trna412-LysCTT | 11.79621185 |
| Tag_15602 | GCATTGGTGGTTCAGTGGTAGAATTCTCGC | Danio_rerio_chr8.trna769-GlyGCC | 10.86223624 |
| Tag_91047 | GCCCGGCTAGCTCAGTCGGTAGAGCATGAGA | Danio_rerio_chr8.trna412-LysCTT* | 9.854182714 |
| Tag_5774 | GTTTCCGTAGTGTAGTGGTTATCACGTTCGCC | Danio_rerio_chr8.trna911-ValCAC | 9.727895996 |
| Tag_88855 | GCATTGGTGGTTCAGTGGTAGAATTCTCGCT | Danio_rerio_chr4.trna5166-GlyCCC | 9.602406884 |
| Tag_17054 | GCATTGGTGGTTCAGTGGTAGAATTCTCGCCTA | Danio_rerio_chr4.trna8052-GlyACC | 8.835590785 |
| Tag_15605 | GCATTGGTGGTTCAGTGGTAGAATTCTCGT | Danio_rerio_chr4.trna8145-GlyCCC | 8.80151904 |
| Tag_45017 | GCCCGGCTAGCTCAGTCGGTAGAGCATGAGACC | Danio_rerio_chr8.trna412-LysCTT | 8.377176833 |
| Tag_27194 | GCCTGGCTAGCTCAGTCGGTAGAGCATGAGACTC | Danio_rerio_chr8.trna558-LysCTT | 8.281685774 |
| Tag_32534 | TCCCTGGTGGTCTAGTGGTTAGGATTCGGCA | Danio_rerio_chr4.trna5066-GluCTC | 8.262087993 |

It should be noted that the pie chart for the tags that were less than 19 nt in length was not included in this report because all of the tags in this range were called as unknown sRNA. An alignment score of at least 18 was required for a tag to be aligned to a sRNA in one of the databases. The tags less than 19 nt did not meet this requirement and so they were labeled as "unknown sRNA". These tags were still included in the rest of these data and results. The pie chart displaying the sRNA types for this range was simply not included as there were no sRNA types to show. After aligning the ten most highly expressed unknown sRNA that were differentially expressed to the zebrafish genome speculation could be made on what class of sRNA a particular unknown tag could be (Table 3). The unknown sequences that did not appear to align to anything in the zebrafish genome were given the symbol "No Alignment".

The canonical sequence for miR-21, which is very well researched, functions as a positive control for my data (Fig. 7). An isomiR of miR-21 had the highest expression, meaning it had the largest average expression of the tags that were differentially expressed, in this dataset (Fig. 7). Of the tRNA derived fragments that were differentially expressed, two of the top three most highly expressed tags were isoforms of each other and differ by only one nucleotide at the 3' end (Fig. 7, Table 2).

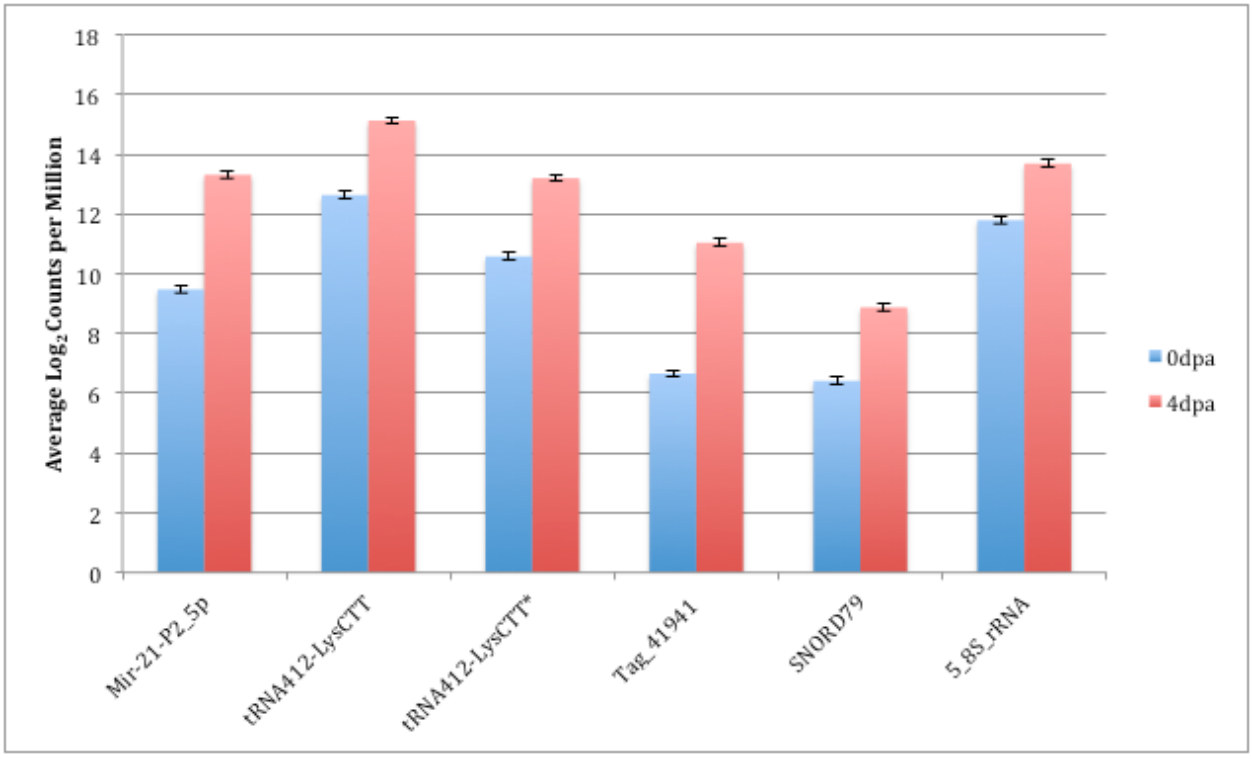


**Figure 7**: **Differential Expression of sRNA From 0 dpa to 4 dpa.** Data shown is in $Log_2$. Standard error bars shown. MiR-21 acts as a positive control for the bioinformatics that was done. The sRNA shown are the most highly expressed tag that aligned to the respective sRNA type and was differentially expressed.

