5-2013

A Comparative Genomics Approach to Using High-Throughput Gene Expression Data to Study Limb Regeneration in Ambystoma Mexicanum and Danio Rerio: Developing a More Completely Annotated Database

Justin Bolinger
University of Maine - Main

A COMPARATIVE GENOMICS APPROACH TO USING HIGH-THROUGHPUT GENE EXPRESSION DATA TO STUDY LIMB REGENERATION IN *AMBYSTOMA MEXICANUM* AND *DANIO RERIO*: DEVELOPING A MORE COMPLETELY ANNOTATED DATABASE

by

Justin Bolinger

A Thesis Submitted in Partial Fulfillment
of the Requirements for a Degree with Honors
(Chemical Engineering)

The Honors College

University of Maine at Orono

May 2013

Advisory Committee:
Keith Hutchison, Department of Biochemistry and Molecular Biology, Advisor
Benjamin King, Staff Scientist, Mount Desert Island Biological Laboratory, co-Advisor
John Hwalek, Department of Chemical Engineering
François Amar, Department of Chemistry
Kevin Roberge, Department of Physics and Astronomy

ABSTRACT

Axolotl (*Ambystoma mexicanum*) and the zebrafish (*Danio rerio*) represent organisms extensively studied because of their remarkable capability of fully regenerating completely functional tissues after a traumatic event takes place. However, the research conducted with regards to the genomics of these two organisms has remained fairly independent of each other. The intent of this study is to bridge this gap and connect genes between axolotl and the zebrafish by use a "Rosetta stone" framework to develop a database comparing gene expression data obtained from both microarray-based experiments and high-throughput DNA sequencing of axolotl and zebrafish mRNA and miRNA. Using gene data of this variety, accessed from a variety of private and public resources, 78 axolotl genes were matched to human genes and found to have homologous zebrafish genes. The function of these genes were organized and discussed from a variety of perspectives, including general gene ontologies, specific mechanisms and functions, expression during regeneration at specific times post amputation, and expression in normal regenerating specimens as compared to specimens exposed to the toxin TCDD. Specific proteins and protein functionalities that appear most frequently or novel significance included ribosomal proteins and mitochondrial processes, neurite regeneration, the presences of proteins NADH ubiquinone oxidoreductase, histone, cystatin, and cathepsin, cell differentiation, apoptosis and cellular maintenance, and the structure of the extracellular matrix.

KEY ABBREVIATIONS

Abbrv.	Definition
BioPr	biological process (gene ontology)
BLAST	Basic Local Alignment Search Tool
CComp	cellular component (gene ontology)
cDNA	complementary DNA
dpa	days post-amputation
DRERI	*Danio rerio* (ortholog organism)
ECM	extracellular matrix
GO	Gene Ontology (annotation nomenclature) (*example*: GO:0000000)
HSAPI	*Homo sapiens* (ortholog organism)
HGNC	HUGO Gene Nomenclature Committee (annotation nomenclature) (*example*: ABCD1_HUMAN)
miRNA	microRNA
MolFn	molecular function (gene ontology)
mRNA	messenger RNA
PCR	polymerase chain reaction
TCDD	2,3,7,8-Tetrachlorodibenzo-*p*-dioxin
UP	UniProt, a protein information catalog
UniProt ID	UniProt/SwissProt Accession (annotation nomenclature) ID (*example*: A1B2C3)

TABLE OF CONTENTS

FIGURES

Figure **Page**

TABLES

1. Introduction

The complete assembly and annotation of the human genome that resulted from the Human Genome Project constituted one of the greatest achievements in the relatively short history of bioinformatics. The ability to qualify and quantify gene expression, and to do so rapidly and accurately, has enabled new ways of assessing the role that specific genes play in human health and disease.[1] One such function, the process of tissue repair and regeneration, is of particular importance to this project. Although humans can regenerate skin and blood cells as well as liver tissues, our ability to regenerate other functional tissues is quite minimal.[2] A study comparing gene expression in regenerating tissues between two model organisms capable of regenerating fully functional tissue will lead to the identification of candidate genes that may be critical to this process.

Ambystoma mexicanum (axolotl) and *Danio rerio* (zebrafish) are well developed model organisms for the study of limb regeneration, both being able to efficiently regenerate complex and fully functioning musculoskeletal systems.[3,4] The development of a database comparing gene expression data obtained from both microarray-based experiments and high-throughput DNA sequencing of axolotl and zebrafish mRNA and miRNA would help researchers identify potential candidate genes that have a role in the process of limb regeneration.[1] This information could be compared with and contrasted against data from the human genome in order to better understand the mechanisms and

various metabolic and development pathways that are at work when vital organs and other aspects of the human musculoskeletal system begin to fail or break down.

Although both organisms have been studied extensively on an individual scale, research to draw connections between and connect homologous gene sequences from related sets of axolotl and zebrafish gene expression data has not been carried out to the same degree. Therefore, this study will focus on developing a more annotated database of gene expression data between axolotl and the zebrafish.

The goal of this study is to analyze public and private datasets of gene expression data from studies regarding regeneration in both the axolotl and the zebrafish to compare the expression of homologous genes that exist in these organisms. Analysis of the data will look to provide the user with a "Rosetta stone" of genetic information that will allow for easier comparisons of axolotl and zebrafish gene expression data to be made. Ideally, these annotations will provide some insight as to what types of gene sequences, biological processes, cellular components, and molecular functions may prove to play common roles in the unique regenerative processes of these two organisms. One group of genes that was examined in detail was those genes associated with the dynamic maintenance and biological processes of the extracellular matrix (ECM).

2. Background

Tissue repair and regeneration has the goal of improving the quality of life of people with chronic diseases, infections, or conditions where tissue damage is a major factor. As the average life expectancy of humans increase, a greater number of deaths are likely to be caused by organ failure or the natural breakdown of tissues as opposed to "curable" diseases. The development of new therapies that can reprogram cells to dedifferentiate and then regenerate tissues is the long-term goal of regeneration research. An ever-aging population is far from the only reason for a desire to further the field, as this research could also be used to further human medicine in regards to the treatment of spinal cord injuries and the handling of amputations during and after traumatic events.

2.1. *Ambystoma mexicanum*, *Danio rerio*, and Previous Studies

Ambystoma mexicanum, binomial nomenclature for the Mexican salamander, is a species of salamander originally indigenous to central Mexico. Known colloquially as axolotl, this organism exhibits a select few traits of particular interest to scientific research, including the phenomenon of neoteny (whereby sexual maturity of an animal is reached in the larval stage, characterized by a lack of metamorphosis) and the ability to completely regenerate fully functioning tissues, as would be necessary in case of such an event as limb amputation. In part because of these unique qualities and in part because of its current status as a critically endangered species as designated by the International Union for Conservation of Nature, the Mexican salamander has primarily been relegated

to captive habitats and is a common model and test organism for fields dealing with developmental and genetic research. The axolotl does not yet have a fully sequenced genome.[5]

Danio rerio, binomial nomenclature for the zebrafish, is a species of freshwater fish native to the Himalayan region of Asia, although it has also been introduced to the United States. The zebrafish also possesses many distinguishing characteristics of interest to scientists. Zebrafish eggs become translucent almost immediately after fertilization, and the majority of initial major organ development happens within 3 days of, fertilization, which makes the fish a favorable model subject for developmental research. Another feature of the zebrafish, similar to axolotl, is its ability to regenerate fully functioning tissues. In contrast with the axolotl, the zebrafish genome has been completely sequenced.[6]

Both organisms are heavily researched individually in the scientific community, more specifically in the fields of genomics and bioinformatics. With a focus on gene expression and regulation, previous studies analyzing the regenerative processes found in axolotl and zebrafish specimens have found certain genes to be expressed at elevated or depressed levels at time intervals past the removal of an appendage.[7] The existence of these upregulated and downregulated genes corresponds to an increase or decrease, respectively, in gene expressions relative to a reference. Gene expression data is recovered from transcriptome assemblies compiled using next-generation sequencing, a process that allows scientists to quantify and annotate transcripts as well as receive

alternative sequence variations and corrections in identified genes without the individual needing prior knowledge of the genes in question.[8]

2.1.1. Effect of TCDD on Gene Expression during Regeneration

2,3,7,8-tetrachlorodibenzo-*p*-dioxin (TCDD) is a type of dibenzo-*p*-dioxin that is polychlorinated and can dissolve in fats, oils, and lipids. TCDD is a persistent chemical compound that possesses the tendency to bioaccumulate in eco- and animal systems. This is especially true for fish, as TCDD has been shown to impede growth and causes heightened mortality rates in these organisms.[9,10]

It has been proven that exposure to TCDD significant affects fin regeneration of the zebrafish. Zodrow *et al.* performed experiments during which zebrafish caudal fins were partially removed and specimens received 2.8, 14, or 70 ng TCDD per g weight via intraperitoneal injections. After 7 days post-amputation (dpa), zebrafish specimens receiving the greatest dosages of TCDD recorded 15% fin regeneration as opposed to 65% fin regeneration as found in untreated specimens. The same study also showed lower rates of cell proliferation for TCDD-exposed zebrafish than for their untreated counterparts.[11]

An agonist of the aryl hydrocarbon receptor (AHR) signaling pathway, a biological process that controls the response of zebrafish to a toxin, TCDD binds with a receptor of AHR and causes translocation of the receptor complex into the nucleus of the

cell, where the complex binds to a gene that transactivates response-specific gene elements. It is through this process that further gene expression is modified.[9,10,12]

2.2. Limb Regeneration – A Brief Cellular Biology Overview

Humans form scar tissue at the wound site rather than regenerate limbs as in axolotl and the zebrafish. In humans, the first stage of wound repair is inflammation that occurs directly after the tissue has been damaged. Inflammation is a fundamental pathological process composed of cytological changes, cellular infiltration, and mediator release. This process occurs in the blood vessels of and the tissues adjacent to those tissues affected by damage.[13] On a cellular level, various components of coagulation process, inflammatory pathways, and the immune system are required to ensure that the organism does not lose too much blood or tissue and does not become infected. A clot of platelets forms and initial plug and is eventually replaced by a matrix of fibrin, a fibrous protein involved with the blood clotting process.[14] The second stage in humans is the formation of new tissue. Keratinocytes migrate over the inner layer of the affected area of skin and angiogenesis occurs. The interaction between myofibroblasts and fibroblasts provides the basis for the extracellular matrix (refer to section 2.3., ECM).[14] The third stage, remodeling, occurs over a longer period of time than the first two stages and is characterized by programmed cell death known as apoptosis and the migration of endothelial cells out of the wound. Tissue at the wound site is mainly composed of collagen and ECM proteins.[14]

For the process of limb regeneration in axolotl and the zebrafish, a different set of stages occur. The first stage is composed of fast wound healing. The ECM breaks down and muscle, skeletal, and connective tissues are broken down from their structures. These freed cells then dedifferentiate and help to form blastema, the second stage of limb regeneration. In this stage, the formation of blastema, clusters of dedifferentiated cells that can act as precursors for the formation of many tissues, enables axolotl and the zebrafish to initiate regeneration of damaged tissue structures. This is because blastema cells bear a morphological resemblance to that of stem-like cells.[13,15] The third stage of regeneration in these organisms is outgrowth, whereby new, functional tissues grow in place of the originally damaged tissue. The mass of blastema cells form buds where the growth occurs, with various levels of bud development (early, medium, and late bud blastema, digital outgrowth) and case-specific factors such as wound epidermis and regenerating nerves characterizing the complete process.[16]

2.3. ECM

The extracellular matrix (ECM), a complex network of macromolecules that supports cells and allow them to move and elongate while maintaining their structure, plays an important role in regeneration. The ECM may also provide cells with sources of structural proteins, specialized proteins, and proteoglycans that can be used by the cell.[17,18] Because of these functions, the ECM is generally a term used to encompass the wide variety of cellular components that can help provide a structural framework for the

cell by surrounding and supporting the cell. The ECM forms an interstitial "glue" that works to hold individual and groups of cells together.[17]

The proteins collagen, elastin, fibrillin, fibronectin, and laminin, all of which can be associated to and appear in connective tissue, are found within the ECM.[18] These proteins possess structural and adhesive roles in the process of cell building. The most common of these proteins, collagen, is one of the dominant constituents of skin and bone. Collagen is a fibrous family of proteins that are secreted by cells in connective tissues systems like muscles, tendons, and bone. The typical collagen molecule possesses a long-chained helical structure that exhibits stiffness and stability, which can help explain its role in these connective tissue systems as these systems contain organs known for their durability and functionality under stress.[18]

In addition to the structural and specialized proteins are the proteoglycans, molecules which form a gel-like, hydrated substrate that contains the fibrous proteins of the cell. These proteoglycans can help control proteins secreted by the cell as well as potentially serve as a primary medium through which intercellular chemical signaling and communication can occur.[17]

2.4. Affymetrix GeneChip DNA Microarrays

Affymetrix GeneChip DNA Microarrays were used to acquire the gene expression data for this study in axolotl and the zebrafish. These devices allow researchers to quickly

characterize the gene expression present in any sample tissue using oligonucleotide probes that are designated to a set of genes for a specific organism. Affymetrix arrays use photolithography to microfabricate the oligonucleotide probes onto a solid surface. The array of oligonucleotides can be thought of as a checkerboard of fragmented DNA strands. Each microarray contains hundreds of thousands of squares, known as features. Features are incredibly small – on the order of 10 microns across, about 20% the width of a human hair. Each feature only contains one unique nucleic acid sequence, a chain of 25 base-pairs known as a probe, but there are millions of identical copies of the same probe in the area occupied by a feature. These chips are constructed in a manner similar to a semiconductor, with highly specialized equipment being able to attach and compound individual nucleotides onto the solid surface of the microarray in order to construct any desired assortment of DNA probes. For scale, a 10K Affymetrix GeneChip DNA Microarray could test, as of a few years ago, for over 10,000 unique probes and contains over 400,000 features; the current numbers are much greater. The entire probe array covers a square less than 2 centimeters across.[19]

The basic principle behind the function of a DNA microarray is a concept known as hybridization, which is the process of using innate attractions between the nucleotide base pairs adenine (A), cytosine (C), guanine (G), and thymine (T) to create complementary sequences. In order to test for the presence of a particular DNA sequence of interest, a feature can be constructed so that it contains the opposite base pairs so as to attract that the desired sequence. For example, if the gene sequence one was interested in

determining the presence of in a sample was |ATTAGCGATC, then a probe with the sequence |TAATCGCTAG would be constructed.[19]

To acquire data from microarrays used in expression studies, RNA is extracted from the target sample and converted to complementary DNA (cDNA). The cDNA strands are chemically fragmented and biotin is attached to each short piece of cDNA. This chemical will later bond with fluorescent molecules and this fluorescence, under a laser, will be used to determine where (which probe) and to what degree (the intensity of expression) the sample RNA strands bind to the complementary oligonucleotide probe on the microarray. The sample cDNA is washed over the microarray for 14 to 16 hours and any sample strands that can find a match to any probe in any feature will attach themselves to the appropriate region of the microarray. The chip is then washed over with a fluorescent material which only attaches to the microarray in those regions with paired matches of sample cDNA and corresponding probe. The amount of expression of a particular gene in a sample of tissue, therefore, can be determined by analyzing which features on the microarray appear brightest.[19]

2.5. Illumina High Throughput Sequencing

Another method of quantifying levels of gene expression used for this study is known as high throughput sequencing. This technique provides genomic technologists with a cheaper and faster method, relative to other techniques, of determining the

sequence and expression level of all RNA molecules in a given sample. High throughput sequencing techniques generally use a polymerase chain reaction in order to amplify cDNA templates. Because millions of these sequencing reactions can be run simultaneously, a high throughput method can yield a full set of cDNA in a relatively short period of time.[20]

Illumina high throughput sequencing was used to obtain transcriptome data for this study. First, cDNA was prepared from RNA samples. Using sonication, acoustic waves fragment the cDNA. Oligonucleotides known as adapters are then attached to the ends of these pieces of cDNA and the strands are fractionated, or broken up into even smaller constituents based upon composition, over a gradient for size. Bands of desired base pair length are selected and these molecules are attached to a slide, also called a flowcell, alongside primers for PCR and amplified with polymerase via bridge amplification. In this process, the primers are attached to the flowcell surface and the sample cDNA is loaded onto the flowcell. The DNA strands act as templates for hybridization to occur to form an elongated copy of the DNA strand on the primer. The original molecule is denatured (a double strand of DNA splits into two strands) and washed away with formamide, leaving behind the elongated copies of DNA on the primers. This elongated strand is then free to bend and form a bridge by hybridizing with another PCR primer. The hybridized primer is extended and the two DNA strands are denatured again, leaving behind a strand of DNA attached to each of the PCR primers. This process can happen continually, resulting in 2^n individual strands of identical DNA

for every *n* number of times the process completely repeats itself.[21,22] Thus, DNA sequences of interest are "amplified".

In order to character the order of base pairs in these amplified sequences, four types of reversible terminator bases are added to the flowcell and any bases not finding a match with an amplified sequence are washed away. Because these added molecules are fluorescently labeled, a camera can record an image of the nucleotides added to the flowcell. The fluorescent material is washed away and the next cycle occurs. Extension of DNA chains occur one nucleotide at a time, with images of every step being captured. This enables the ability to record rapid, sequential information for very large colonies of DNA.[21,22]

2.6. Bioinformatics

The field of bioinformatics merges the fields of genetics, genomics, statistics, and often computer programming into an interdisciplinary field aimed at acquiring, organizing, assessing, and analyzing data from biological systems. Many principles of bioinformatics were used in the development of the research carried out by this study and many of these will be explained in further detail.