Of the unknown sRNA tags that were differentially expressed Tag_41941 was the most highly expressed in these data (Table 3). After aligning Tag_41941 to the zebrafish genome in NCBI's map viewer and UCSC Genome Browser it appears this is a tRNA derived fragment that aligns just beyond the 3' end of a tRNA sequence.

**Table 3**: **Ten Most Highly Expressed Unknown Small RNA.** Tag_41941 and Tag_18259 are isoforms of each other and are suspected to be tRNA derived fragments that arise from a novel location following the 3' end of a tRNA transcript. LogCPM stands for "counts per million" and represents the average expression of each tag.

| Tag # | Sequence | UCSC Genome Browser/NCBI Map Viewer | LogCPM |
|---|---|---|---|
| Tag_41941 | GGGATGAGTCCCTATCACGGCACAGGAGTGTGACTTT | tRNA-His-CAY | 7.564625881 |
| Tag_83108 | TTACGACCTCAGATCAGACGAGACAACCCGCT | rRNA | 7.43486585 |
| Tag_73835 | AAATGACTCGTAGACGACCTGATTC | rRNA | 7.046799037 |
| Tag_47567 | TTACGACCTCAGATCAGACGAGACA | rRNA | 6.748483564 |
| Tag_18259 | GGGATGAGTCCCTATCACGGCACAGGAGTGTGACTTTT | tRNA-His-CAY | 6.511064918 |
| Tag_22908 | TATCAGACTGGTGTTGGC | miR-21 | 5.702317597 |
| Tag_62814 | ACCGGGCGGAAACACCA | tRNA-Val-CAC/AAC | 5.46477622 |
| Tag_22801 | TAGACGACCTGATTC | No Alignment | 5.283316109 |
| Tag_48310 | CCTGTCTGAGCGTCGCTC | No Alignment | 5.258546622 |
| Tag_62744 | CGTAGACGACCTGATTC | No Alignment | 5.2107215 |

**Table 4: Top Five Most Highly Expressed rRNA Derived Fragments.** LogCPM stands for "counts per million" and represents the average expression of each tag.

| Tag # | Sequence | Symbol | 5' vs. 3' end of rRNA and Sense(+)/Anti-sense(-) | LogCPM |
|---|---|---|---|---|
| Tag_40455 | ACGCCTGTCTGAGCGTCGCTT | 5_8S_rRNA | 3' - | 10.45135088 |
| Tag_37691 | ACGCCTGTCTGAGCGTCGCTC | 5_8S_rRNA | 3' + | 7.399685524 |
| Tag_95762 | CACGCCTGTCTGAGCGTCGCTT | 5_8S_rRNA | 3' - | 7.391079492 |
| Tag_87565 | GAGACCGCCTGGGAATACCAGGTGCTGTAAGCTTT | 5S_rRNA | 3' - | 6.611736591 |
| Tag_82334 | CAACTCTTAGCGGTGGATCACTCGGCTCGTGCGTT | 5_8S_rRNA | 5' - | 5.694029251 |

Of the five most highly expressed tags that mapped to rRNA derived fragments and were differentially expressed, most aligned to the 3' end of the rRNA transcript. However, there was one tag that aligned to the 5' end (Table 4). Additionally, 4 of the 5 most highly

expressed rRNA derived fragments were found on the antisense strand while only one was found on the sense strand.