2.6.1. Searching for Sequence Similarity

A common goal of a study using bioinformatics is to identify instances where a particular protein sequence or gene of interest possesses a sequence similar to a test or desired sequence; for this study, a protein sequence or gene acquired from a sample organism will be referred to as a contig, and a known protein sequence or gene accessed from a genomic database will be referred to as a target. While multiple methods exist for determining these relationships between combinations of sample data and database data, all follow the basic premise of matching nucleotide base pairs and observing how well and to what extent alignments can be made.

2.6.1.1. Local Sequence Alignment

Structural alignment methods use the shape and three-dimensional conformation of a protein structure to make homologous connections between two or more structures.[23] On a more basic level, local sequence alignment methods search for segments of two or more protein sequences that appear to match well. This type of alignment differs from global alignment in that the latter assumes that the protein sequences of approximately equal length and overlap each other over this length. Local alignment, on the other hand, does not force an entire protein sequence into a match between sequences, instead simply comparing the parts of the sequence that exhibit strong similarity. While global sequence alignments may be more effective for identifying unobvious similarities in an overall

sequence, local sequence alignments may provide more accurate and useful results on the whole. A visualization of a global and local alignment is shown in Figure 1 below.[24]

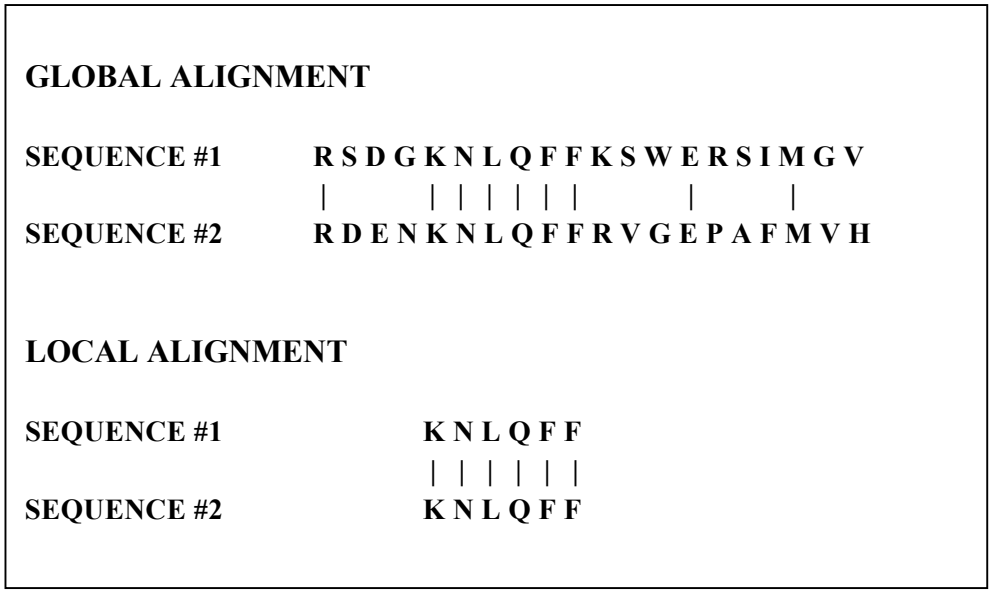

Fig. 1. *Example Illustration of Global and Local Alignment*

Another concept that will be mentioned in the Methods section is the notion of a "percent identity" and a "percent length". These values determine the mapping accuracy (identity) and the extent of coverage between a contig and a target (length) that is to be allowed by the program. Figure 2, below, illustrates an example of how these values can be visualized. The percent identity of this mapping is 95% (19 of the 20 base pair matches are correct) and the percent length of this mapping is 80% (20 of the target's 25 bp sequence is matched up). Moreover, the example in Fig. 2 would be classified as fitting a 95/80 "threshold level".

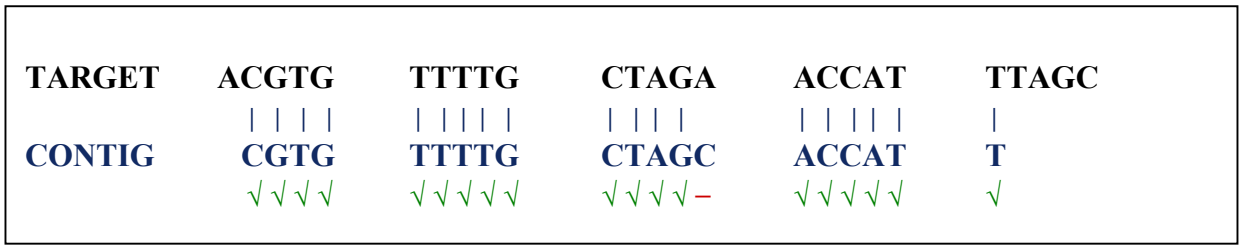

Fig. 2. *Example Illustration of Percent Identity and Percent Length Concepts*

2.6.1.2. BLAST

A popular tool in the bioinformatics community for determining and further characterizing the extent to which local alignments occur between protein sequences is the program Basic Local Alignment Search Tool (BLAST). This program is able to identify regions of local alignment similarity between protein sequences as well as quantify the statistical significance of any observed alignment. BLAST is commonly used to group known gene sequences together based on genetic families and to identify any inter-functional, ontological, evolutionary, or metabolic pathway relationships that may exist between these sequences.[25] BLAST can be used with many types of database files and is a versatile tool for compiling valuable gene mapping information. For the purposes of this study, a BLAST package was downloaded from the internet.[25]

2.6.2. Databases for Protein Sequences and Genes

In order to complete a study of genetic expression data through the lens of comparative genomics, the expression data must either be compiled by the tester or acquired from a secondary source. The focus of this study is data analysis rather than data

acquisition, and so all data files were obtained from outside sources. Many of the files used in this study were downloaded from a private database set up by Benjamin King (MDIBL) on the MDIBL server; this server was used because a few data sets are unpublished and are not subject to public domain. Other files with annotated gene descriptions and other ontological information were downloaded from the Ensembl Genome Browser via BioMart and NCBI via the Gene Expression Omnibus (GEO) DataSet Brower.[26,27]

2.6.3. Programming (Perl)

In order to properly manipulate the data to acquire the results demanded by this study, a programming language must be used in order to parse and sort results from BLAST. Perl is a multipurpose programming language commonly used for, among other applications, handling bioinformatics data. As with typical programming languages, Perl allows users to write scripts, or blocks of carefully formatted texts with a specific syntax, that will can execute commands and display output information in a command prompt window as well as create files of data. The script for this study was written using Perl as the programming language. ActivePerl, an industry-standard distribution of Perl programming language software, was downloaded from ActiveState.[28]

2.6.3.1. Associative Arrays

A key concept in the use of writing a script to manipulate and sort tables of data is the idea of keeping every descriptive value associated with what it's describing. This functionality is carried out through the use of an associative array. These associative arrays allow the user to establish a matrix of values where every individual value in a line of data is automatically "associated" with each other, much as rows in an Excel spreadsheet would be kept together. These input files are tab-delimited so that the script is able to identify this spacing format and associate all corresponding data values to one another, which forms the associative array.

2.6.3.2. Merging Data Sets via Alignment Criteria

As will be discussed in the Methods section, the major purpose of using a programming language with associative arrays to sort through massive amounts of gene name, description, and expression data is to enable the user to quickly make connections between values from different data sets without needing to check every individual gene for a possible match. The main focus of the Perl script written for this study was to use associative arrays and Perl language to merge data sets based on specific criteria. These criteria may include the presence of a common gene name or description, desired statistical significance level, or 1:1 mappings between data sets. For a further explanation of the Perl scripting functionalities used in as well as the overall methodology of the steps taken to acquire data for this study, refer to the Methods section.

3. Methods

The majority of the physical work completed for this study was compiling a script in Perl designed to input axolotl and zebrafish data and output results that align homologous genes between the two organisms, show connections between the organisms, and identify genes that may be of interest. Before a description of the script is discussed, a framing of the methods of this study must be undertaken.

3.1. Framing the Methods from a "Rosetta Stone" Perspective

Figure 3 highlights key concepts for the overall methodology of this study and provides a frame through which all research and data collection can be viewed. A GeneChip microarray is used to map probe sets, or groups of oligonucleotide fragments designed to test for the presence of desired sequences, to the Zv9 (zebrafish) genome for the zebrafish and to axolotl EST (expressed sequence tags) and then to a RNA-Seq assembly for the axolotl. The results of these steps are the production of data files that list all found matches of sample organism gene sequences to known or database gene sequences for those organisms in addition to probe set identification and gene symbol information. These files can be converged and compared to each other as well as to other data sets using Perl and other online resources like NCBI via the Gene Expression Omnibus (GEO) DataSet Brower, UniProt via UniProtKB, and Université de Genève's Department of Microbiology and Molecular Medicine via OrthoDB.[27,29,30]

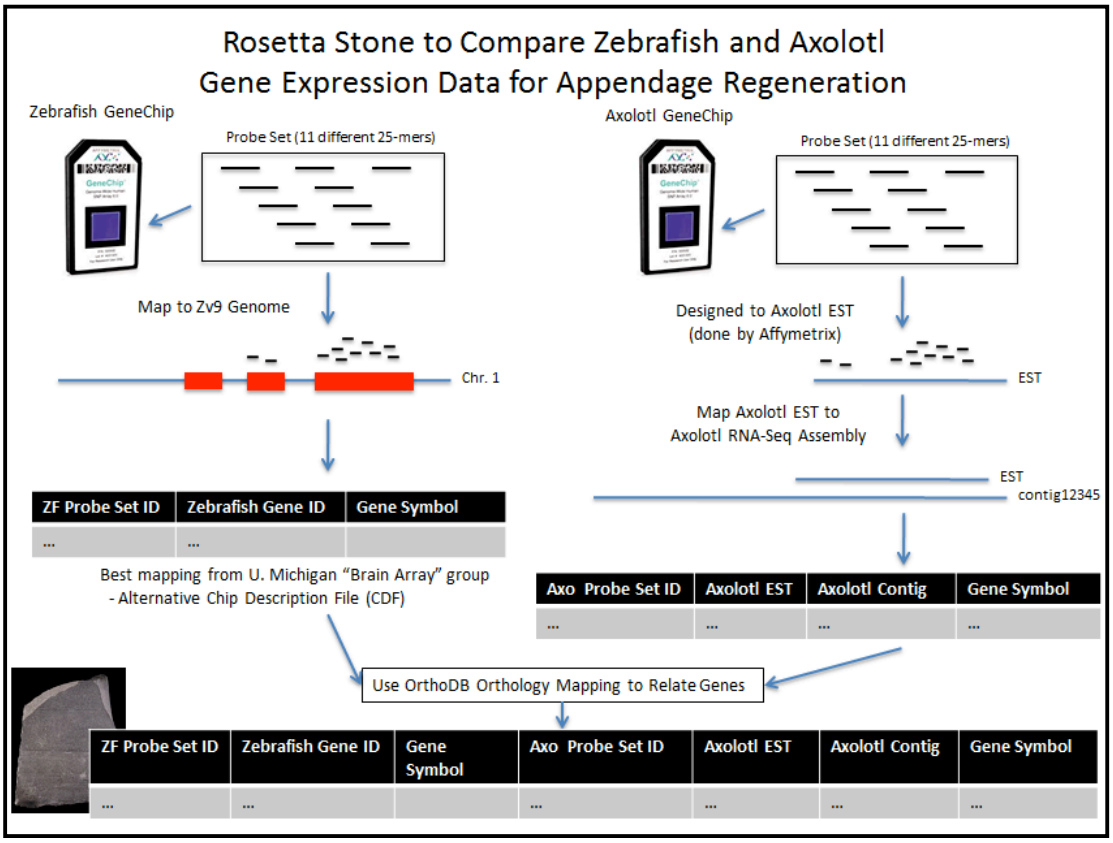

Fig. 3. *Overview of the Rosetta Stone Methodology of this Study*

(source: Benjamin King, MDIBL)

Using a combination of genetic data from outside sources, Perl programming for computational and comparative purposes, and consultation of gene homology resources will the facilitate this study in its aim of creating a "Rosetta stone" of gene ontology for those genes expressed during the process of limb regeneration in axolotl and the zebrafish.

3.2. Perl Script Construction

(See Appendix A for the Perl scripts used in this study.)

First, any FASTA-formatted information file was converted into a tab-delimited text file so that all data files could be properly manipulated by the Perl script. Command prompts were run using Perl programming language to use the BLAST software to run initial comparisons between the axolotl data and the axolotl database genetic data; these data were used to create a file which mapped contigs against targets and a file which mapped the reverse of this case, targets against contigs. This was done so that mappings of genetic sequences could be looked at from two perspectives.

After these initial steps, the primary script used by this study was generated. On-screen prompts for "% IDENTITY" and "% LENGTH" were created, allowing the program user to enter a value between 0 and 100 for each property.

Next, the text files containing axolotl contig lengths from the sample data, axolotl target lengths from the database, and axolotl gene annotations from the database were all read into the Perl script and converted into associative arrays. The input text file for the targets versus contigs comparison was read into the script and converted into an associative array. The script was written to test for incidences of targets mapping to contigs where the threshold level matched those values for percent identity and percent length entered by the user. A counting function was created to assess the number of instances where one target aligned to exactly one, greater than one, and equal to or greater than one contigs. The input text file for the contigs versus targets comparison was

then read into the script and a similar procedure was followed. The data used for these counts was used to give a number of 1:1 mappings for the given threshold level. A 1:1 mapping exists where one target is mapped to only one contig and where that same contig is mapped to only that same target.

Additional programming language was added to the script to read in the text file containing human ECM gene listing data, this also being converted into an associative array. Any 1:1 mappings that share UPSPA names with those found in the human ECM gene listing data were identified.

4. Results and Discussion

Once every input data file was obtained and ran through the completed programs previously described, output data of interest could be acquired. A count of 1:1 mappings between sample contigs from the axolotl data and targets from an axolotl gene database for a variety of alignment identity/length threshold values is easily obtained. Table 1 provides a scope of how much of the total data will be covered by the results of this project.

For example, at a threshold level of 95/30, there are 3900 contigs that align to exactly one target, and 220 contigs that align to two or more targets. At the same threshold level, there are 1462 targets that align to exactly one contig, and 54 targets that align to two or more contigs. (Across all threshold levels, more contig gene sequences

were statistically aligned to target sequences than target sequences were to contig sequences because it is much easier to find a match for a shorter sequence of nucleotides in a larger sequence, as would be the case for the contigs, which are comparatively smaller in length than the targets.) When these matches are compared against each other to identify only those instances where one contig sequence aligns to one target sequence and the same target aligns to the same contig, 1054 such cases were found.

Table 1. *No. of 1:1 Mappings between Sample Contigs and Database Targets, By Alignment Identity and Alignment Length*

		% alignment identity				
		≥ 98	**≥ 95**	**≥ 90**	**≥ 85**	**≥ 80**
% alignment length	**≥ 96**	1	1	1	1	1
	≥ 90	6	8	11	12	14
	≥ 70	114	132	146	154	
	≥ 50	419	461	480		
	≥ 30	997	1054			

Table 1 shows that fewer 1:1 mappings exist as the required accuracies for alignment identity and alignment length are increased. This is to be expected, as a contig that aligns to a target to 86% alignment identity would be counted in the in the "≥ 85" column but not in the "≥ 90" column; the same logic applies for the percent alignment length. This chart of aggregate threshold values can be used to determine threshold values

that ought to be used for various areas of interest for this study, as sometimes selecting a certain threshold value could provide too few or many results to be practically discussed. Threshold values for percent alignment identity are weighted heavier than those values for percent alignment length were because an accurately mapped gene with limited coverage is of more statistical value than a gene with limited mapping accuracy and strong coverage; this is why the increments for percent alignment identity are smaller than the increments for percent alignment length.

In order to determine the threshold values to be used to find axolotl gene data that can be matched to homologous human gene data and zebrafish gene data, the final script is run at a variety of threshold levels in order to find limits not only where these thresholds were still fairly high and but also would provide a good overall number of matches for reporting purposes. A threshold value of 90/50 (90% alignment length accuracy and 50% alignment length accuracy) is selected and the axolotl genes that corresponded to human gene annotation data are given in Table 2 (pages 25-55).