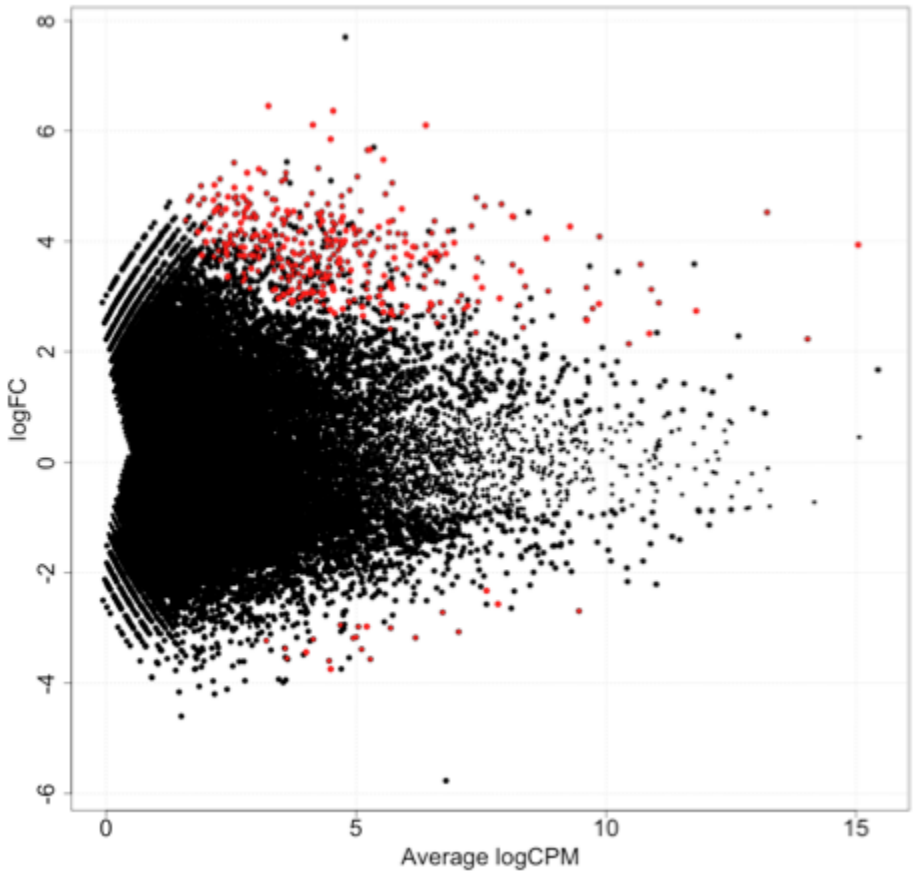


**Figure 8: MA Plot of The Differentially Expression sRNA.** The red dots are tags with an FDR < 0.00001. Black dots are tags with an FDR > 0.00001.

**Table 5: Coordinates for Tags of Particular Interest in Fig. 8.** LogFC (log of the the fold change) is the Y-axis coordinate in Fig. 8 and LogCPM is the X-axis coordinate in Fig. 8. LogCPM stands for "counts per million" and represents the average expression of each tag.

| Tag # | Sequence | Symbol | LogFC | LogCPM |
|---|---|---|---|---|
| Tag_65875 | CCGCAGCTAGGAATAATGGAATAGGAC TCCGGTTCCATTTT | Dre-Mir-133-P2_5p* | 1.026920947 | 1.831664454 |
| Tag_6402 | TAGCTTATCAGACTGGTGTTGGC | Dre-Mir-21-P2_5p | 3.937909731 | 15.04229364 |
| Tag_60878 | GCCCGGCTAGCTCAGTCGGTAGAGCATG AGAC | tRNA412-LysCTT | 2.737231879 | 11.79621185 |
| Tag_15602 | GCATTGGTGGTTCAGTGGTAGAATTCT CGC | tRNA769-GlyGCC | 2.328041739 | 10.86223624 |
| Tag_41941 | GGGATGAGTCCCTATCACGGCACAGGA GTGTGACTTT | Unknown sRNA | 4.637553228 | 7.564625881 |
| Tag_18442 | TGTTTATGATGATTGTTAAATGAAACA ATGGGAATCTCTCTGAATT | SNORD79 | 2.71722535 | 5.547094804 |

**Discussion**

In the 19-26 nt length range 85% of the differentially expressed tags were miRNA (Fig. 4B). Previous studies have shown miRNAs regulate genes involved in the regeneration of the zebrafish caudal fin (Yin VP et al, Thatcher EJ et al.). As a result it would make sense that there would be an abundance of miRNA that are being differentially expressed following the amputation of the caudal fin. If there are other kinds of sRNA, besides these miRNA, that play a regulatory role in the regeneration pathway then it would be expected to see an abundance of such sRNA being differentially expressed.

In the tag length range 27-35 nt 86% of those differentially expressed tags aligned to tRNA derived fragments(Fig. 5B). From previous experiments it has been found that tRNA derived fragments have a regulatory role in various other body systems (Chen Q et

al. and Sharma U et al.). If tRNA derived fragments do indeed have a regulatory mechanism for silencing genes and they appear to be differentially expressed as a result of amputation of the caudal fin then it stands to reason that these tRNA derived fragments may have a regulatory role in the regeneration pathway.

It should again be mentioned that of the 196 differentially expressed miRNA (Fig. 4B), 159 of them were isomiRs of miR-21. MiR-21 is a very well researched miRNA that targets the gene for programmed cell death 4 (PDCD4). The expression of PDCD4 leads to apoptosis. MiR-21 has most notably shown high expression levels in cancerous tissues thus preventing the destruction of cancerous cells and allowing them to continue to divide (Lu Z et al.). The high expression observed here in the regeneration of zebrafish fin tissue can be attributed to preventing proliferating cells from undergoing apoptosis as proliferation and outgrowth of the cells is the primary goal of regeneration (Lu Z et al.).

IsomiRs are sRNA with a variation in the canonical sequence particularly at the 3' end of the miRNA or as single nucleotide alterations in the canonical miRNA transcript (McGahon MK et al). This variation is thought to add an additional level of specificity to the regulation of particular mRNA strands (McGahon MK et al). This abundance of miR-21 isomiRs skews the relative abundance of miRNA found in that particular tag length range. Since there are a large number of isomiRs of tags aligning to miR-21 and all of them are upregulated it gives the appearance that there are lots of miRNA that are being upregulated when really it's only miR-21 isomiRs (Fig. 8). Looking also at the 27-35 nt length range these data show that while there are still some isoforms of these tRNA derived fragments they are not nearly as abundant as miR-21 isomiRs that were in the 19-

26 nt range. Due to time restraints I was not able to examine these various isomiRs more closely. However, it would be interesting to go through and remove all of the sRNA isoforms to see the abundance of each sRNA type when each sRNA is only counted once.

The two tags that were differentially expressed and aligned to snoRNA both mapped to SNORD79 (Fig. 7). While the expression of these tags was not exceptionally high the differential expression of these two sequences suggest that they may play a regulatory role in the regeneration pathway. In order to understand what specific role these snoRNA are playing in fin regeneration it is necessary to first identify their gene targets.