Of the 480 1:1 mappings occurring in the axolotl data set, 78 are matches to homologous human genes. These genes are given in the second column of Table 2 by way of a six digit UniProt ID.[29] Each of these 78 human gene UniProt IDs were individually queried in OrthoDB in order to search for orthologous zebrafish genes. The results of those searches yielded other orthologous human genes as well as orthologous zebrafish genes that related to the queried gene; UniProt IDs and descriptions of each of the 162 homologs found are given. The OrthoDB searches also provided gene ontology

information, which is also included in Table 2 and will be used to further characterize these sets of homologous genes.[30]

Table 2. *Homologous Gene Data for 1:1 Human-Matched Axolotl Data to Zebrafish with Gene Ontologies, at 90/70 Threshold*

	Matched UniProt ID	Ortholog Organism	Ortholog UniProt ID	Ortholog Description	GO Type	GO Descriptions
1	A2AJT9	HSAPI	A2AJT9	Uncharacterized protein CXorf23	[CComp]	GO:0005739 - mitochondrion
2	B4DZF2					
3	C9K0J4					
4	O00483	HSAPI	O00483	NADH dehydrogenase [ubiquinone] 1 alpha subcomplex subunit 4; NDUFA4; NDUA4	[BioPr]	GO:0006810 - transport / GO:0022900 - electron transport chain / GO:0044281 - small molecule metabolic process / GO:0006120 - mitochondrial electron transport, NADH to ubiquinone
		HSAPI	G3V560	NADH dehydrogenase [ubiquinone] 1 alpha subcomplex subunit 4-like 2; NDUFA4L2	[CComp]	GO:0005747 - mitochondrial respiratory chain complex I / GO:0005743 - mitochondrial inner membrane / GO:0070469 - respiratory chain
		DRERI	Q6PBH5	NADH dehydrogenase [ubiquinone] 1 alpha subcomplex subunit 4	[MolFn]	GO:0008137 - NADH dehydrogenase (ubiquinone) activity
		DRERI	E7F8X1	Uncharacterized protein		

Table 2. *Homologous Gene Data for 1:1 Human-Matched Axolotl Data to Zebrafish with Gene Ontologies, at 90/70 Threshold*

	Matched UniProt ID	Ortholog Organism	Ortholog UniProt ID	Ortholog Description	GO Type	GO Descriptions
		DRERI	E7F4U9	Uncharacterized protein; ndufa4l2		
5	O14879	HSAPI	O14879	IFIT-3; IFIT3; IFI60, IFIT4; IFIT3	[BioPr]	GO:0035457 - cellular response to interferon-alpha / GO:0009615 - response to virus / GO:0071360 - cellular response to exogenous dsRNA / GO:0071357 - cellular response to type I interferon / 10 others
		HSAPI	P09913	IFIT-2; IFIT2; G10P2, IFI54; IFIT2	[CComp]	GO:0005829 - cytosol / GO:0005737 - cytoplasm
		HSAPI	Q13325	IFIT-5; IFIT5; RI58; IFIT5		
		HSAPI	P09914	IFIT-1; IFIT1; G10P1, IFI56, IFNAI1; IFIT1		
		HSAPI	Q5T764	Interferon-induced protein with tetratricopeptide repeats 1B; IFIT1B		
		DRERI	E7FBH4	Uncharacterized protein		

Table 2. *Homologous Gene Data for 1:1 Human-Matched Axolotl Data to Zebrafish with Gene Ontologies, at 90/70 Threshold*

	Matched UniProt ID	Ortholog Organism	Ortholog UniProt ID	Ortholog Description	GO Type	GO Descriptions
		DRERI	E7F8D8	Uncharacterized protein		
		DRERI	F6P8G1	Uncharacterized protein; ifit2		
		DRERI	F1QG25	Uncharacterized protein		
		DRERI	B8A535	Novel protein (Zgc:123282); Uncharacterized protein; DKEY-27M7.2		
		DRERI	E7FA13	Uncharacterized protein		
		DRERI	F1QW56	Uncharacterized protein		
		DRERI	E7FCM5	Uncharacterized protein		

Table 2. *Homologous Gene Data for 1:1 Human-Matched Axolotl Data to Zebrafish with Gene Ontologies, at 90/70 Threshold*

	Matched UniProt ID	Ortholog Organism	Ortholog UniProt ID	Ortholog Description	GO Type	GO Descriptions
		DRERI	F1QGN7	Uncharacterized protein		
6	O43181					
7	O43676					
8	O43678	HSAPI	O43678	NADH dehydrogenase [ubiquinone] 1 alpha subcomplex subunit 2; NDUFA2; NDUA2	[BioPr]	GO:0006810 - transport / GO:0022900 - electron transport chain / GO:0044281 - small molecule metabolic process / GO:0006120 - mitochondrial electron transport, NADH to ubiquinone
		DRERI	Q4VBI5	NADH dehydrogenase [ubiquinone] 1 alpha subcomplex subunit 2; ndufa2	[CComp]	GO:0005743 - mitochondrial inner membrane / GO:0070469 - respiratory chain / GO:0005747 - mitochondrial respiratory chain complex I
					[MolFn]	GO:0008137 - NADH dehydrogenase (ubiquinone) activity
9	O75792	HSAPI	O75792	RNase H2 subunit A; RNASEH2A; RNASEHI, RNHIA; RNH2A	[BioPr]	GO:0090305 - nucleic acid phosphodiester bond hydrolysis / GO:0016070 - RNA metabolic process / GO:0006401 - RNA catabolic process / GO:0006260 - DNA replication

Table 2. *Homologous Gene Data for 1:1 Human-Matched Axolotl Data to Zebrafish with Gene Ontologies, at 90/70 Threshold*

	Matched UniProt ID	Ortholog Organism	Ortholog UniProt ID	Ortholog Description	GO Type	GO Descriptions
		DRERI	Q6TNR0	Ribonuclease; rnaseh2a; CH211-145G9.4, RNASEH2A; CH211-145G9.4-001	[CComp]	GO:0032299 - ribonuclease H2 complex / GO:0005634 - nucleus
					[MolFn]	GO:0003723 - RNA binding / GO:0004523 - ribonuclease H activity / GO:0046872 - metal ion binding
10	O94888	HSAPI	O94888	UBX domain-containing protein 7; UBXN7; KIAA0794, UBXD7; UBXN7	[CComp]	GO:0034098 - Cdc48p-Npl4p-Ufd1p AAA ATPase complex
		DRERI	Q6P3G3	Uncharacterized protein; Zgc:92437; ubxn7		
11	O95168	HSAPI	O95168	NADH dehydrogenase [ubiquinone] 1 beta subcomplex subunit 4; NDUFB4; NDUB4	[BioPr]	GO:0006979 - response to oxidative stress / GO:0006810 - transport / GO:0022990 - electron transport chain / GO:0044281 - small molecule metabolic process / 1 other
		DRERI	Q6PBK0	NADH dehydrogenase (Ubiquinone) 1 beta subcomplex, 4; Uncharacterized	[CComp]	GO:0005739 - mitochondrion / GO:0005747 - mitochondrial respiratory chain complex I / GO:0016021 - integral to membrane
					[MolFn]	GO:0008137 - NADH dehydrogenase (ubiquinone) activity

Table 2. *Homologous Gene Data for 1:1 Human-Matched Axolotl Data to Zebrafish with Gene Ontologies, at 90/70 Threshold*

	Matched UniProt ID	Ortholog Organism	Ortholog UniProt ID	Ortholog Description	GO Type	GO Descriptions
12	O95171	HSAPI	O95171	Sciellin; SCEL; SCEL	[BioPr]	GO:0030216 - keratinocyte differentiation / GO:0009790 - embryo development
		DRERI	A8WHU6	Sciellin; Uncharacterized protein; scel; CH211-221B16.2-001	[CComp]	GO:0005886 - plasma membrane / GO:0005737 - cytoplasm / GO:0001533 - cornified envelope
					[MolFn]	GO:0008270 - zinc ion binding
13	P01034	HSAPI	P01034	Cystatin-C; CST3; CYTC	[BioPr]	GO:0060311 - negative regulation of elastin catabolic process / GO:0060313 - negative regulation of blood vessel remodeling / GO:0043206 - fibril organization / GO:0010711 - negative regulation of collagen catabolic process / 34 others
		HSAPI	P01036	Cystatin-S; CST4; CYTS	[CComp]	GO:0005576 - extracellular region / GO:0005615 - extracellular region / GO:0005783 - endoplasmic reticulum / GO:0033267 - axon part / 10 others
		HSAPI	P28325	Cystatin-D; CST5; CYTD	[MolFn]	GO:0004869 - cysteine-type endopeptidase inhibitor activity
		HSAPI	P09228	Cystatin-SA; CST2; CYTT		

Table 2. *Homologous Gene Data for 1:1 Human-Matched Axolotl Data to Zebrafish with Gene Ontologies, at 90/70 Threshold*

	Matched UniProt ID	Ortholog Organism	Ortholog UniProt ID	Ortholog Description	GO Type	GO Descriptions
		HSAPI	P01037	Cystatin-SN; CST1; CYTN		
		DRERI	B8A4D0	Novel protein (Zgc:136227); Uncharacterized protein; DKEY-76K16.3		
14	P04424	HSAPI	P04424	ASAL; ASL; ARLY	[BioPr]	GO:0042450 - arginine biosynthetic process via ornithine / GO:0007626 - locomotory behavior / GO:0009791 - post-embryonic development / GO:0019676 - ammonia assimilation cycle / 4 others
		DRERI	E9QEZ3	Uncharacterized protein; asl	[CComp]	GO:0005829 - cytosol
					[MolFn]	GO:0004056 - argininosuccinate lyase activity
15	P05388	HSAPI	P05388	60S acidic ribosomal protein P0; RPLP0; RLA0	[BioPr]	GO:0042254 - ribosome biogenesis / GO:0006414 - translational elongation / GO:0006413 - translational initiation / GO:0006415 - translational termination / 5 others
		DRERI	Q6P5K3	60S acidic ribosomal protein P0; Rplp0 protein; rplp0	[CComp]	GO:0005840 - ribosome / GO:0005634 - nucleus / GO:0022625 - cytosolic large ribosomal subunit / GO:0030529 - ribonucleoprotein complex / 1 other

Table 2. *Homologous Gene Data for 1:1 Human-Matched Axolotl Data to Zebrafish with Gene Ontologies, at 90/70 Threshold*

	Matched UniProt ID	Ortholog Organism	Ortholog UniProt ID	Ortholog Description	GO Type	GO Descriptions
					[MolFn]	GO:0003735 structural constituent of ribosome / GO:0003723 - RNA binding
16	P10589	HSAPI	P10588	Nuclear receptor subfamily 2 group F member 6; NR2F6; EAR2, ERBAL2; NR2F6	[BioPr]	GO:0030522 - intracellular receptor mediated signaling pathway / GO:0000122 - negative regulation of transcription from RNA polymerase II promoter / GO:0030900 - forebrain development / GO:0001764 - neuron migration / 23 others
		HSAPI	P10589	COUP-TF1; NR2F1; EAR3, ERBAL3, TFCOUP1; COT1	[CComp]	GO:0005634 - nucleus
		HSAPI	P24468	COUP-TF2; NR2F2; ARP1, TFCOUP2; COT2	[MolFn]	GO:0003707 - steroid hormone receptor activity / GO:0004879 - ligand-activated sequence-specific DNA bind RNA polymerase II transcription factor activity / GO:0043565 - sequence-specific DNA binding / GO:0008270 - zinc ion binding / 5 others
		DRERI	Q6P115	Nuclear receptor subfamily 2, group F, member 6b; Uncharacterized protein		
		DRERI	Q6P117	Novel protein (Zgc:77259); Nuclear receptor subfamily 2, group F		
		DRERI	Q6PH18	Nuclear receptor subfamily 2 group F member 1-B; nr2f1b; nr2f1l		

Table 2. *Homologous Gene Data for 1:1 Human-Matched Axolotl Data to Zebrafish with Gene Ontologies, at 90/70 Threshold*

	Matched UniProt ID	Ortholog Organism	Ortholog UniProt ID	Ortholog Description	GO Type	GO Descriptions
		DRERI	Q06726	Nuclear receptor subfamily 2 group F member 5; nr2f5; svp46; NR2F5		
		DRERI	Q91430	Drosophila seven-up homolog/mammalian ARP-1 homolog		
		DRERI	Q06725	Nuclear receptor subfamily 2 group F member 1-A; nr2f1a; nr2f1; svp44		
17	P16112					
18	P20472	HSAPI	P20472		[CComp]	GO:0005737 - cytoplasm / GO:0030424 - axon
		DRERI	Q9I8V0	Parvalbumin-2; pvalb2; pvalb, pvalb1; PRV2	[MolFn]	GO:0005509 - calcium ion binding
		DRERI	Q804W1	Parvalbumin isoform 4b; Pvalb6 protein; Uncharacterized protein; pvalb6		

Table 2. *Homologous Gene Data for 1:1 Human-Matched Axolotl Data to Zebrafish with Gene Ontologies, at 90/70 Threshold*

	Matched UniProt ID	Ortholog Organism	Ortholog UniProt ID	Ortholog Description	GO Type	GO Descriptions
		DRERI	Q7ZT36	Parvalbumin 3; Parvalbumin isoform 1b; Pvalb3 protein; Uncharacterized protein		
		DRERI	Q6IMW7	Parvalbumin 4; Uncharacterized protein; pvalb4		
		DRERI	Q804W2	Parvalbumin-7; pvalb7; pvalb; PRV7		
		DRERI	Q804W0	Parvalbumin 1; Parvalbumin isoform 1d; Uncharacterized protein; pvalb1		
19	P20674	HSAPI	P20674	Cytochrome c oxidase subunit 5A, mitochondrial; COX5A; COX5A	[BioPr]	GO:0048568 - embryonic organ development / GO:0021522 - spinal cord motor neuron differentiation / GO:0044281 - small molecule metabolic process
		DRERI	Q4VBU7	Cytochrome c oxidase subunit Vaa; Uncharacterized protein; cox5aa	[CComp]	GO:0005743 - mitochondrial inner membrane
		DRERI	F1QJ99	Uncharacterized protein; cox5ab	[MolFn]	GO:0004129 - cytochrome-c oxidase activity / GO:0046872 - metal ion binding / GO:0009055 - electron carrier activity

Table 2. *Homologous Gene Data for 1:1 Human-Matched Axolotl Data to Zebrafish with Gene Ontologies, at 90/70 Threshold*

	Matched UniProt ID	Ortholog Organism	Ortholog UniProt ID	Ortholog Description	GO Type	GO Descriptions
		DRERI	Q4VBU7	Cytochrome c oxidase subunit Vaa; Uncharacterized protein; cox5aa		
20	P23396	HSAPI	P23396	40S ribosomal protein S3; RPS3; OK/SW-cl.26; RS3	[BioPr]	GO:0006412 - translation / GO:0006919 - activation of cysteine-type endopeptidase activity involved in apoptotic process / GO:0032088 - negative regulation of NF-kappaB transcription factor activity / GO:0006974 - response to DNA damage stimulus / 9 others
		DRERI	Q6TLG8	Ribosomal protein S3; Uncharacterized protein; rps3	[CComp]	GO:0015935 - small ribosomal subunit / GO:0022627 - cytosolic small ribosomal subunit / GO:0032587 - ruffle membrane / GO:0005634 - nucleus / 5 others
					[MolFn]	GO:0003735 - structural constituent of ribosome / GO:0003723 - RNA binding / GO:0003729 - mRNA binding / GO:0004519 - endonuclease activity / 3 others
21	P27449	HSAPI	P27449	V-ATPase 16 kDa proteolipid subunit; ATP6V0C; ATP6C, ATP6L, ATPL; VATL	[BioPr]	GO:0015991 - ATP hydrolysis coupled proton transport / GO:0044419 - interspecies interaction between organisms / GO:0006879 - cellular iron ion homeostasis / GO:0007035 - vacuolar acidification / 3 others
		DRERI	Q6P041	Uncharacterized protein; Zgc:77708; Zgc:77708 protein; atp6v0cb	[CComp]	GO:0033179 - proton-transporting V-type ATPase, V0 domain / GO:0016021 - integral to membrane / GO:0005774 - vacuolar membrane / GO:0010008 - endosome membrane / 5 others
		DRERI	F1QSP4	Uncharacterized protein; atp6v0ca	[MolFn]	GO:0015078 - hydrogen ion transmembrane transporter activity / GO:0046933 - hydrogen ion transporting ATP synthase activity, rotational mechanism / GO:0046961 - proton-transporting ATPase activity / GO:0042625 - ATPase activity, coupled to transmembrane movement of ions / 2 others

Table 2. *Homologous Gene Data for 1:1 Human-Matched Axolotl Data to Zebrafish with Gene Ontologies, at 90/70 Threshold*

	Matched UniProt ID	Ortholog Organism	Ortholog UniProt ID	Ortholog Description	GO Type	GO Descriptions
22	P35611					
23	P36578	HSAPI	P36578	60S ribosomal protein L4; RPL4; RPL1; RL4	[BioPr]	GO:0006412 - translation / GO:0006413 - translational initiation / GO:0006415 - translaional termination / GO:0006414 - translational elongation / 3 others
		DRERI	Q7ZW95	Ribosomal protein L4; Uncharacterized protein; rpl4	[CComp]	GO:0005840 - ribosome / GO:0005730 - nucleolus / GO:0022625 - cytosolic large ribosomal subunit / GO:0030529 - ribonucleoprotein complex
					[MolFn]	GO:0003735 - structural constituent of ribosome / GO:0003723 - RNA binding
24	P42766					
25	P43235	HSAPI	P43235	Cathepsin K; CTSK; CTSO, CTSO2; CATK	[BioPr]	GO:0006508 - proteolysis / GO:0097067 - cellular response to thyroid hormone stimulus / GO:0002250 - adaptive immune response / GO:0045453 - bone resorption / 3 others
		HSAPI	P25774	Cathepsin S; CTSS; CATS	[CComp]	GO:0005764 - lysosome / GO:0016020 - membrane / GO:0005615 - extracellular space / GO:0005576 - extracellular region / 2 others

Table 2. *Homologous Gene Data for 1:1 Human-Matched Axolotl Data to Zebrafish with Gene Ontologies, at 90/70 Threshold*

	Matched UniProt ID	Ortholog Organism	Ortholog UniProt ID	Ortholog Description	GO Type	GO Descriptions
		DRERI	A2BF64	Novel protein (Zgc:111950); Uncharacterized protein; ctssb.2	[MolFn]	GO:0008234 - cysteine-type peptidase activity / GO:0004197 - cysteine-type endopeptidase activity / GO:0008233 - peptidase activity
		DRERI	F1Q8A0	Uncharacterized protein; ctsk		
		DRERI	Q502A6	Cathepsin S, b.1; Novel protein (Zgc:112226); Uncharacterized protein		
26	P46783	HSAPI	P46783	40S ribosomal protein S10; RPS10; RS10	[BioPr]	GO:0006415 - translational termination / GO:0019083 - viral transcription / GO:0006413 - translational initiation / GO:0006614 - SRP-dependent cotranslational protein targeting to membrane / 2 others
		DRERI	Q7T1J9	Ribosomal protein S10; Uncharacterized protein; rps10	[CComp]	GO:0022627 - cytosolic small ribosomal subunit / GO:0005730 - nucleolus / GO:0005840 - ribosome
27	P56181	HSAPI	P56181	NADH dehydrogenase [ubiquinone] flavoprotein 3, mitochondrial; NDUFV3; NDUV3	[BioPr]	GO:0044281 - small molecule metabolic process / GO:0006810 - transport / GO:0006120 - mitochondrial electron transport, NADH to ubiquinone
					[CComp]	GO:0005747 - mitochondrial respiratory chain complex I / GO:0005634 - nucleus / GO:0005739 - mitochondrion

Table 2. *Homologous Gene Data for 1:1 Human-Matched Axolotl Data to Zebrafish with Gene Ontologies, at 90/70 Threshold*

	Matched UniProt ID	Ortholog Organism	Ortholog UniProt ID	Ortholog Description	GO Type	GO Descriptions
					[MolFn]	GO:0008137 - NADH dehydrogenase (ubiquinone) activity
28	P56556	HSAPI	P56556	NADH dehydrogenase [ubiquinone] 1 alpha subcomplex subunit 6; NDUFA6	[BioPr]	GO:0006810 - transport / GO:0022900 - electron transport chain / GO:0006979 - response to oxidative stress / GO:0044281 - small molecule metabolic process / 1 other
		DRERI	F1RBK6	NADH dehydrogenase [ubiquinone] 1 alpha subcomplex subunit 6; ndufa6	[CComp]	GO:0005743 - mitochondrial inner membrane / GO:0070469 - respiratory chain / GO:0005747 - mitochondrial respiratory chain complex I / GO:0005739 - mitochondrion
					[MolFn]	GO:0008137 - NADH dehydrogenase (ubiquinone) activity
29	P60174	HSAPI	P60174	TIM; TPI1; TPI; TPIS	[BioPr]	GO:0009790 - embryo development / GO:0019682 - glyceraldehyde-3-phosphate metabolic process / GO:0006098 - pnetose-phosphate shunt / GO:0006094 - gluconeogenesis / 2 others
		DRERI	Q1MTI4	TIM-A; tpi1a; si:dkey-89b17.2; TPISA	[CComp]	GO:0005829 - cytosol / GO:0005625 - soluble fraction / GO:0005634 - nucleus
		DRERI	E9QBF0	Triosephosphate isomerase; tpi1b	[MolFn]	GO:0004807 - triose-phosphate isomerase activity

Table 2. *Homologous Gene Data for 1:1 Human-Matched Axolotl Data to Zebrafish with Gene Ontologies, at 90/70 Threshold*

	Matched UniProt ID	Ortholog Organism	Ortholog UniProt ID	Ortholog Description	GO Type	GO Descriptions
30	P60842	HSAPI	P60842	eIF-4A-I; EIF4A1; DDX2A, EIF4A; IF4A1	[BioPr]	GO:0000289 - nuclear-transcribed mRNA poly(A) tail shortening / GO:0019221 - cytokine-mediated signaling pathway / GO:0006413 - translational initiation / GO:0031100 - organ regeneration
		DRERI	Q7ZU67	Eukaryotic translation initiation factor 4A isoform 1B	[CComp]	GO:0005829 - cytosol / GO:0016281 - eukaryotic translation initiation factor 4F complex
		DRERI	F1R166	Uncharacterized protein; eif4a2	[MolFn]	GO:0005524 - ATP binding / GO:0008026 - ATP-dependent helicase activity / GO:0003676 - nucleic acid binding / GO:0003743 - translation initiation factor activity / 4 others
		DRERI	Q802C9	Eukaryotic translation initiation factor 4A, isoform 1A		
31	P61803	HSAPI	P61803	Oligosaccharyl transferase subunit DAD1	[BioPr]	GO:0006486 - protein glycosylation / GO:0006916 - anti-apoptosis / GO:0001824 - blastocyst development / GO:0006915 - apoptotic process / 4 others
		DRERI	A7E2L0	Dad1 protein; Novel protein similar to vertebrate defender against cell death 1 (DAD1)	[CComp]	GO:0016021 - integral to membrane / GO:0008250 - oligosaccharyltransferase complex
					[MolFn]	GO:0004579 - dolichyl-diphosphooligosaccharide-protein glycotransferase activity / GO:0016740 - transferase activity

Table 2. *Homologous Gene Data for 1:1 Human-Matched Axolotl Data to Zebrafish with Gene Ontologies, at 90/70 Threshold*

	Matched UniProt ID	Ortholog Organism	Ortholog UniProt ID	Ortholog Description	GO Type	GO Descriptions
32	P62072	HSAPI	P62072	Mitochondrial import inner membrane translocase subunit Tim10; TIMM10	[BioPr]	GO:0045039 - protein import into mitochondrial inner membrane / GO:0072321 - chaperone-meditated protein transport / GO:0044267 - cellular protein metabolic process / GO:0007605 - sensory preception of sound / 7 others
		DRERI	Q6DI06	Mitochondrial import inner membrane translocase subunit Tim10; timm10	[CComp]	GO:0042719 - mitochondrial intermebrane space protein transporter complex / GO:0005743 - mitochondrial inner membrane / GO:0005744 - mitochondrial inner membrane presequence translocase complex / GO:0009507 - chloroplast
					[MolFn]	GO:0046872 - metal ion binding / GO:0005215 - transporter activity / GO:0008565 - protein transporter activity / GO:0008270 - zinc ion binding / 2 others
33	P62269	HSAPI	P62269	40S ribosomal protein S18; RPS18; D6S218E; RS18	[BioPr]	GO:0006412 - translation / GO:0042254 - ribosome biogenesis / GO:0006417 - regulation of translation / GO:0051726 - regulation of cell cycle / 7 others
		DRERI	Q8JGS9	40S ribosomal protein S18; rps18; ke3; RS18	[CComp]	GO:0005840 - ribosome / GO:0022627 - cytosolic small ribosomal subunit / GO:0015935 - small ribosomal subunit / GO:0005829 - cytosol
					[MolFn]	GO:0003735 - structural constituent of ribosome / GO:0003723 - RNA binding / GO:0019843 - rRNA binding
34	P62277	HSAPI	P62277	40S ribosomal protein S13; RPS13; RS13	[BioPr]	GO:0006412 - translation / GO:0033119 - negative regulation of RNA splicing / GO:0006415 - translational termination / GO:0019083 - viral transcription / 4 others

Table 2. *Homologous Gene Data for 1:1 Human-Matched Axolotl Data to Zebrafish with Gene Ontologies, at 90/70 Threshold*

	Matched UniProt ID	Ortholog Organism	Ortholog UniProt ID	Ortholog Description	GO Type	GO Descriptions
		DRERI	Q6IMW6	Ribosomal protein S13; Uncharacterized protein; rps13	[CComp]	GO:0005840 - ribosome / GO:0022627 - cytosolic small ribosomal subunit / GO:0015935 - small ribosomal subunit / GO:0005730 - nucleolus
					[MolFn]	GO:0003735 - structural constituent of ribosome / GO:0003729 - mRNA binding
35	P62308	HSAPI	P62308	snRNP-G; SNRPG; PBSCG; RUXG	[BioPr]	GO:0008334 - histone mRNA metabolic process / GO:0000245 - spliceosomal complex assembly / GO:0034660 - ncRNA metabolic process / GO:0000387 - spliceosomal snRNP assmebly / 3 others
		HSAPI	F5H5R7	Uncharacterized protein	[CComp]	GO:0005683 - U7 snRNP / GO:0071013 - catalytic step 2 spliceosome / GO:0005689 - U12-type spliceosomal complex / GO:0005654 - nucleoplasm / 3 others
		DRERI	F1QPY7	Uncharacterized protein; zgc:103688	[MolFn]	GO:0003723 - RNA binding / GO:0003676 - nucleic acid binding
36	P62310	HSAPI	P62310	U6 snRNA-associated Sm-like protein LSm3; LSM3; MDS017; LSM3	[BioPr]	GO:0008380 - RNA splicing / GO:0006397 - mRNA processing / GO:0000398 - mRNA splicing / GO:0043928 - exonucleolytic nuclear-transcribed mRNA catabolic process involved in deadenylation-dependent decay
		DRERI	E7EZE6	Uncharacterized protein; CABZ01028296.1	[CComp]	GO:0071013 - catalytic step 2 spliceosome / GO:0005829 - cytosol

Table 2. *Homologous Gene Data for 1:1 Human-Matched Axolotl Data to Zebrafish with Gene Ontologies, at 90/70 Threshold*

	Matched UniProt ID	Ortholog Organism	Ortholog UniProt ID	Ortholog Description	GO Type	GO Descriptions
					[MolFn]	GO:0003723 - RNA binding
37	P62312	HSAPI	P62306	snRNP-F; SNRPF; PBSCF; RUXF	[BioPr]	GO:0006396 - RNA processing / GO:0008380 - RNA splicing / GO:0006397 - mRNA processing / GO:0006364 - rRNA processing / 6 others
		HSAPI	P62312	U6 snRNA-associated Sm-like protein LSm6; LSM6; LSM6	[CComp]	GO:0005634 - nucleus / GO:0005683 - U7 snRNP / GO:0005689 - U12-type spliceosomal complex / GO:0071013 - catalytic step 2 spliceosome / 6 others
		DRERI	H9GXB1	Uncharacterized protein; snrpf	[MolFn]	GO:0003723 - RNA binding / GO:0003676 - nucleic acid binding
		DRERI	Q6IMW9	LSM6 homolog, U6 small nuclear RNA associated (S. cerevisiae); Uncharacterized protein		
38	Q14061					
39	Q15004	HSAPI	Q15004	PCNA-associated factor; PAF; KIAA0101, NS5ATP9; L5; PAF	[CComp]	GO:0005634 - nucleus / GO:0005739 - mitochondrion

Table 2. *Homologous Gene Data for 1:1 Human-Matched Axolotl Data to Zebrafish with Gene Ontologies, at 90/70 Threshold*

	Matched UniProt ID	Ortholog Organism	Ortholog UniProt ID	Ortholog Description	GO Type	GO Descriptions
40	Q15633					
41	Q16718					
42	Q2M385	HSAPI	Q2M385	Macrophage gene 1 protein; MPEG1; MPEG1	[CComp]	GO:0016021 - integral to membrane
		DRERI	A9C3Q1	Novel protein similar to H.sapiens MPEG1, macrophage expressed gene 1 (MPEG1, zgc:110354)		
		DRERI	B0R0K7	Novel protein similar to MPEG1, macrophage expressed gene 1 (MPEG1, zgc:66409)		
		DRERI	B0R063	Novel protein similar to mouse and rat macrophage expressed gene 1 (Mpeg1)		
43	Q30201					

Table 2. *Homologous Gene Data for 1:1 Human-Matched Axolotl Data to Zebrafish with Gene Ontologies, at 90/70 Threshold*

	Matched UniProt ID	Ortholog Organism	Ortholog UniProt ID	Ortholog Description	GO Type	GO Descriptions
44	Q4U2R6	HSAPI	Q4U2R6	L51mt; MRPL51; MRP64; CDA09, HSPC241; RM51	[BioPr]	GO:0006412 - translation
		DRERI	Q5BJJ8	L51mt; mrpl51; zgc:110255; RM51	[CComp]	GO:0005762 - mitochondrial large ribosomal subunit / GO:0005761 - mitochondrial ribosome
					[MolFn]	GO:0003735 - structural constituent of ribosome
45	Q5U5M0					
46	Q5VWZ2	HSAPI	Q5VWZ2	Lysophospholipase-like protein 1; LYPLAL1; LYPL1	[CComp]	GO:0005737 - cytoplasm
		DRERI	F1QBS9	Uncharacterized protein; zgc:110848	[MolFn]	GO:0016787 - hydrolase activity / GO:0004622 - lysophosphate activity
47	Q6UW78	HSAPI	Q6UW78	UPF0723 protein C11orf83; UNQ655/PRO1286; CK083	[CComp]	GO:0005576 - extracellular region

Table 2. *Homologous Gene Data for 1:1 Human-Matched Axolotl Data to Zebrafish with Gene Ontologies, at 90/70 Threshold*

	Matched UniProt ID	Ortholog Organism	Ortholog UniProt ID	Ortholog Description	GO Type	GO Descriptions
		DRERI	B3DFP2	UPF0723 protein C11orf83 homolog; si:dkey-88p24.9; CK083		
48	Q6UXB3	HSAPI	P55000	SLURP-1; SLURP1; ARS; SLUR1	[BioPr]	GO:0001775 - cell activation / GO:0007155 - cell adhesion
		HSAPI	Q6UXB3	Ly6/PLAUR domain-containing protein 2; LYPD2; LYPDC2; UNQ430/PRO788; LYPD2	[CComp]	GO:0005615 - extracellular space / GO:0005886 - plasma membrane / GO:0031225 - anchored to membrane / GO:0005576 - extracellular region
49	Q71DI3	HSAPI	P68431	Histone H3.1; HIST1H3A; H3FA; H31	[BioPr]	GO:0006334 - nucleosome assembly / GO:0060968 - regulation of gene silencing /GO:0007596 - blood coagulation / GO:0051320 - S phase / 7 others
50		HSAPI	P84243	Histone H3.3; H3F3A; H3.3A, H3F3; PP781; H33	[CComp]	GO:0000786 - nucleosome / GO:0005634 - nucleus / GO:0005654 - nucleoplasm / GO:005576 - extracellular region
		HSAPI	Q16695	H3/t; HIST3H3; H3FT; H31T	[MolFn]	GO:0003677 - DNA binding
		HSAPI	Q71DI3	Histone H3.2; HIST2H3A; H32		

Table 2. *Homologous Gene Data for 1:1 Human-Matched Axolotl Data to Zebrafish with Gene Ontologies, at 90/70 Threshold*

	Matched UniProt ID	Ortholog Organism	Ortholog UniProt ID	Ortholog Description	GO Type	GO Descriptions
		HSAPI	Q5TEC6	Histone H3; HIST2H3PS2; RP5-998N21.6-001		
		DRERI	Q6PI20	Histone H3.3; h3f3a; zgc:56193, zgc:86731; H33		
		DRERI	A8KBJ5	Histone H3; zgc:173552		
		DRERI	G1K2S9	Histone H3; h3f3b.1		
		DRERI	Q4QRF4	Histone H3.2; zgc:113984; H32		
51	Q7L2H7					
52	Q7Z4S6	HSAPI	Q7Z4S6	Kinesin-like protein KIF21A; KIAA1708, KIF2; KI21A	[BioPr]	GO:0007018 - microtubule-based movement

Table 2. *Homologous Gene Data for 1:1 Human-Matched Axolotl Data to Zebrafish with Gene Ontologies, at 90/70 Threshold*

	Matched UniProt ID	Ortholog Organism	Ortholog UniProt ID	Ortholog Description	GO Type	GO Descriptions
		DRERI	F1QWX6	Uncharacterized protein; kif21a	[CComp]	GO:0005874 - microtubule / GO:0005737 - cytoplasm
		DRERI	F8W3W5	Uncharacterized protein	[MolFn]	GO:0003777 - microtubule motor activity / GO:0005524 - ATP binding
53	Q8N6V9					
54	Q8N6Y1	HSAPI	Q8N6Y1	Protocadherin-20; PCDH20; PCDH13; PCD20	[BioPr]	GO:0007156 - homophilic cell adhesion
		DRERI	E7F9B5	Uncharacterized protein; pcdh20	[CComp]	GO:0016021 - integral to membrane / GO:0005886 - plasme membrane
					[MolFn]	GO:0005509 - calcium ion binding / GO:0003723 - RNA binding
55	Q8WVI0					

Table 2. *Homologous Gene Data for 1:1 Human-Matched Axolotl Data to Zebrafish with Gene Ontologies, at 90/70 Threshold*

	Matched UniProt ID	Ortholog Organism	Ortholog UniProt ID	Ortholog Description	GO Type	GO Descriptions
56	Q96BP2					
57	Q96E66					
58	Q96MD7	HSAPI	Q96MD7	Uncharacterized protein C9orf85; CI085		
		DRERI	A3KPQ4	Novel protein (Zgc:101016); Uncharacterized protein; DKEY-24K20.3		
59	Q9BPX1	HSAPI	Q9BPX1	17-beta-HSD 14; HSD17B14; DHRS10, SDR3; UNQ502/PRO474; DHB14	[BioPr]	GO:0055114 - oxidation-reduction process / GO:0006706 - steroid catabolic process
		DRERI	Q6DEH9	Uncharacterized protein; Zgc:100900; hsd17b14	[CComp]	GO:0005813 - centrosome / GO:0005829 - cytosol
					[MolFn]	GO:0000166 - nucleotide binding / GO:0016491 - oxidoreductase activity / GO:0047045 - testosterone 17-beta-dehydrogenase (NADP+) activity / GO:0004303 - estradioil 17-beta-dehydrogenase activity