It was not possible to align 18% of the differentially expressed tags to any of the sRNA databases that were used to annotate the tags (Fig. 3). While 18% is not an exceptionally large amount it is enough to warrant some curiosity as to what these unknown sRNA are. The majority of these unknown sRNA were less than 26 nt in length. However, Tag_41941, which was the most highly expressed unknown sRNA that was also differentially expressed, was 37 nt in length (Fig. 7, Table 3). When aligning this tag sequence to the zebrafish genome using UCSC Genome Browser and NCBI's Map Viewer, this sequence appears immediately following the 3' end of a tRNA sequence suggesting that it might be a tRNA derived fragment.

In recent research it was found that two tRNA derived fragments that functioned as miRNA and were transcribed from the 3' UTR end of a tRNA sequence (Pekarsky Y et al.). These tRNA derived fragments are terminated by a series of four thymine nucleotides (Pekarsky Y et al.). Tag_41941, which is the more highly expressed tag

actually only has three thymine nucleotides at its 3' end; however, Tag_18259 has the four thymine nucleotides at its 3' end (Table 3). These two "unknown sRNA" could actually be tRNA derived fragments with miRNA like functions (Pekarsky Y et al.).

To discover the actual class that these "unknown sRNA" belong to more research into how these sequences are interacting during the regeneration process is needed. Since this sequence did not align to the actual tRNA, but rather it began exactly at the 3' end of the tRNA sequence in the zebrafish genome it makes sense that it was not called by the tRNA database that was used to identify the tags. This phenomenon raises the question of what other sRNA could potentially be tRNA derived fragments that are missed because they appear just beyond the 3' end of the tRNA sequence?

It would be interesting to study the function of tRNA412-LysCTT (Tag_60878) during regeneration. Tag_60878 was the most highly expressed tRNA derived fragment that was differentially expressed (Table 2). Furthermore, it's isoform; Tag_91047 was also highly expressed (Fig. 7, Table 2). These two tag sequences only differ by one nucleotide and have largely the same expression pattern except Tag_60878 has a higher expression than Tag_91047. This difference in expression level could be caused by the high specificity required of sRNA to bind to mRNA strands to silence translation (Fig. 7). Both of these tags align to the 5' end of the tRNA transcript, which is considered more of the typical place to find these tRNA derived fragments (Shigematsu and Kirino. 2015).

The rRNA derived fragments were also mapped to both the 5' and the 3' end of the rRNA transcript (Table 4). It appears that these tRNA derived fragments and the rRNA derived fragments can arise from both the 5' and 3' end of their respective RNA

transcripts. Additionally, the tRNA derived fragments can also arise from just beyond the 3' end of the tRNA sequence. As a result of this the question of, "What other sRNA are being transcribed from these "unusual" locations?" is provoked. This is a question that only further research into these sRNA can answer.

**Conclusion**

The results from this research shows that sRNA types, besides miRNA, were differentially expressed following the amputation of the zebrafish caudal fin. Based on differential expression of the tRNA derived fragments, Tag_41941 and tRNA412-LysCTT are perhaps the most promising sRNA sequences to continue to study. This is because Tag_41941 appears immediately on the 3' end of the tRNA sequence in the zebrafish genome and highlights a novel location for these tRNA derived fragments. Furthermore, looking at Fig. 7 it can be seen that Tag_41941 had a greater up regulation than even the tags that aligned to known tRNA derived fragments. The tRNA412-LysCTT has the highest expression of the tRNA derived fragments in these data and it has an isoform that, shows a lower expression but is only a single nucleotide different. These sRNA in conjunction with the differential expression of SNORD79 and particular rRNA derived fragments support the evidence suggesting that these types of sRNA are indeed playing some kind of role during the regeneration process. That said it is also clear that the expression of these sRNA are not at the same level as miR-21. Suggesting that miR-21 plays a more prominent role in the regeneration process than the other sRNA.

Nevertheless, only further research with these other sRNA will reveal their actual role during this process.

The appearance of potential tRNA derived fragments just beyond the 3' end of tRNA transcripts provides us with an additional location to research while searching for potential sRNA. The differential expression of rRNA derived fragments at both the 3' and 5' ends of the rRNA transcript also demonstrates that these sRNA sequences do not necessarily appear in the more traditional locations. The expression levels of these rRNA derived fragments shown in this research suggest that these sRNA require further study into where they are derived from and how they are affecting the regeneration process. This future research will give us insight into how to better control this pathway in other organisms. In future research working with these data it would be interesting to also examine the different sRNA isoforms to see what variation in expression patterns, if any, arise during regeneration. A significant change in the differential expression between two sRNA isoforms could suggest that sequence specificity is a critical aspect in the expression of a particular sRNA.

This research only looked at the sRNA that were expressed at 0 dpa and at 4 dpa. It would be intriguing to determine and analyze the expression levels of these sRNA at more time intervals up to complete regeneration of the fin. These data would reveal a more complete understanding of the expression pattern of these sRNA during the regeneration process. The differential expression patterns of the sRNA in this analysis are consistent with tRNA derived fragments and rRNA derived fragments having a regulatory role during the regeneration of the zebrafish caudal fin. The differential

expression pattern of snoRNA might suggest that this type has a regulatory role as well, however; more data is needed to make this claim with greater confidence. Additionally, this research supports the notion that the 3' UTR of tRNA transcripts is a location of sRNA with potential gene regulatory functions. Understanding the variety of ways in which regeneration is regulated is essential in facilitating our mastery of the process and eventual application to human patients.

**References**:

Bartel, D. P. (2004). MicroRNAs: genomics, biogenesis, mechanism, and function. *Cell, 116*(2), 281-297.