Table 2. *Homologous Gene Data for 1:1 Human-Matched Axolotl Data to Zebrafish with Gene Ontologies, at 90/70 Threshold*

	Matched UniProt ID	Ortholog Organism	Ortholog UniProt ID	Ortholog Description	GO Type	GO Descriptions
60	Q9BX69	HSAPI	Q9BX69	Caspase recruitment domain-containing protein 6; CARD6; CARD6	[BioPr]	GO:0042981 - regulation of apoptotic process / GO:0006915 - apoptotic process / GO:0043122 - regulation of I-kappaB kinase/NF-kappaB cascade
					[CComp]	GO:0005622 - intracellular / GO:0005737 - cytoplasm
61	Q9C029	HSAPI	Q9C029	Tripartite motif-containing protein 7; TRIM7; GNIP, RNF90; TRIM7	[CComp]	GO:0005622 - intracellular / GO:0005634 - nucleus / GO:0005737 - cytoplasm
		HSAPI	Q96A61	Tripartite motif-containing protein 52; TRIM52; RNF102; TRI52	[MolFn]	GO:0008270 - zinc ion binding
62	Q9H4I9					
63	Q9NQT5	HSAPI	Q9NQT5	Exosome complex component RRP40; EXOSC3; CGI-102; EXOS3	[BioPr]	GO:0045190 - isotype switching / GO:0006364 - rRNA processing / GO:0045006 - DNA deamination / GO:0043928 - exonucleolytic nuclear-transcribed mRNA catabolic process involved in deadenylation-dependent decay / 1 other
		DRERI	F1R2H8	Uncharacterized protein; exosc3	[CComp]	GO:0000177 - cytoplasmic exosome (Rnase complex) / GO:0005730 - nucleolus / GO:0000176 - nuclear exosome (Rnase complex) / GO:0035327 - transcriptionally active chromatin / 2 others

Table 2. *Homologous Gene Data for 1:1 Human-Matched Axolotl Data to Zebrafish with Gene Ontologies, at 90/70 Threshold*

	Matched UniProt ID	Ortholog Organism	Ortholog UniProt ID	Ortholog Description	GO Type	GO Descriptions
					[MolFn]	GO:0003723 - RNA binding / GO:0000175 - 3'-5'-exoribonuclease activity
64	Q9NRA8					
65	Q9P0J6	HSAPI	Q9P0J6	L36mt; MRPL36; BRIP1; RM36	[BioPr]	GO:0006412 - translation
		DRERI	Q1LWG3	L36mt; mrpl36; si:dkey-261i16.3; RM36	[CComp]	GO:0005840 - ribosome / GO:0005739 - mitochondrion / GO:0005762 - mitochondrial large ribosomal subunit
					[MolFn]	GO:0003735 - structural constituent of ribosome
66	Q9P0U1	HSAPI	Q9P0U1	Mitochondrial import receptor subunit TOM7 homolog; TOMM7	[BioPr]	GO:0006886 - intracellular protein trasnport / GO:0006626 - protein targeting to mitochondrion / GO:0044267 - cellular protein metabolic process
		DRERI	Q0P4E8	Uncharacterized protein; Zgc:152998; tomm7	[CComp]	GO:0005741 - mitochondrial outer membrane / GO:0005742 - mitochondrial outer membrane translocase complex / GO:0016021 - integral to membrane / GO:0005739 - mitochondrion

Table 2. *Homologous Gene Data for 1:1 Human-Matched Axolotl Data to Zebrafish with Gene Ontologies, at 90/70 Threshold*

	Matched UniProt ID	Ortholog Organism	Ortholog UniProt ID	Ortholog Description	GO Type	GO Descriptions
					[MolFn]	GO:0015450 - P-P-bond-hydrolysis-driven protein transmembrane transporter activity / GO:0008320 - protein transmembrane transporter activity
67	Q9P1F3	HSAPI	Q9P1F3	Costars family protein ABRACL; C6orf115, HSPC280, PRO2013; ABRAL		
		DRERI	Q6TGV7	Costars family protein ABRACL; abracl; si:dkey-34f16.3; ABRAL		
68	Q9UDV7	HSAPI	Q9UDV7	Zinc finger protein 282; ZNF282; HUB1; ZN282	[BioPr]	GO:0006355 - regulation of transcription, DNA-dependent / GO:0006351 - transcription, DNA-dependent / GO:0045892 - negative regulation of transcription, DNA-dependent
					[CComp]	GO:0005622 - intracellular / GO:0005634 - nucleus
					[MolFn]	GO:0003676 - nucleic acid binding / GO:0008270 - zinc ion binding / GO:0003677 - DNA binding
69	Q9UI09					

Table 2. *Homologous Gene Data for 1:1 Human-Matched Axolotl Data to Zebrafish with Gene Ontologies, at 90/70 Threshold*

	Matched UniProt ID	Ortholog Organism	Ortholog UniProt ID	Ortholog Description	GO Type	GO Descriptions
70	Q9UII2	HSAPI	Q9UII2	ATPase inhibitor, mitochondrial; ATPIF1; ATPI; ATIF1	[BioPr]	GO:0045980 - negative regulation of nucleotide metabolic process / GO:0043086 - negative regulation of catalytic activity / GO:0001937 - negative regulation of endothelial cell proliferation / GO:0051346 - negative regulation of hydrolase activity / 6 others
		DRERI	A3KNL5	Uncharacterized protein; Zgc:162207 protein; atpif1	[CComp]	GO:0005739 - mitochondrion / GO:0009986 - cell surface / GO:0005753 - mitochondrial proton-transporting ATP synthase complex / GO:0005634 - nucleus
					[MolFn]	GO:0004857 - enzyme inhibitor activity / GO:0042030 - ATPase inhibitor activity / GO:0003677 - DNA binding
71	Q9ULW0	HSAPI	Q9ULW0	Targeting protein for Xklp2; TPX2; C20orf1, C20orf2, DIL2, HCA519; TPX2	[BioPr]	GO:0060236 - regulation of mitotic spindle organization / GO:0006915 - apoptotic process / GO:0007067 - mitosis / GO:0051301 - cell division / 2 others
		DRERI	E9QE51	Uncharacterized protein; tpx2	[CComp]	GO:0005634 - nucleus / GO:0005819 - spindle / GO:0000922 - spindle pole /GO:0005737 - cytoplasm / 1 other
					[MolFn]	GO:0005524 - ATP binding / GO:0005525 - GTP binding
72	Q9UN76	HSAPI	Q9UN76	Sodium- and chloride-dependent neutral and basic amino acid transporter B(0+); SLC6A14	[BioPr]	GO:0009636 - response to toxin / GO:0006811 - ion transport / GO:0006520 - cellular amino acid metabolic process / GO:0006865 - amino acid transport

Table 2. *Homologous Gene Data for 1:1 Human-Matched Axolotl Data to Zebrafish with Gene Ontologies, at 90/70 Threshold*

	Matched UniProt ID	Ortholog Organism	Ortholog UniProt ID	Ortholog Description	GO Type	GO Descriptions
		DRERI	Q1ECY9	Transporter; LOC100150665	[CComp]	GO:0016021 - integral to membrane / GO:0031526 - brush border membrane / GO:0005886 - plasma membrane
					[MolFn]	GO:0005328 - neurotransmitter:sodium symporter activity / GO:0015171 - amino acid transmembrane transporter activity
73	Q9UNX3	HSAPI	Q9UNX3	60S ribosomal protein L26-like 1; RPL26L1; RPL26P1; RL26L	[BioPr]	GO:0006412 - transform / GO:0006364 - rRNA processing / GO:0042273 - ribosomal large subunit biogenesis / GO:0006415 - translational termination / 5 others
		HSAPI	P61254	60S ribosomal protein L26; RPL26; RL26	[CComp]	GO:0015934 - large ribosomal subunit / GO:0022625 - cytosolic large ribosomal subunit / GO:0005829 - cytosol
		DRERI	Q7SXA1	Ribosomal protein L26; Uncharacterized protein; rpl26	[MolFn]	GO:0003735 - structural constituent of ribosome / GO:0003723 - RNA binding
74	Q9Y279					
75	Q9Y2Q9	HSAPI	Q9Y2Q9	MRP-S28; MRPS28; MRPS35; HSPC007; RT28	[CComp]	GO:0005739 - mitochondrion / GO:0005840 - ribosome / GO:0005763 - mitochondrial small ribosomal subunit

Table 2. *Homologous Gene Data for 1:1 Human-Matched Axolotl Data to Zebrafish with Gene Ontologies, at 90/70 Threshold*

	Matched UniProt ID	Ortholog Organism	Ortholog UniProt ID	Ortholog Description	GO Type	GO Descriptions
		DRERI	A5PMP2	Novel protein similar to vertebrate mitochondrial ribosomal protein S28 (MPRS28); Uncharacterized protein		
76	Q9Y5U4					
77	Q9Y6Q3	HSAPI	Q9Y6Q3	Zfp-37; ZFP37; ZFP37	[BioPr]	GO:0006355 - regulation of transcription, DNA-dependent / GO:0006351 - transcription, DNA-dependent / GO:0008283 - cell proliferation / GO0030154 - cell differentiation / 3 others
					[CComp]	GO:0005622 - intracellular / GO:0005634 - nucleus
					[MolFn]	GO:0008270 - zinc ion binding / GO:0003676 - nucleic acid binding / GO:0003677 - DNA binding / GO:0003700 - sequence-specific DNA binding transcription factor activity
78	Q9Y6R7	HSAPI	Q9Y6R7	IgGFc-binding protein; FCGBP; FCGBP	[BioPr]	GO:0007160 - cell-matrix adhesion
		DRERI	E7F2A5	Uncharacterized protein; CU459152.1	[CComp]	GO:0005576 - extracellular region / GO:0005737 - cytoplasm / GO:0005578 - proteinaceous extracellular matrix

Table 2. *Homologous Gene Data for 1:1 Human-Matched Axolotl Data to Zebrafish with Gene Ontologies, at 90/70 Threshold*

	Matched UniProt ID	Ortholog Organism	Ortholog UniProt ID	Ortholog Description	GO Type	GO Descriptions
		DRERI	F1RDU8	Uncharacterized protein; si:dkey-65b12.6		

One of the larger sets of orthologous genes with a common function appear to be those relating to different varieties of ribosomal proteins; of the 159 gene homologs given, 15 genes represent proteins related to this organelle. The 40S and 60S ribosomal proteins represent the small and large subunits, respectively, required by the ribosome to carry out protein synthesis via translation. Ribosomes create proteins by stringing together mRNA-determined sequences of amino acids. The small subunit is responsible for reading the mRNA, while the large subunit uses this information to create polypeptides from the amino acids.[26] For example, gene P05388 (60S acidic ribosomal protein P0) is homologous to zebrafish gene Q5P6K3 (60S acidic ribosomal protein P0), gene P23396 (40S ribosomal protein S3) is homologous to zebrafish gene Q6TLG8 (ribosomal protein S3), gene P46783 (40S ribosomal protein S10) is homologous to zebrafish gene Q7T1J9 (ribosomal protein S10), gene P62269 (40S ribosomal protein S18) is homologous to zebrafish gene Q8JGS9 (40S ribosomal protein S18), gene P62277 (40S ribosomal protein S13) is homologous to zebrafish gene Q6IMW6 (ribosomal protein S13), and genes Q9UNX3 (60S ribosomal protein L26-like 1) and P61254 (60S ribosomal protein L26) are homologous to zebrafish gene Q78XA1 (ribosomal protein L26). Gene P36578 (60S ribosomal protein L4) is homologous to zebrafish gene Q7ZW95 (ribosomal protein L4); ribosomal protein L4 plays a functional role in the regulation of neurite regeneration and is regulated via translation during this phenomenon. As scar tissue does not contain nerve endings, and because both axolotl and the zebrafish see an absence of scar tissue during the regenerative process, ribosomal protein L4 is likely plays an important role in the maintenance of proper nerve development in regenerating tissues.[32]

Another set of orthologous gene descriptions in Table 2 is that of NADH dehydrogenase [ubiquinone] 1 alpha and beta; of the 159 gene homologs given, 10 genes relate to this function. As an entry enzyme for the mitochondrial respiratory chain, NADH dehydrogenase [ubiquinone], acts as a catalyst for electron transfer between NADH and coenzyme Q (CoQ).[33] Also known as NADH ubiquinone oxidoreductase, this protein complex is one of four that is responsible for the pumping of protons across the mitochondrial cell membrane.[34] Studies have shown that the Complex I type of NADH ubiquinone oxidoreductase may play a role in initiating apoptosis, a mechanism of programmed cell death which other studies have linked to the process of limb regeneration via providing the regenerating cell with a way to prevent the growth of unwanted cellular masses.[35,36]

Various forms of histone genes comprise 9 of the 162 gene homologs found in Table 2. Histone H3 proteins are core histones responsible for arranging DNA into nucleosomes. Histones are primary components of chromatin and involved with gene regulation. Gene ontology for histone H3.1 includes the regulation of gene silencing and blood coagulation, and that of histone H3.3 includes the extracellular region. These ontologies connect histones with the ECM in that extracellular histones have been found to be mediators in the processes of inflammation and thrombosis.[37]

Other gene homolog groups of interest include parvalbumin, which are calcium-binding proteins that play a role in cell-cycle regulation especially in fast-contracting

muscle tissue, and the IFIT family of proteins, which have been shown to exhibit antiviral traits and could therefore be potentially capable of helping mediate innate immune responses.[38,39]

Some gene ontologies from Table 2 appear to be potentially related to the process of regeneration or ECM. These ontologies include GO:0001824 - blastocyst development, GO:0005576 - extracellular region, GO:0005578 - proteinaceous extracellular matrix, GO:0005615 - extracellular space, GO:0006915 - apoptotic process, GO:0006916 - anti-apoptosis, GO:0007275 - multicellular organismal development, GO:0030154 - cell differentiation, GO:0031100 - organ regeneration, GO:0042981 - regulation of apoptotic process, and GO:0060968 - regulation of gene silencing. A more statistical approach, however, to facilitate the identification of any significant gene ontologies for all 159 gene homologs is found in Table 3 (pages 59-60), which lists all gene ontology descriptions from Table 2 found more than once.

Table 3. *List of Most Common GO Descriptions for Table 2 Genes*

GO Description	Count
GO:0005634 - nucleus	14
GO:0003723 - RNA binding	11
GO:0005737 - cytoplasm	9
GO:0005739 - mitochondrion	9
GO:0003735 - structural constituent of ribosome	8
GO:0005829 - cytosol	8
GO:0005840 - ribosome	7
GO:0006412 - translation	6
GO:0005576 - extracellular region	5
GO:0006810 - transport	5
GO:0008137 - NADH dehydrogenase (ubiquinone) activity	5
GO:0016021 - integral to membrane	5
GO:0044281 - small molecule metabolic process	5
GO:0005622 - intracellular	4
GO:0005743 - mitochondrial inner membrane	4
GO:0005747 - mitochondrial respiratory chain complex I	4
GO:0006413 - translational initiation	4
GO:0006415 - translational termination	4
GO:0008270 - zinc ion binding	4
GO:0022627 - cytosolic small ribosomal subunit	4
GO:0003676 - nucleic acid binding	3
GO:0003677 - DNA binding	3
GO:0005730 - nucleolus	3
GO:0006120 - mitochondrial electron transport, NADH to ubiquinone	3
GO:0006364 - rRNA processing	3

Table 3. *List of Most Common GO Descriptions for Table 2 Genes*

GO Description	Count
GO:0006915 - apoptotic process	3
GO:0015935 - small ribosomal subunit	3
GO:0022625 - cytosolic large ribosomal subunit	3
GO:0022900 - electron transport chain	3
GO:0071013 - catalytic step 2 spliceosome	3
GO:0003676 - nucleic acid binding	2
GO:0005524 - ATP binding	2
GO:0005615 - extracellular space	2
GO:0005654 - nucleoplasm	2
GO:0005689 - U12-type spliceosomal complex	2
GO:0005886 - plasma membrane	2
GO:0006351 - transcription, DNA-dependent	2
GO:0006355 - regulation of transcription, DNA-dependent	2
GO:0006397 - mRNA processing	2
GO:0006414 - translational elongation	2
GO:0006979 - response to oxidative stress	2
GO:0008270 - zinc ion binding	2
GO:0008380 - RNA splicing	2
GO:0016021 - integral to membrane	2
GO:0019083 - viral transcription	2
GO:0042254 - ribosome biogenesis	2
GO:0044267 - cellular protein metabolic process	2
GO:0046872 - metal ion binding	2
GO:0070469 - respiratory chain	2

Many of the more common gene ontology descriptions for the 162 gene homologs are structural in nature: nucleus, cytoplasm, mitochondrion, cytosol, ribosome, etc. Other gene ontologies occurring in great number have already been touched upon: RNA binding, translation, transport, NADH dehydrogenase [ubiquinone] activity, mitochondrial respiratory chain complex I.

There are numerous genes that relate to the ECM as determined by their gene ontologies, and because the formation and maintenance of which would be an integral part in the regeneration of fully functional tissue, this study will now look at these genes. As many of these homologous zebrafish genes are larger uncharacterized, a more thorough understanding of the mechanisms these genes take part in, especially in mediating the ECM with respect to regeneration, should be acquired.

Gene P01036 (Cystatin-S; CST4) is homologous to zebrafish gene B8A4D0, a largely uncharacterized protein. The gene ontology for this group of homologous genes consists of negative regulation of blood vessel remodeling, fibril organization, negative regulation of the collagen catabolic process, and negative regulation of elastin catabolic process, among others. Axolotl gene expression data from MDIBL showed the downregulation (decreased gene expression as compared to a reference) of collagen genes very early in the process of regeneration, only for upregulation (increased gene expression as compared to a reference) of the same genes to occur at a time after, so the presence of these genes in the zebrafish may signal a similar mechanism for outgrowth.

Genes P43235 (Cathepsin K; CTSK) and P25774 (Cathepsin S; CTSS) are homologous to zebrafish genes F1Q8A0 (CTSK), A2BF64 (novel protein), and Q502A6 (Cathepsin S); these three zebrafish genes are also uncharacterized proteins. Cathepsins are proteases that can, in response to a signal under certain circumstances, trigger apoptosis through numerous pathways, such as via the release of mitochondrial proapoptotic factors. These proteases can thus play a role in controlling cell turnover

within an organism.[40] Upon the occurrence of an incision or amputation, damaged cells must be removed before the ECM can begin to rebuild itself and new, functional tissues can be regenerated.

One such example of a gene homolog pair that researchers may choose to study more in-depth is gene Q9Y6R7 (IgGFc-binding protein; FCGBP) and its homologous zebrafish genes E7F2A5 and F1RDU8, both of which are uncharacterized proteins. The related gene ontology terms for this homologous pair include cell-matrix adhesion, extracellular region, cytoplasm, and proteinaceous extracellular matrix. In particular, cell-matrix adhesion (GO:0007160) in particular may be of interest to this particular homologous pair of genes because this gene ontology only occurs once in the data in Table 2. Because the ECM provides a framework for cellular support in tissue and organ systems, cell adhesion is necessary in order to allow this molecular scaffold and the structure of the cell's surface to become more tightly linked. Cell-matrix adhesion can also provide signaling to work with biological processes such as wound healing and cell proliferation.[41]

As a brief aside, a directory of all genes related to the ECM was downloaded from the Ensembl Genome Browser via BioMart and compared against the axolotl data.[26] Of the 470 unique genes in this list, 322 intersections between the axolotl data and the ECM gene list were found. Genes with 1:1 mappings from these intersections at an 80/30 threshold are listed in Table 4 (page 63); Table 4 will not be subject to in-depth discussion but has been given for reference purposes.

Table 4. *List of Axolotl Data – Human ECM Data Intersections at 80/30 Threshold*

	UniProt ID	Description	HGNC Symbol for *Homo sapiens*
1	O00602	ficolin (collagen/fibrinogen domain containing) 1	FCN1_HUMAN
2	O75443	tectorin alpha	TECTA_HUMAN
3	P01033	TIMP metallopeptidase inhibitor	TIMP1_HUMAN
4	P01034	cystatin C	CYTC_HUMAN
5	P02747	complement component 1, q subcomponent, C chain	C1QC_HUMAN
6	P03956	matrix metallopeptidase 1 (interstitial collagenase)	MMP1_HUMAN
7	P07355	annexin A2	ANXA2_HUMAN
8	P08294	superoxide dismutase 3, extracellular	SODE_HUMAN
9	P09382	lectin, galactoside-binding, soluble, 1	LEG1_HUMAN
10	P12109	collagen, type VI, alpha 1	CO6A1_HUMAN
11	P12111	collagen, type VI, alpha 3	CO6A3_HUMAN
12	P27658	collagen, type VIII, alpha 1	CO8A1_HUMAN
13	P45452	matrix metallopeptidase 13 (collagenase 3)	MMP13_HUMAN
14	P48307	tissue factor pathway inhibitor 2	TFPI2_HUMAN
15	P62269	ribosomal protein S18	RS18_HUMAN
16	Q07507	dermatopontin	DERM_HUMAN
17	Q08380	lectin, galactoside-binding, soluble, 3 binding protein	LG3BP_HUMAN
18	Q15465	sonic hedgehog	SHH_HUMAN
19	Q15848	adiponectin, C1Q and collagen domain containing	ADIPO_HUMAN
20	Q6NUI6	chondroadherin-like	CHADL_HUMAN

Another version of Table 2 can also be created to show changes in gene expression of these axolotl genes 1, 3, 5, and 7 dpa; this data was obtained from MDIBL. Table 5 (pages 65-76) shows these changes in levels of gene expression as fold changes, a commonly used scale in bioinformatics. A fold change of *n* after *x* dpa can be considered a 2^n times change in gene expression from day 0 to day *x* (i.e. a fold change of -0.15 equals a 0.90, or 90%, expression rate as compared to the reference). A positive fold change indicates upregulation of a gene, and a negative fold change indicates downregulation of a gene.

Table 5 is sorted from greatest to least value for the parameter "7 dpa Fold change."

Table 5. *Gene Expression Fold Changes for Table 2 Genes*

	Affymetric ID	Matched UniProt ID	Ortholog Organism	Ortholog UniProt ID	1 dpa Fold change	1 dpa adjusted p-val	3 dpa Fold change	3 dpa adjusted p-val	5 dpa Fold change	5 dpa adjusted p-val	7 dpa Fold change	7 dpa adjusted p-val
17	SRV_03349_at	P16112			4.622	0.0000	5.449	0.0000	5.165	0.0000	4.652	0.0000
25	SRV_00328_at	P43235	HSAPI	P43235	-0.420	0.2275	0.506	0.1357	2.054	0.0000	3.110	0.0000
25a			HSAPI	P25774	-0.420	0.2275	0.506	0.1357	2.054	0.0000	3.110	0.0000
25b			DRERI	A2BF64	-0.420	0.2275	0.506	0.1357	2.054	0.0000	3.110	0.0000
25c			DRERI	F1Q8A0	-0.420	0.2275	0.506	0.1357	2.054	0.0000	3.110	0.0000
25d			DRERI	Q502A6	-0.420	0.2275	0.506	0.1357	2.054	0.0000	3.110	0.0000
39	SRV_03588_at	Q15004	HSAPI	Q15004	-0.476	0.0024	1.455	0.0000	1.625	0.0000	1.766	0.0000
67	SRV_05942_at	Q9P1F3	HSAPI	Q9P1F3	0.427	0.0000	0.851	0.0000	0.947	0.0000	1.161	0.0000
67a			DRERI	Q6TGV7	0.427	0.0000	0.851	0.0000	0.947	0.0000	1.161	0.0000
68a					0.427	0.0000	0.851	0.0000	0.947	0.0000	1.161	0.0000
68b					0.427	0.0000	0.851	0.0000	0.947	0.0000	1.161	0.0000
71	SRV_03257_at	Q9ULW0	HSAPI	Q9ULW0	-0.653	0.0004	-0.208	0.1145	0.778	0.0000	0.948	0.0000
71a			DRERI	E9QE51	-0.653	0.0004	-0.208	0.1145	0.778	0.0000	0.948	0.0000
71b					-0.653	0.0004	-0.208	0.1145	0.778	0.0000	0.948	0.0000
9	SRV_02906_a_at	O75792	HSAPI	O75792	-0.416	0.0022	0.298	0.0141	0.453	0.0005	0.753	0.0000
9a			DRERI	Q6TNR0	-0.416	0.0022	0.298	0.0141	0.453	0.0005	0.753	0.0000
9b					-0.416	0.0022	0.298	0.0141	0.453	0.0005	0.753	0.0000
58	SRV_05596_at	Q96MD7	HSAPI	Q96MD7	-0.008	0.5245	0.359	0.0002	0.450	0.0000	0.588	0.0000
58a			DRERI	A3KPQ4	-0.008	0.5245	0.359	0.0002	0.450	0.0000	0.588	0.0000

Table 5. *Gene Expression Fold Changes for Table 2 Genes*

	Affymetric ID	Matched UniProt ID	Ortholog Organism	Ortholog UniProt ID	1 dpa Fold change	1 dpa adjusted p-val	3 dpa Fold change	3 dpa adjusted p-val	5 dpa Fold change	5 dpa adjusted p-val	7 dpa Fold change	7 dpa adjusted p-val
65	SRV_04954_at	Q9P0J6	HSAPI	Q9P0J6	-0.093	0.1624	0.284	0.0004	0.492	0.0000	0.583	0.0000
65a			DRERI	Q1LWG3	-0.093	0.1624	0.284	0.0004	0.492	0.0000	0.583	0.0000
65b					-0.093	0.1624	0.284	0.0004	0.492	0.0000	0.583	0.0000
37	SRV_03149_at	P62312	HSAPI	P62306	-0.118	0.1104	0.118	0.0823	0.396	0.0000	0.582	0.0000
37a			HSAPI	P62312	-0.118	0.1104	0.118	0.0823	0.396	0.0000	0.582	0.0000
37b			DRERI	H9GXB1	-0.118	0.1104	0.118	0.0823	0.396	0.0000	0.582	0.0000
37c			DRERI	Q6IMW9	-0.118	0.1104	0.118	0.0823	0.396	0.0000	0.582	0.0000
32	SRV_03340_at	P62072	HSAPI	P62072	0.363	0.0223	0.658	0.0001	0.455	0.0037	0.479	0.0015
32a			DRERI	Q6DI06	0.363	0.0223	0.658	0.0001	0.455	0.0037	0.479	0.0015
32b					0.363	0.0223	0.658	0.0001	0.455	0.0037	0.479	0.0015
5	SRV_03329_at	O14879	HSAPI	O14879	1.593	0.0001	0.860	0.0133	0.670	0.0395	0.453	0.1069
5a			HSAPI	P09913	1.593	0.0001	0.860	0.0133	0.670	0.0395	0.453	0.1069
5b			HSAPI	Q13325	1.593	0.0001	0.860	0.0133	0.670	0.0395	0.453	0.1069
5c			HSAPI	P09914	1.593	0.0001	0.860	0.0133	0.670	0.0395	0.453	0.1069
5d			HSAPI	Q5T764	1.593	0.0001	0.860	0.0133	0.670	0.0395	0.453	0.1069
5e			DRERI	E7FBH4	1.593	0.0001	0.860	0.0133	0.670	0.0395	0.453	0.1069
5f			DRERI	E7F8D8	1.593	0.0001	0.860	0.0133	0.670	0.0395	0.453	0.1069
5g			DRERI	F6P8G1	1.593	0.0001	0.860	0.0133	0.670	0.0395	0.453	0.1069
5h			DRERI	F1QG25	1.593	0.0001	0.860	0.0133	0.670	0.0395	0.453	0.1069

Table 5. *Gene Expression Fold Changes for Table 2 Genes*

	Affymetric ID	Matched UniProt ID	Ortholog Organism	Ortholog UniProt ID	1 dpa Fold change	1 dpa adjusted p-val	3 dpa Fold change	3 dpa adjusted p-val	5 dpa Fold change	5 dpa adjusted p-val	7 dpa Fold change	7 dpa adjusted p-val
5i			DRERI	B8A535	1.593	0.0001	0.860	0.0133	0.670	0.0395	0.453	0.1069
5j			DRERI	E7FA13	1.593	0.0001	0.860	0.0133	0.670	0.0395	0.453	0.1069
5k			DRERI	F1QW56	1.593	0.0001	0.860	0.0133	0.670	0.0395	0.453	0.1069
5l			DRERI	E7FCM5	1.593	0.0001	0.860	0.0133	0.670	0.0395	0.453	0.1069
5m			DRERI	F1QGN7	1.593	0.0001	0.860	0.0133	0.670	0.0395	0.453	0.1069
3	SRV_03881_a_at	C9K0J4			-0.708	0.0388	0.271	0.2212	0.070	0.3870	0.387	0.1287
57	SRV_05313_at	Q96E66			-0.288	0.0352	-0.396	0.0035	0.114	0.2064	0.386	0.0033
63	SRV_03835_at	Q9NQT5	HSAPI	Q9NQT5	0.040	0.3989	0.267	0.0037	0.285	0.0023	0.380	0.0001
63a			DRERI	F1R2H8	0.040	0.3989	0.267	0.0037	0.285	0.0023	0.380	0.0001
63b					0.040	0.3989	0.267	0.0037	0.285	0.0023	0.380	0.0001
53	SRV_06076_a_at	Q8N6V9			-0.417	0.0128	-0.138	0.2046	0.234	0.0828	0.375	0.0132
31	SRV_00867_a_at	P61803	HSAPI	P61803	0.008	0.4912	0.201	0.0001	0.341	0.0000	0.351	0.0000
31a			DRERI	A7E2L0	0.008	0.4912	0.201	0.0001	0.341	0.0000	0.351	0.0000
31b					0.008	0.4912	0.201	0.0001	0.341	0.0000	0.351	0.0000
35	SRV_01722_at	P62308	HSAPI	P62308	-0.016	0.4601	0.235	0.0002	0.268	0.0000	0.348	0.0000
35a			HSAPI	F5H5R7	-0.016	0.4601	0.235	0.0002	0.268	0.0000	0.348	0.0000
35b			DRERI	F1QPY7	-0.016	0.4601	0.235	0.0002	0.268	0.0000	0.348	0.0000
36	SRV_03544_at	P62310	HSAPI	P62310	0.014	0.4705	0.205	0.0008	0.210	0.0006	0.337	0.0000
36a			DRERI	E7EZE6	0.014	0.4705	0.205	0.0008	0.210	0.0006	0.337	0.0000

Table 5. *Gene Expression Fold Changes for Table 2 Genes*

	Affymetric ID	Matched UniProt ID	Ortholog Organism	Ortholog UniProt ID	1 dpa Fold change	1 dpa adjusted p-val	3 dpa Fold change	3 dpa adjusted p-val	5 dpa Fold change	5 dpa adjusted p-val	7 dpa Fold change	7 dpa adjusted p-val
36b					0.014	0.4705	0.205	0.0008	0.210	0.0006	0.337	0.0000
44	SRV_03979_at	Q4U2R6	HSAPI	Q4U2R6	0.073	0.1168	0.191	0.0004	0.264	0.0000	0.312	0.0000
44a			DRERI	Q5BJJ8	0.073	0.1168	0.191	0.0004	0.264	0.0000	0.312	0.0000
44b					0.073	0.1168	0.191	0.0004	0.264	0.0000	0.312	0.0000
30	SRV_01263_at	P60842	HSAPI	P60842	0.571	0.0006	-0.031	0.3948	0.131	0.2001	0.192	0.1054
30a			DRERI	Q7ZU67	0.571	0.0006	-0.031	0.3948	0.131	0.2001	0.192	0.1054
30b			DRERI	F1R166	0.571	0.0006	-0.031	0.3948	0.131	0.2001	0.192	0.1054
30c			DRERI	Q802C9	0.571	0.0006	-0.031	0.3948	0.131	0.2001	0.192	0.1054
1	SRV_02446_a_at	A2AJT9	HSAPI	A2AJT9	-0.166	0.0837	0.041	0.3315	0.007	0.4269	0.176	0.0462
27	SRV_04503_at	P56181	HSAPI	P56181	-0.104	0.0788	-0.020	0.3551	0.100	0.0636	0.174	0.0039
27a					-0.104	0.0788	-0.020	0.3551	0.100	0.0636	0.174	0.0039
27b					-0.104	0.0788	-0.020	0.3551	0.100	0.0636	0.174	0.0039
60	SRV_04970_a_at	Q9BX69	HSAPI	Q9BX69	0.049	0.4108	0.064	0.2945	0.023	0.3916	0.134	0.1330
60a					0.049	0.4108	0.064	0.2945	0.023	0.3916	0.134	0.1330
55	SRV_09790_at	Q8WVI0			-0.121	0.1300	-0.056	0.2633	-0.078	0.1940	0.088	0.1545
76	SRV_03884_at	Q9Y5U4			-0.043	0.2600	0.011	0.3926	-0.064	0.1136	0.066	0.1003
19	SRV_02129_at	P20674	HSAPI	P20674	-0.053	0.1588	0.082	0.0353	0.054	0.1129	0.053	0.1088
19a			DRERI	Q4VBU7	-0.053	0.1588	0.082	0.0353	0.054	0.1129	0.053	0.1088
19b			DRERI	F1QJ99	-0.053	0.1588	0.082	0.0353	0.054	0.1129	0.053	0.1088