Carthew, R. W., and Sontheimer, E. J. (2009). Origins and Mechanisms of miRNAs and siRNAs. *Cell, 136*(4), 642-655. doi:10.1016/j.cell.2009.01.035

Chan, P. P., and Lowe, T. M. (2009). GtRNAdb: a database of transfer RNA genes detected in genomic sequence. *Nucleic Acids Res, 37*(Database issue), D93-97. doi:10.1093/nar/gkn787

Chen, C. H., Merriman, A. F., Savage, J., Willer, J., Wahlig, T., Katsanis, N., Yin, V. P., and Poss, K. D. (2015). Transient laminin beta 1a Induction Defines the Wound Epidermis during Zebrafish Fin Regeneration. *PLoS Genet, 11*(8), e1005437. doi:10.1371/journal.pgen.1005437

Chen, Q., Yan, M., Cao, Z., Li, X., Zhang, Y., Shi, J., Feng, G. H., Peng, H., Zhang, X., Qian, J., Duan, E., Zhai, Q., and Zhou, Q. (2016). Sperm tsRNAs contribute to intergenerational inheritance of an acquired metabolic disorder. *Science, 351*(6271), 397-400. doi:10.1126/science.aad7977

Ender, C., Krek, A., Friedlander, M. R., Beitzinger, M., Weinmann, L., Chen, W., Pfeffer, S., Rajewsky, N., and Meister, G. (2008). A human snoRNA with microRNA-like functions. *Mol Cell, 32*(4), 519-528. doi:10.1016/j.molcel.2008.10.017

Ender, C., and Meister, G. (2010). Argonaute proteins at a glance. *J Cell Sci, 123*(Pt 11), 1819-1823. doi:10.1242/jcs.055210

Eulalio, A., Huntzinger, E., and Izaurralde, E. (2008). Getting to the root of miRNA-mediated

gene silencing. *Cell, 132*(1), 9-14. doi:10.1016/j.cell.2007.12.024

Fromm, B., Billipp, T., Peck, L. E., Johansen, M., Tarver, J. E., King, B. L., Newcomb, J. M., Sempere, L. F., Flatmark, K., Hovig, E., and Peterson, K. J.. A Uniform System for the Annotation of Vertebrate microRNA Genes and the Evolution of the Human microRNAome. Annu Rev Genet. 2015;49:213-42. doi: 10.1146/annurev-genet-120213-092023. Epub 2015 Oct 14. PubMed PMID: 26473382; PubMed Central PMCID: PMC4743252.

Geer, L.Y., Marchler-Bauer, A., Geer, R.C., Han, L., He, J., He, S., Liu, C., Shi, W., and Bryant, S.H.. The NCBI BioSystems database. *Nucleic Acids Res.* 2010 Jan; 38(Database issue):D492-6. *(Epub 2009 Oct 23)* [PubMed PMID: 19854944]

He, L., and Hannon, G. J. (2004). MicroRNAs: small RNAs with a big role in gene regulation. *Nat Rev Genet, 5*(7), 522-531.

Jazwinska, A., and Sallin, P. (2016). Regeneration versus scarring in vertebrate appendages and heart. *J Pathol, 238*(2), 233-246. doi:10.1002/path.4644

Kent, W.J.. BLAT - the BLAST-like alignment tool. *Genome Res.* 2002 Apr;12(4):656-64.Release of zebrafish data: Jul. 2010 (Zv9/danRer7)

King, B.L., and Yin, V.P.. A Conserved MicroRNA Regulatory Circuit is Differentially Controlled during Limb/Appendage Regeneration. PLoS ONE (*in press*).

Kinsella, R. J., Kahari, A., Haider, S., Zamora, J., Proctor, G., Spudich, G., Almeida-King, J., Staines, D., Derwent, P., Kerhornou, A., Kersey, P., and Flicek, P. (2011). Ensembl BioMarts: a hub for data retrieval across taxonomic space. *Database (Oxford), 2011*, bar030. doi:10.1093/database/bar030

Kiriakidou, M., Tan, G. S., Lamprinaki, S., De Planell-Saguer, M., Nelson, P. T., and Mourelatos, Z. (2007). An mRNA m7G cap binding-like motif within human Ago2 represses translation. *Cell, 129*(6), 1141-1151. doi:10.1016/j.cell.2007.05.016

Kozomara, A., and Griffiths-Jones, S. (2014). miRBase: annotating high confidence microRNAs using deep sequencing data. *Nucleic Acids Res, 42*(Database issue), D68-73. doi:10.1093/nar/gkt1181

Lu, Z., Liu, M., Stribinskis, V., Klinge, C. M., Ramos, K. S., Colburn, N. H., and Li, Y. (2008). MicroRNA-21 promotes cell transformation by targeting the programmed cell death 4 gene. *Oncogene, 27*(31), 4373-4379. doi:10.1038/onc.2008.72

Mardis, E. R. (2008). The impact of next-generation sequencing technology on genetics. *Trends Genet, 24*(3), 133-141. doi:10.1016/j.tig.2007.12.007

Matera, A.G., Terns, R.M., and Terns, M.P. (2007). Non-coding RNAs: lessons from the small nuclear and small nucleolar RNAs. Nat. Rev. Mol. Cell Biol. 8, 209–220.

McGahon, M. K., Yarham, J. M., Daly, A., Guduric-Fuchs, J., Ferguson, L. J., Simpson, D. A., and Collins, A. (2013). Distinctive profile of IsomiR expression and novel microRNAs in rat heart left ventricle. *PLoS One, 8*(6), e65809. doi:10.1371/journal.pone.0065809

Kersey, P.J., Allen, J. E., Armean, I., Boddu, S., Bolt, B. J., Carvalho-Silva, D., Christensen, M., Davis, P., Falin, L. J., Grabmueller, C., Humphrey, J., Kerhornou, A., Khobova, J., Aranganathan, N. K., Langridge, N., Lowy, E., McDowall, M. D., Maheswari, U., Nuhn, M., Ong, C. K., Overduin, B., Paulini, M., Pedro, H., Perry, E., Spudich, G., Tapanari, E., Walts, B., Williams, G., Tello–Ruiz, M., Stein, J., Wei, S., Ware, D., Bolser, D. M., Howe, K.L., Kulesha,

E., Lawson, D., Maslen, G. and Staines, D. M.. Ensembl Genomes 2016: more genomes, more complexity. *Nucleic Acids Res. 2015* doi: 10.1093/nar/gkv1209

Pekarsky, Y., Balatti, V., Palamarchuk, A., Rizzotto, L., Veneziano, D., Nigita, G., Rassenti, L. Z., Pass, H. I., Kipps, T. J., Liu, C. G., and Croce, C. M. (2016). Dysregulation of a family of short noncoding RNAs, tsRNAs, in human cancer. *Proc Natl Acad Sci U S A, 113*(18), 5071-5076. doi:10.1073/pnas.1604266113

Poleo, G., Brown, C. W., Laforest, L., and Akimenko, M. A. (2001). Cell proliferation and movement during early fin regeneration in zebrafish. *Dev Dyn, 221*(4), 380-390. doi:10.1002/dvdy.1152

Poss, K. D., Keating, M. T., and Nechiporuk, A. (2003). Tales of regeneration in zebrafish. *Dev Dyn, 226*(2), 202-210. doi:10.1002/dvdy.10220

Quast, C., Pruesse, E., Yilmaz, P., Gerken, J., Schweer, T., Yarza, P., Peplies, J., and Glöckner, F. O. (2013) The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic Acids Res. 41 (D1): D590-D596.*

Robinson, M.D., McCarthy, D.J., and Smyth, G.K. (2010). edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics 26, 139-140

Robinson, M.D. and Oshlack, A. (2010). A scaling normalization method for differential expression analysis of RNA-seq data. Genome Biology 11, R25.

Schoorlemmer, J., Perez-Palacios, R., Climent, M., Guallar, D., and Muniesa, P. (2014). Regulation of Mouse Retroelement MuERV-L/MERVL Expression by REX1 and

Epigenetic Control of Stem Cell Potency. *Front Oncol, 4*, 14. doi:10.3389/fonc.2014.00014

Sharma, U., Conine, C. C., Shea, J. M., Boskovic, A., Derr, A. G., Bing, X. Y., Belleannee, C., Kucukural, A., Serra, R. W., Sun, F., Song, L., Carone, B. R., Ricci, E. P., Li, X. Z., Fauquier, L., Moore, M. J., Sullivan, R., Mello, C. C., Garber, M., and Rando, O. J. (2016). Biogenesis and function of tRNA fragments during sperm maturation and fertilization in mammals. *Science, 351*(6271), 391-396. doi:10.1126/science.aad6780

Shigematsu, M., and Kirino, Y. (2015). tRNA-Derived Short Non-coding RNA as Interacting Partners of Argonaute Proteins. *Gene Regul Syst Bio, 9*, 27-33. doi:10.4137/grsb.s29411

Song, K., Wang, Y., and Sassoon, D. (1992). Expression of Hox-7.1 in myoblasts inhibits terminal differentiation and induces cell transformation. *Nature, 360*(6403), 477-481. doi:10.1038/360477a0

Thatcher, E. J., Paydar, I., Anderson, K. K., and Patton, J. G. (2008). Regulation of zebrafish fin regeneration by microRNAs. *Proc Natl Acad Sci U S A, 105*(47), 18384-18389. doi:10.1073/pnas.0803713105

Thomson, D. W., Pillman, K. A., Anderson, M. L., Lawrence, D. M., Toubia, J., Goodall, G. J., and Bracken, C. P. (2015). Assessing the gene regulatory properties of Argonaute-bound small RNAs of diverse genomic origin. *Nucleic Acids Res, 43*(1), 470-481. doi:10.1093/nar/gku1242

Yin, V. P., Lepilina, A., Smith, A., and Poss, K. D. (2012). Regulation of zebrafish heart regeneration by miR-133. *Dev Biol, 365*(2), 319-327. doi:10.1016/j.ydbio.2012.02.018

Yin, V. P., Thomson, J. M., Thummel, R., Hyde, D. R., Hammond, S. M., and Poss, K. D. (2008). Fgf-dependent depletion of microRNA-133 promotes appendage regeneration in zebrafish. *Genes Dev, 22*(6), 728-733. doi:10.1101/gad.1641808

**Appendix A**
**LEN_Final_Version_0dpa+4dpa.pyw**

```
import re
fasta_desc = re.compile('^>\d+\-(\d+)$')
tags = {}
tags_A = {}
tags_B = {}
tags_C = {}
tags_D = {}
tags_E = {}

for line in open('Fin_0dpa_PD2_1_clipped.fa'):
    line = line.strip()
    match = fasta_desc.match(line)
    if (match):
        freq = match.group(1)
    else:
        tags[line] = freq

for line in open('Fin_0dpa_PD2_2_clipped.fa'):
    line = line.strip()
    match = fasta_desc.match(line)
    if (match):
        freq = match.group(1)
    else:
        tags_A[line] = freq

for line in open('Fin_0dpa_PD2_3_clipped.fa'):
    line = line.strip()
    match = fasta_desc.match(line)
    if (match):
        freq = match.group(1)
    else:
        tags_B[line] = freq

for line in open('Fin_4dpa_PD2_1_clipped.fa'):
    line = line.strip()
    match = fasta_desc.match(line)
    if (match):
        freq = match.group(1)
    else:
        tags_C[line] = freq

for line in open('Fin_4dpa_PD2_2_clipped.fa'):
```

```
    line = line.strip()
    match = fasta_desc.match(line)
    if (match):
        freq = match.group(1)
    else:
        tags_D[line] = freq

for line in open('Fin_4dpa_PD2_3_clipped.fa'):
    line = line.strip()
    match = fasta_desc.match(line)
    if (match):
        freq = match.group(1)
    else:
        tags_E[line] = freq

for key in tags:
    if
((tags.has_key(key))&(tags_A.has_key(key))&(tags_B.has_key(key))&(tags_C.has_key(
key))&(tags_D.has_key(key))&(tags_E.has_key(key))):
        print key, "\t",len(key)
```