Table 5. *Gene Expression Fold Changes for Table 2 Genes*

	Affymetric ID	Matched UniProt ID	Ortholog Organism	Ortholog UniProt ID	1 dpa Fold change	1 dpa adjusted p-val	3 dpa Fold change	3 dpa adjusted p-val	5 dpa Fold change	5 dpa adjusted p-val	7 dpa Fold change	7 dpa adjusted p-val
19c			DRERI	Q4VBU7	-0.053	0.1588	0.082	0.0353	0.054	0.1129	0.053	0.1088
20	SRV_00071_s_at	P23396	HSAPI	P23396	-0.022	0.1588	0.028	0.0696	0.021	0.1271	0.034	0.0341
20a			DRERI	Q6TLG8	-0.022	0.1588	0.028	0.0696	0.021	0.1271	0.034	0.0341
20b					-0.022	0.1588	0.028	0.0696	0.021	0.1271	0.034	0.0341
2	SRV_00936_a_at	B4DZF2			-0.383	0.1141	0.269	0.1656	-0.067	0.3736	0.032	0.3748
51	SRV_02889_a_at	Q7L2H7			-0.093	0.0604	0.052	0.1627	0.088	0.0493	0.020	0.3017
23	SRV_00581_at	P36578	HSAPI	P36578	-0.083	0.2468	0.016	0.4051	-0.007	0.4254	0.015	0.3635
23a			DRERI	Q7ZW95	-0.083	0.2468	0.016	0.4051	-0.007	0.4254	0.015	0.3635
23b					-0.083	0.2468	0.016	0.4051	-0.007	0.4254	0.015	0.3635
56	SRV_07869_at	Q96BP2			-0.017	0.4455	-0.082	0.0873	-0.020	0.3454	-0.005	0.3852
24	SRV_03185_a_at	P42766			-0.021	0.1157	0.016	0.1433	0.018	0.1162	-0.007	0.2831
33	SRV_04622_a_at	P62269	HSAPI	P62269	0.004	0.4385	0.007	0.3003	0.004	0.3646	-0.007	0.2709
33a			DRERI	Q8JGS9	0.004	0.4385	0.007	0.3003	0.004	0.3646	-0.007	0.2709
33b					0.004	0.4385	0.007	0.3003	0.004	0.3646	-0.007	0.2709
64	SRV_04361_a_at	Q9NRA8			-0.099	0.2773	0.049	0.3351	-0.076	0.2700	-0.008	0.3897
26	SRV_00671_at	P46783	HSAPI	P46783	-0.019	0.2157	-0.010	0.3014	-0.002	0.4131	-0.023	0.1162
26a			DRERI	Q7T1J9	-0.019	0.2157	-0.010	0.3014	-0.002	0.4131	-0.023	0.1162
45	SRV_03891_at	Q5U5M0			-0.224	0.1198	-0.419	0.0069	-0.278	0.0467	-0.025	0.3658
73	SRV_03860_at	Q9UNX3	HSAPI	Q9UNX3	0.043	0.4853	-0.165	0.2296	-0.017	0.4230	-0.037	0.3623

Table 5. *Gene Expression Fold Changes for Table 2 Genes*

	Affymetric ID	Matched UniProt ID	Ortholog Organism	Ortholog UniProt ID	1 dpa Fold change	1 dpa adjusted p-val	3 dpa Fold change	3 dpa adjusted p-val	5 dpa Fold change	5 dpa adjusted p-val	7 dpa Fold change	7 dpa adjusted p-val
73a			HSAPI	P61254	0.043	0.4853	-0.165	0.2296	-0.017	0.4230	-0.037	0.3623
73b			DRERI	Q7SXA1	0.043	0.4853	-0.165	0.2296	-0.017	0.4230	-0.037	0.3623
34	SRV_00675_a_at	P62277	HSAPI	P62277	-0.023	0.0930	-0.070	0.0000	-0.046	0.0018	-0.052	0.0003
34a			DRERI	Q6IMW6	-0.023	0.0930	-0.070	0.0000	-0.046	0.0018	-0.052	0.0003
34b					-0.023	0.0930	-0.070	0.0000	-0.046	0.0018	-0.052	0.0003
16	SRV_12295_at	P10589	HSAPI	P10588	-0.277	0.0242	0.043	0.3465	0.015	0.4100	-0.053	0.2878
16a			HSAPI	P10589	-0.277	0.0242	0.043	0.3465	0.015	0.4100	-0.053	0.2878
16b			HSAPI	P24468	-0.277	0.0242	0.043	0.3465	0.015	0.4100	-0.053	0.2878
16c			DRERI	Q6P115	-0.277	0.0242	0.043	0.3465	0.015	0.4100	-0.053	0.2878
16d			DRERI	Q6P117	-0.277	0.0242	0.043	0.3465	0.015	0.4100	-0.053	0.2878
16e			DRERI	Q6PH18	-0.277	0.0242	0.043	0.3465	0.015	0.4100	-0.053	0.2878
16f			DRERI	Q06726	-0.277	0.0242	0.043	0.3465	0.015	0.4100	-0.053	0.2878
16g			DRERI	Q91430	-0.277	0.0242	0.043	0.3465	0.015	0.4100	-0.053	0.2878
16h			DRERI	Q06725	-0.277	0.0242	0.043	0.3465	0.015	0.4100	-0.053	0.2878
15	AQ_s_at	P05388	HSAPI	P05388	-0.091	0.0106	0.006	0.4026	-0.026	0.2279	-0.060	0.0448
15a			DRERI	Q6P5K3	-0.091	0.0106	0.006	0.4026	-0.026	0.2279	-0.060	0.0448
15b					-0.091	0.0106	0.006	0.4026	-0.026	0.2279	-0.060	0.0448
11	SRV_02251_at	O95168	HSAPI	O95168	-0.076	0.1674	-0.278	0.0001	-0.185	0.0038	-0.061	0.1610
11a			DRERI	Q6PBK0	-0.076	0.1674	-0.278	0.0001	-0.185	0.0038	-0.061	0.1610

Table 5. *Gene Expression Fold Changes for Table 2 Genes*

	Affymetric ID	Matched UniProt ID	Ortholog Organism	Ortholog UniProt ID	1 dpa Fold change	1 dpa adjusted p-val	3 dpa Fold change	3 dpa adjusted p-val	5 dpa Fold change	5 dpa adjusted p-val	7 dpa Fold change	7 dpa adjusted p-val
11b					-0.076	0.1674	-0.278	0.0001	-0.185	0.0038	-0.061	0.1610
61	SRV_05065_at	Q9C029	HSAPI	Q9C029	0.025	0.4620	-0.072	0.2317	0.015	0.4011	-0.070	0.2137
61a			HSAPI	Q96A61	0.025	0.4620	-0.072	0.2317	0.015	0.4011	-0.070	0.2137
8	SRV_01452_at	O43678	HSAPI	O43678	-0.174	0.0036	-0.164	0.0035	-0.110	0.0319	-0.092	0.0561
8a			DRERI	Q4VBI5	-0.174	0.0036	-0.164	0.0035	-0.110	0.0319	-0.092	0.0561
8b					-0.174	0.0036	-0.164	0.0035	-0.110	0.0319	-0.092	0.0561
49	SRV_04497_at	Q71DI3	HSAPI	P68431	0.177	0.1931	-0.154	0.1822	-0.182	0.1414	-0.098	0.2453
50			HSAPI	P84243	0.177	0.1931	-0.154	0.1822	-0.182	0.1414	-0.098	0.2453
50a			HSAPI	Q16695	0.177	0.1931	-0.154	0.1822	-0.182	0.1414	-0.098	0.2453
50b			HSAPI	Q71DI3	0.177	0.1931	-0.154	0.1822	-0.182	0.1414	-0.098	0.2453
50c			HSAPI	Q5TEC6	0.177	0.1931	-0.154	0.1822	-0.182	0.1414	-0.098	0.2453
50d			DRERI	Q6PI20	0.177	0.1931	-0.154	0.1822	-0.182	0.1414	-0.098	0.2453
50e			DRERI	A8KBJ5	0.177	0.1931	-0.154	0.1822	-0.182	0.1414	-0.098	0.2453
50f			DRERI	G1K2S9	0.177	0.1931	-0.154	0.1822	-0.182	0.1414	-0.098	0.2453
50g			DRERI	Q4QRF4	0.177	0.1931	-0.154	0.1822	-0.182	0.1414	-0.098	0.2453
72	SRV_03196_a_at	Q9UN76	HSAPI	Q9UN76	0.269	0.0893	-0.124	0.2356	-0.119	0.2424	-0.115	0.2252
72a			DRERI	Q1ECY9	0.269	0.0893	-0.124	0.2356	-0.119	0.2424	-0.115	0.2252
72b					0.269	0.0893	-0.124	0.2356	-0.119	0.2424	-0.115	0.2252
12	SRV_01994_at	O95171	HSAPI	O95171	0.619	0.0833	-0.236	0.2648	-0.075	0.3882	-0.137	0.3089

Table 5. *Gene Expression Fold Changes for Table 2 Genes*

	Affymetric ID	Matched UniProt ID	Ortholog Organism	Ortholog UniProt ID	1 dpa Fold change	1 dpa adjusted p-val	3 dpa Fold change	3 dpa adjusted p-val	5 dpa Fold change	5 dpa adjusted p-val	7 dpa Fold change	7 dpa adjusted p-val
12a			DRERI	A8WHU6	0.619	0.0833	-0.236	0.2648	-0.075	0.3882	-0.137	0.3089
12b					0.619	0.0833	-0.236	0.2648	-0.075	0.3882	-0.137	0.3089
38	SRV_02637_at	Q14061			0.241	0.0146	-0.111	0.1335	-0.149	0.0677	-0.140	0.0749
28	SRV_01454_at	P56556	HSAPI	P56556	-0.190	0.0009	-0.158	0.0028	-0.107	0.0280	-0.140	0.0057
28a			DRERI	F1RBK6	-0.190	0.0009	-0.158	0.0028	-0.107	0.0280	-0.140	0.0057
28b					-0.190	0.0009	-0.158	0.0028	-0.107	0.0280	-0.140	0.0057
68	SRV_01894_a_at	Q9UDV7	HSAPI	Q9UDV7	-0.225	0.0374	0.003	0.4481	-0.105	0.1744	-0.159	0.0752
74	SRV_03206_at	Q9Y279			1.329	0.0000	0.196	0.1609	-0.040	0.3854	-0.165	0.1806
43	SRV_00060_x_at	Q30201			-0.146	0.0873	0.123	0.0989	-0.085	0.1829	-0.194	0.0205
70	SRV_10299_at	Q9UII2	HSAPI	Q9UII2	-0.156	0.0002	-0.131	0.0007	-0.154	0.0001	-0.207	0.0000
70a			DRERI	A3KNL5	-0.156	0.0002	-0.131	0.0007	-0.154	0.0001	-0.207	0.0000
70b					-0.156	0.0002	-0.131	0.0007	-0.154	0.0001	-0.207	0.0000
21	SRV_01135_a_at	P27449	HSAPI	P27449	-0.129	0.0997	0.105	0.1180	-0.176	0.0250	-0.262	0.0014
21a			DRERI	Q6P041	-0.129	0.0997	0.105	0.1180	-0.176	0.0250	-0.262	0.0014
21b			DRERI	F1QSP4	-0.129	0.0997	0.105	0.1180	-0.176	0.0250	-0.262	0.0014
10	SRV_10181_at	O94888	HSAPI	O94888	-0.190	0.1364	-0.193	0.0988	-0.204	0.0857	-0.265	0.0361
10a			DRERI	Q6P3G3	-0.190	0.1364	-0.193	0.0988	-0.204	0.0857	-0.265	0.0361
48	SRV_02627_at	Q6UXB3	HSAPI	P55000	-0.022	0.4883	-0.184	0.0642	-0.250	0.0204	-0.280	0.0099
48a			HSAPI	Q6UXB3	-0.022	0.4883	-0.184	0.0642	-0.250	0.0204	-0.280	0.0099

Table 5. *Gene Expression Fold Changes for Table 2 Genes*

	Affymetric ID	Matched UniProt ID	Ortholog Organism	Ortholog UniProt ID	1 dpa Fold change	1 dpa adjusted p-val	3 dpa Fold change	3 dpa adjusted p-val	5 dpa Fold change	5 dpa adjusted p-val	7 dpa Fold change	7 dpa adjusted p-val
4	SRV_01453_a_at	O00483	HSAPI	O00483	-0.199	0.0000	-0.203	0.0000	-0.214	0.0000	-0.281	0.0000
4a			HSAPI	G3V560	-0.199	0.0000	-0.203	0.0000	-0.214	0.0000	-0.281	0.0000
4b			DRERI	Q6PBH5	-0.199	0.0000	-0.203	0.0000	-0.214	0.0000	-0.281	0.0000
4c			DRERI	E7F8X1	-0.199	0.0000	-0.203	0.0000	-0.214	0.0000	-0.281	0.0000
4d			DRERI	E7F4U9	-0.199	0.0000	-0.203	0.0000	-0.214	0.0000	-0.281	0.0000
75	SRV_03417_at	Q9Y2Q9	HSAPI	Q9Y2Q9	-0.264	0.0005	-0.331	0.0000	-0.308	0.0000	-0.333	0.0000
75a			DRERI	A5PMP2	-0.264	0.0005	-0.331	0.0000	-0.308	0.0000	-0.333	0.0000
47	SRV_10473_at	Q6UW78	HSAPI	Q6UW78	-0.361	0.0000	-0.346	0.0000	-0.338	0.0000	-0.345	0.0000
47a			DRERI	B3DFP2	-0.361	0.0000	-0.346	0.0000	-0.338	0.0000	-0.345	0.0000
78	SRV_07368_at	Q9Y6R7	HSAPI	Q9Y6R7	0.813	0.1238	-0.352	0.2704	0.022	0.4340	-0.346	0.2470
78a			DRERI	E7F2A5	0.813	0.1238	-0.352	0.2704	0.022	0.4340	-0.346	0.2470
78b			DRERI	F1RDU8	0.813	0.1238	-0.352	0.2704	0.022	0.4340	-0.346	0.2470
41	SRV_02403_at	Q16718			-0.256	0.0002	-0.354	0.0000	-0.355	0.0000	-0.368	0.0000
66	SRV_04332_at	Q9P0U1	HSAPI	Q9P0U1	-0.191	0.0006	-0.278	0.0000	-0.352	0.0000	-0.399	0.0000
66a			DRERI	Q0P4E8	-0.191	0.0006	-0.278	0.0000	-0.352	0.0000	-0.399	0.0000
66b					-0.191	0.0006	-0.278	0.0000	-0.352	0.0000	-0.399	0.0000
29	SRV_00312_a_at	P60174	HSAPI	P60174	-0.253	0.0015	-0.135	0.0417	-0.335	0.0000	-0.405	0.0000
29a			DRERI	Q1MTI4	-0.253	0.0015	-0.135	0.0417	-0.335	0.0000	-0.405	0.0000
29b			DRERI	E9QBF0	-0.253	0.0015	-0.135	0.0417	-0.335	0.0000	-0.405	0.0000

Table 5. *Gene Expression Fold Changes for Table 2 Genes*

	Affymetric ID	Matched UniProt ID	Ortholog Organism	Ortholog UniProt ID	1 dpa Fold change	1 dpa adjusted p-val	3 dpa Fold change	3 dpa adjusted p-val	5 dpa Fold change	5 dpa adjusted p-val	7 dpa Fold change	7 dpa adjusted p-val
18	SRV_01625_at	P20472	HSAPI	P20472	0.076	0.3969	-0.212	0.1227	-0.246	0.0883	-0.414	0.0114
18a			DRERI	Q9I8V0	0.076	0.3969	-0.212	0.1227	-0.246	0.0883	-0.414	0.0114
18b			DRERI	Q804W1	0.076	0.3969	-0.212	0.1227	-0.246	0.0883	-0.414	0.0114
18c			DRERI	Q7ZT36	0.076	0.3969	-0.212	0.1227	-0.246	0.0883	-0.414	0.0114
18d			DRERI	Q6IMW7	0.076	0.3969	-0.212	0.1227	-0.246	0.0883	-0.414	0.0114
18e			DRERI	Q804W2	0.076	0.3969	-0.212	0.1227	-0.246	0.0883	-0.414	0.0114
18f			DRERI	Q804W0	0.076	0.3969	-0.212	0.1227	-0.246	0.0883	-0.414	0.0114
14	SRV_00134_a_at	P04424	HSAPI	P04424	-0.269	0.0008	-0.329	0.0001	-0.343	0.0000	-0.442	0.0000
14a			DRERI	E9QEZ3	-0.269	0.0008	-0.329	0.0001	-0.343	0.0000	-0.442	0.0000
14b					-0.269	0.0008	-0.329	0.0001	-0.343	0.0000	-0.442	0.0000
6	SRV_01459_at	O43181			-0.355	0.0000	-0.456	0.0000	-0.476	0.0000	-0.500	0.0000
7	SRV_01455_at	O43676			-0.355	0.0000	-0.456	0.0000	-0.476	0.0000	-0.500	0.0000
54	SRV_04640_at	Q8N6Y1	HSAPI	Q8N6Y1	-0.275	0.0213	-0.459	0.0003	-0.468	0.0001	-0.530	0.0000
54a			DRERI	E7F9B5	-0.275	0.0213	-0.459	0.0003	-0.468	0.0001	-0.530	0.0000
54b					-0.275	0.0213	-0.459	0.0003	-0.468	0.0001	-0.530	0.0000
42	SRV_10075_at	Q2M385	HSAPI	Q2M385	1.179	0.0957	-0.614	0.2165	0.130	0.3960	-0.539	0.2184
42a			DRERI	A9C3Q1	1.179	0.0957	-0.614	0.2165	0.130	0.3960	-0.539	0.2184
42b			DRERI	B0R0K7	1.179	0.0957	-0.614	0.2165	0.130	0.3960	-0.539	0.2184
42c			DRERI	B0R063	1.179	0.0957	-0.614	0.2165	0.130	0.3960	-0.539	0.2184