**Appendix B**

**Tab_organized_data.pyw**

```
hits={}
symbols={}
alignment={}
align_prct={}
hits_r={}
symbols_r={}
alignment_r={}
align_prct_r={}
hits_t={}
symbols_t={}
alignment_t={}
align_prct_t={}
hits_Mt_rRNA={}
symbols_Mt_rRNA={}
alignment_Mtr={}
align_prct_Mtr={}
hits_Mt_tRNA={}
symbols_Mt_tRNA={}
alignment_Mtt={}
align_prct_Mtt={}
hits_sno={}
symbols_sno={}
alignment_sno={}
align_prct_sno={}
hits_silva_rRNA={}
symbols_silva_rRNA={}
alignment_silva_rRNA={}
align_prct_silva_rRNA={}
for line in open("tags_vs_dre_mature_tab_filtered.txt"):
    line = line.strip()
    fields=line.split("\t")
    tagid=fields[0]
    hits[tagid]=line
    symbols[tagid]=fields[1]
    if int(fields[3]) > 18:
        alignment[tagid]=fields[3]
    else:
        alignment[tagid] = ""
    if int(float(fields[2])) >= 95.000:
```

```
        align_prct[tagid] = fields[2]
    else:
        align_prct[tagid] = ""


for line in open("tags_vs_rRNA_tab_filtered.fa"):
    line.strip()
    fields=line.split("\t")
    tagid=fields[0]
    hits_r[tagid]=line
    symbols_r[tagid]=fields[1]
    if int(fields[3]) > 18:
        alignment_r[tagid]=fields[3]
    else:
        alignment_r[tagid] = ""
    if int(float(fields[2])) >= 95.000:
        align_prct_r[tagid] = fields[2]
    else:
        align_prct_r[tagid] = ""


for line in open("tags_vs_Danio_rerio_tRNA_tab_filtered.fa"):
    line.strip()
    fields=line.split("\t")
    tagid=fields[0]
    hits_t[tagid]=line
    symbols_t[tagid]=fields[1]
    if int(fields[3]) > 18:
        alignment_t[tagid]=fields[3]
    else:
        alignment_t[tagid] = ""
    if int(float(fields[2])) >= 95.000:
        align_prct_t[tagid] = fields[2]
    else:
        align_prct_t[tagid] = ""


for line in open("tags_vs_Mt_rRNA_tab_filtered.fa"):
    line.strip()
    fields=line.split("\t")
    tagid=fields[0]
    hits_Mt_rRNA[tagid]=line
    symbols_Mt_rRNA[tagid]=fields[1]
    if int(fields[3]) > 18:
```

```
        alignment_Mtr[tagid]=fields[3]
    else:
        alignment_Mtr[tagid] = ""
    if int(float(fields[2])) >= 95.000:
        align_prct_Mtr[tagid] = fields[2]
    else:
        align_prct_Mtr[tagid] = ""


for line in open("tags_vs_Mt_tRNA_tab_filtered.fa"):
    line.strip()
    fields=line.split("\t")
    tagid=fields[0]
    hits_Mt_tRNA[tagid]=line
    symbols_Mt_tRNA[tagid]=fields[1]
    if int(fields[3]) > 18:
        alignment_Mtt[tagid]=fields[3]
    else:
        alignment_Mtt[tagid] = ""
    if int(float(fields[2])) >= 95.000:
        align_prct_Mtt[tagid] = fields[2]
    else:
        align_prct_Mtt[tagid] = ""

for line in open("tags_vs_snoRNA_tab_filtered.fa"):
    line.strip()
    fields=line.split("\t")
    tagid=fields[0]
    hits_sno[tagid]=line
    symbols_sno[tagid]=fields[1]
    if int(fields[3]) > 18:
        alignment_sno[tagid]=fields[3]
    else:
        alignment_sno[tagid] = ""
    if int(float(fields[2])) >= 95.000:
        align_prct_sno[tagid] = fields[2]
    else:
        align_prct_sno[tagid] = ""


for line in open("tags_vs_silva_rRNA_tab_filtered.txt"):
    line.strip()
    fields=line.split("\t")
    tagid=fields[0]
    hits_silva_rRNA[tagid]=line
```

```
    symbols_silva_rRNA[tagid]=fields[1]
    if int(fields[3]) > 18:
        alignment_silva_rRNA[tagid]=fields[3]
    else:
        alignment_silva_rRNA[tagid] = ""
    if int(float(fields[2])) >= 95.000:
        align_prct_silva_rRNA[tagid] = fields[2]
    else:
        align_prct_silva_rRNA[tagid] = ""


for line in open('Tag_num_Final_0dpa+4dpa.fa'):
    line = line.strip()
    fields=line.split("\t")
    tagid=fields[0]
    sequence = fields[1]
    seq_num = len(sequence)-1
    counts =
fields[2]+"\t"+fields[3]+"\t"+fields[4]+"\t"+fields[5]+"\t"+fields[6]+"\t"+fields[7]

    if align_prct.has_key(tagid):
        align_prcts=align_prct[tagid]

    if align_prct_r.has_key(tagid):
        align_prcts_r=align_prct_r[tagid]

    if align_prct_t.has_key(tagid):
        align_prcts_t=align_prct_t[tagid]

    if align_prct_Mtr.has_key(tagid):
        align_prcts_Mtr=align_prct_Mtr[tagid]

    if align_prct_Mtt.has_key(tagid):
        align_prcts_Mtt=align_prct_Mtt[tagid]

    if align_prct_sno.has_key(tagid):
        align_prcts_sno=align_prct_sno[tagid]

    if align_prct_silva_rRNA.has_key(tagid):
        align_prcts_silva_rRNA=align_prct_silva_rRNA[tagid]


    if alignment.has_key(tagid):
        alignments = alignment[tagid]
```

```
if alignment_r.has_key(tagid):
   alignments_r=alignment_r[tagid]

if alignment_t.has_key(tagid):
   alignments_t=alignment_t[tagid]

if alignment_Mtr.has_key(tagid):
   alignments_Mtr=alignment_Mtr[tagid]

if alignment_Mtt.has_key(tagid):
   alignments_Mtt=alignment_Mtt[tagid]

if alignment_sno.has_key(tagid):
   alignments_sno=alignment_sno[tagid]

if alignment_silva_rRNA.has_key(tagid):
   alignments_silva_rRNA=alignment_silva_rRNA[tagid]


if symbols.has_key(tagid):
   symbol=symbols[tagid]
else:
   symbol = ""

if symbols_r.has_key(tagid):
   symbol_r=symbols_r[tagid]
else:
   symbol_r=""

if symbols_t.has_key(tagid):
   symbol_t=symbols_t[tagid]
else:
   symbol_t=""

if symbols_Mt_rRNA.has_key(tagid):
   symbol_Mt_rRNA=symbols_Mt_rRNA[tagid]
else:
   symbol_Mt_rRNA=""

if symbols_Mt_tRNA.has_key(tagid):
   symbol_Mt_tRNA=symbols_Mt_tRNA[tagid]
else:
   symbol_Mt_tRNA=""

if symbols_sno.has_key(tagid):
```

```
        symbol_sno=symbols_sno[tagid]

    else:
        symbol_sno=""

    if symbols_silva_rRNA.has_key(tagid):
        symbol_silva_rRNA=symbols_silva_rRNA[tagid]

    else:
        symbol_silva_rRNA=""

    if symbol != "" and alignment[tagid] != "" and align_prct[tagid] != "":
        print tagid,"\t",sequence,"\t",seq_num,"\t",symbol,"\t", align_prcts,"\t",
alignments,"\t",counts, "\t", "Micro_RNA"

    elif symbol_t != "" and alignment_t[tagid] != "" and align_prct_t[tagid] != "":
        print tagid,"\t",sequence,"\t",seq_num,"\t",symbol_t,"\t",align_prcts_t,"\t",
alignments_t,"\t",counts, "\t", "t_RNA"

    elif symbol_sno != "" and alignment_sno[tagid] != "" and align_prct_sno[tagid] != "":
        print tagid,"\t",sequence,"\t",seq_num,"\t",symbol_sno,"\t",align_prcts_sno,"\t",
alignments_sno,"\t",counts, "\t", "sno_RNA"

    elif symbol_Mt_tRNA != "" and alignment_Mtt[tagid] != "" and align_prct_Mtt[tagid]
!= "":
        print
tagid,"\t",sequence,"\t",seq_num,"\t",symbol_Mt_tRNA,"\t",align_prcts_Mtt,"\t",
alignments_Mtt,"\t",counts, "\t", "Mt_tRNA"

    elif symbol_Mt_rRNA != "" and alignment_Mtr[tagid] != "" and align_prct_Mtr[tagid]
!= "":
        print
tagid,"\t",sequence,"\t",seq_num,"\t",symbol_Mt_rRNA,"\t",align_prcts_Mtr,"\t",alignme
nts_Mtr,"\t",counts, "\t", "Mt_rRNA"

    elif symbol_r != "" and alignment_r[tagid] != "" and align_prct_r[tagid] != "":
        print
tagid,"\t",sequence,"\t",seq_num,"\t",symbol_r,"\t",align_prcts_r,"\t",alignments_r,"\t",co
unts, "\t", "r_RNA"

    elif symbol_silva_rRNA != "" and alignment_silva_rRNA[tagid] != "" and
align_prct_silva_rRNA[tagid] != "":
        print
tagid,"\t",sequence,"\t",seq_num,"\t",symbol_silva_rRNA,"\t",align_prcts_silva_rRNA,"\
t",alignments_silva_rRNA,"\t",counts, "\t", "r_RNA"
```

```
else:
    print tagid,"\t",sequence,"\t",seq_num,"\t",'Unknown sRNA',"\t",counts
```

Author's Biography

Jefferson Adams was born in Bangor, Maine on March 3, 1994. He was raised in Hampden, Maine and graduated from Hampden Academy in 2012. He is majoring in Biology with a concentration Pre-Medical Studies. He is an active member of Black Bear Catholic, the Catholic student group on campus, and is also a Tour Supervisor for Team Maine in the office of Admissions.

Upon graduation, Jefferson will be taking a year off to work and study for the GRE before applying to graduate schools for a degree in Computational Biology and Bioinformatics.