Table 5. *Gene Expression Fold Changes for Table 2 Genes*

	Affymetric ID	Matched UniProt ID	Ortholog Organism	Ortholog UniProt ID	1 dpa Fold change	1 dpa adjusted p-val	3 dpa Fold change	3 dpa adjusted p-val	5 dpa Fold change	5 dpa adjusted p-val	7 dpa Fold change	7 dpa adjusted p-val
77	SRV_05077_at	Q9Y6Q3	HSAPI	Q9Y6Q3	-0.630	0.0000	-0.983	0.0000	-0.631	0.0000	-0.564	0.0000
77a					-0.630	0.0000	-0.983	0.0000	-0.631	0.0000	-0.564	0.0000
77b					-0.630	0.0000	-0.983	0.0000	-0.631	0.0000	-0.564	0.0000
69	SRV_04292_at	Q9UI09			-0.322	0.0005	-0.517	0.0000	-0.488	0.0000	-0.589	0.0000
46	SRV_05237_at	Q5VWZ2	HSAPI	Q5VWZ2	-0.239	0.0028	-0.691	0.0000	-0.616	0.0000	-0.666	0.0000
46a			DRERI	F1QBS9	-0.239	0.0028	-0.691	0.0000	-0.616	0.0000	-0.666	0.0000
62	SRV_05056_at	Q9H4I9			-0.348	0.0001	-0.430	0.0000	-0.524	0.0000	-0.676	0.0000
59	SRV_03913_at	Q9BPX1	HSAPI	Q9BPX1	-0.479	0.0007	-0.837	0.0000	-0.778	0.0000	-0.790	0.0000
59a			DRERI	Q6DEH9	-0.479	0.0007	-0.837	0.0000	-0.778	0.0000	-0.790	0.0000
59b					-0.479	0.0007	-0.837	0.0000	-0.778	0.0000	-0.790	0.0000
22	SRV_00743_at	P35611			-0.528	0.0002	-0.509	0.0002	-0.629	0.0000	-0.833	0.0000
40	SRV_05198_a_at	Q15633			-0.593	0.0000	-0.656	0.0000	-0.828	0.0000	-0.958	0.0000
52	SRV_04047_at	Q7Z4S6	HSAPI	Q7Z4S6	-0.775	0.0000	-1.178	0.0000	-1.325	0.0000	-1.685	0.0000
52a			DRERI	F1QWX6	-0.775	0.0000	-1.178	0.0000	-1.325	0.0000	-1.685	0.0000
52b			DRERI	F8W3W5	-0.775	0.0000	-1.178	0.0000	-1.325	0.0000	-1.685	0.0000
13	SRV_00148_at	P01034	HSAPI	P01034								
13a			HSAPI	P01036								
13b			HSAPI	P28325								
13c			HSAPI	P09228								

Table 5. *Gene Expression Fold Changes for Table 2 Genes*

	Affymetric ID	Matched UniProt ID	Ortholog Organism	Ortholog UniProt ID	1 dpa Fold change	1 dpa adjusted p-val	3 dpa Fold change	3 dpa adjusted p-val	5 dpa Fold change	5 dpa adjusted p-val	7 dpa Fold change	7 dpa adjusted p-val
13d			HSAPI	P01037								
13e			DRERI	B8A4D0								

Table 5 also gives adjusted p-values for all fold change values, which gives the data some statistical test significance. A p-value can essentially be thought of as a statistic measuring the incidence of false positives of a test labeled as statistically significant, which could also be described as a rate of false discovery.[42] Adjusted p-values were calculated from p-value data in order provide a more accurate metric for statistical testing with multiple comparisons.[43]

The primary genes of interest from Table 5 are those with fold changes $\geq \pm 1.5$ and adjusted p-values < 0.05. Gene P16112 (Aggrecan core protein) shows fold changes of greater than 4.6 for all time periods (1, 3, 5, 7 dpa) and therefore remains highly expressed in the early stages of response to an incision or amputation and limb regeneration; this gene is a cartilage-specific proteoglycan that likely plays a major role in regeneration with regards to building collagen-based tissues. Genes P43235 (Cathepsin K; CTSK) and Q15004 (PCNA-associated factor) both show initial downregulation for the 1 dpa time period (although the adjusted p-value of P43235 at 1 dpa is > 0.05) followed by upregulation of fold changes over 1.5 for the 5 dpa and 7 dpa time periods. This is indicative of genes that can be said to "turn down" early in the wound response process, only to be "turned back up" when their functions are deemed necessary by the cell. As a final example, gene Q7Z4S6 (probable squalene synthase) shows a trend in fold changes, from -0.775 at 1 dpa to -1.178 at 3 dpa to -1.325 at 5 dpa to -1.685 at 7 dpa. Squalene synthase takes part in the isoprenoid biosynthetic pathway, catalyzing the first stage of a multi-step reaction that eventually works to produce sterols, or steroid alcohols, from squalene.[44] The corresponding homologous zebrafish genes F1QWX6 and

F8W3W5 are both uncharacterized proteins, so further research into these genes may provide useful information into a possible connection between the functionalities of these genes and how they could potentially relate to the regenerative process.

Comparisons of physical sequence similarities, as opposed to simply homologous function, can also be made using control-vs-TCDD zebrafish gene data and predetermined annotation information within that data regarding sequence structure similarities to corresponding human gene sequences, which can be used in correspondence with Tables 2 and 5 to assess which homologous sets of genes may be of further interest. The result of such a comparison is shown here as Table 6, where homologous gene data is given in addition to fold change and adjusted p-value data for comparisons between control zebrafish and zebrafish exposed to TCDD for the values at 1, 3, and 5 dpa for both subsets. Positive fold change values represent cases in which genes exhibited higher amounts of expression in the control zebrafish for a given dpa time period as compared to the expression of that gene for the TCDD-dosed zebrafish, and vice versa. Another way to explain this concept is to assume all positive fold change values in Table 6 represent upregulation of a given gene in a control organism relative to the one dosed with TCDD. Similarly, all negative fold change values can be viewed as a downregulation of a given gene in a control organism relative to one dosed with TCDD. Greater fold change values imply that the presence of TCDD results in the downregulation of a gene, and lesser fold change values imply that the presence of TCDD results in the upregulation of a gene. Graphical representations of the fold change data given in Table 6 (page 79) can be seen in Figures 4, 5, and 6 (pages 80-82).

Table 6. *Fold Change Values for Matches between Zebrafish CTRL-vs-TCDD Data at 1, 3, 5 dpa and Table 2*

Zebrafish UniProt ID	Extent of Sequence Similarity	Human UniProt ID	CTRL vs TCDD comparison	Fold change	adjusted p-val
Q502E4	Moderate	O43181	1 dpa - 1 dpa	0.243	0.7361
			3 dpa - 3 dpa	0.119	0.8134
			5 dpa - 5 dpa	0.227	0.4263
Q4VBU7	Moderate	P20674	1 dpa - 1 dpa	0.190	0.5506
			3 dpa - 3 dpa	0.108	0.6538
			5 dpa - 5 dpa	0.128	0.3876
Q0P4E8	Moderate	Q9P0U1	1 dpa - 1 dpa	0.345	0.4235
			3 dpa - 3 dpa	0.119	0.7525
			5 dpa - 5 dpa	0.078	0.7227
Q7SY44	Weak	P04424	1 dpa - 1 dpa	-0.114	0.9302
			3 dpa - 3 dpa	0.724	0.1233
			5 dpa - 5 dpa	0.442	0.2671
Q75T39	Weak	P16112	1 dpa - 1 dpa	0.162	0.9591
			3 dpa - 3 dpa	-2.817	0.0400
			5 dpa - 5 dpa	-1.793	0.1301
Q5BJA2	Weak	Q16718	1 dpa - 1 dpa	0.152	0.8891
			3 dpa - 3 dpa	0.203	0.7089
			5 dpa - 5 dpa	0.129	0.6853
Q566P2	Weak	Q30201	1 dpa - 1 dpa	0.137	0.9304
			3 dpa - 3 dpa	-0.272	0.7319
			5 dpa - 5 dpa	0.302	0.5236
A2CEX8	Weak	Q9UDV7	1 dpa - 1 dpa	-0.054	0.9446
			3 dpa - 3 dpa	0.139	0.7525
			5 dpa - 5 dpa	0.146	0.5729
Q502D9	Weak	Q9ULW0	1 dpa - 1 dpa	-0.826	0.5967
			3 dpa - 3 dpa	0.068	0.9497
			5 dpa - 5 dpa	0.379	0.5816

Fig. 4. *Graphical Representation of Table 6 – Genes Q502E4, Q4VBU7, and Q0P4E8*

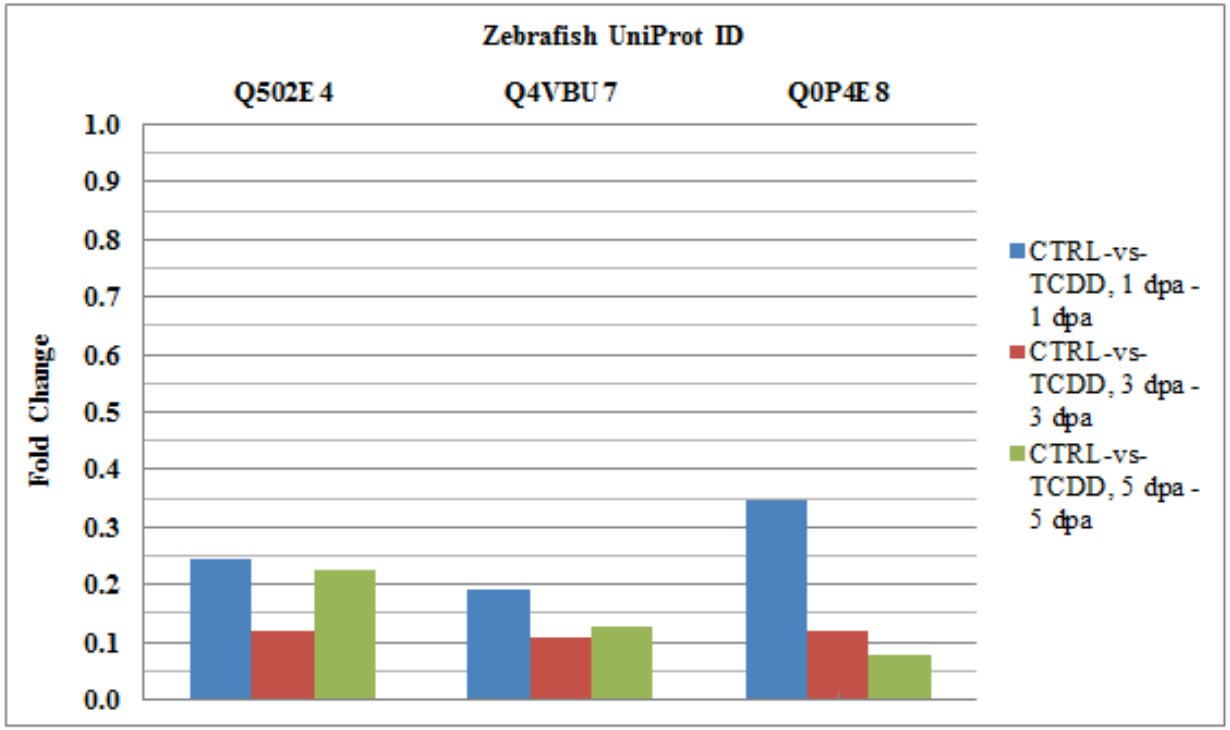

Fig. 5. *Graphical Representation of Table 6 – Genes Q7SY44, Q75T39, and Q5BJA2*

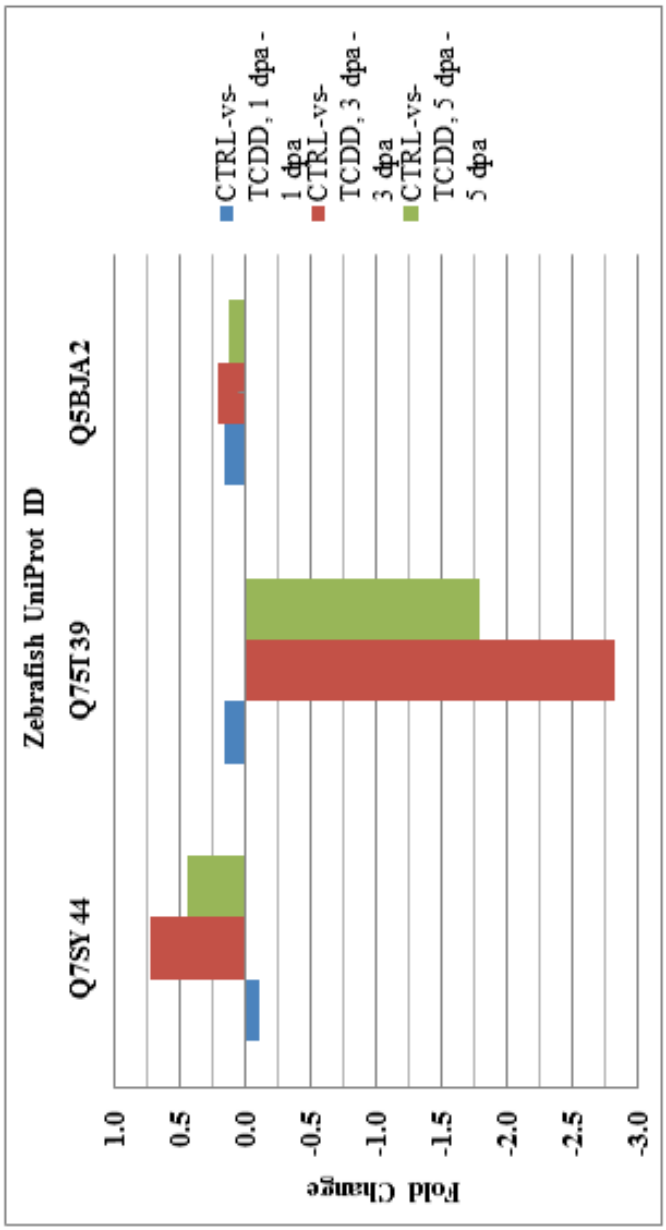

Fig. 6. *Graphical Representation of Table 6 – Genes Q566P2, A2CEX8, and Q502D9*

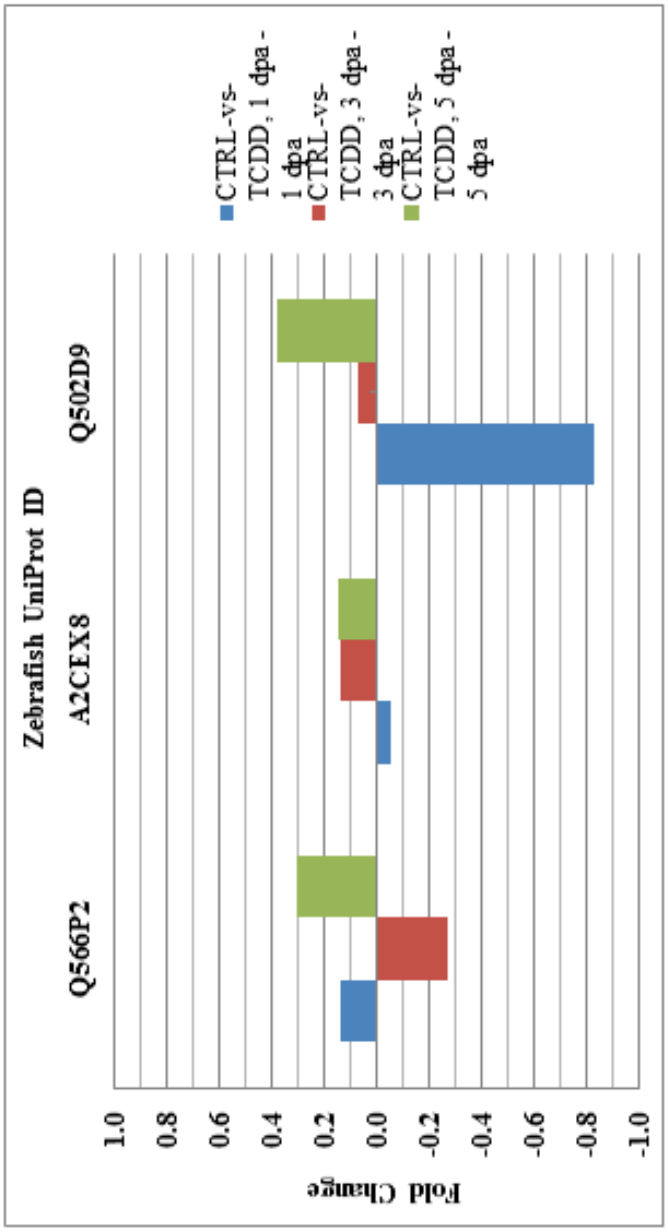

Table 6 shows that of the 9 zebrafish genes in the zebrafish CTRL-vs-TCDD data, 3 showed moderate sequence similarity and 6 showed weak sequence similarity. It should be noted that the majority of p-values given for the fold change values in Table 6 are not < 0.05, so some degree of caution should be exerted when taking these fold change values at face value.

Table 6 and Figure 4 show that zebrafish genes Q502E4, Q4VBU7, and Q0P4E8 are not significantly altered with regards to comparing gene expression from a control organism to a TCDD-dosed organism; the values are all positive, however, so the presence of TCDD likely caused these genes to become downregulated because, again, a positive fold change signifies greater expression of a gene in the control animal as compared to the TCDD-dosed case.

Table 6 and Figure 5, however show vastly different results. Zebrafish gene Q7SY44 shows an insignificant amount of upregulation in the presence of TCDD at 1 dpa followed by a quick transition to a markedly more downregulated state at 3 dpa when compared to the control organism. Zebrafish gene Q75T39 shows an insignificant amount of downregulation at 1 dpa followed by transitions to states of nearly 3 fold and 2 fold upregulation at 3 dpa and 5 dpa, respectively, in the presence of TCDD when compared to the control organism. The implication of this behavior is that the presence of TCDD causes gene Q75T39 to become significantly more expressed in the zebrafish during regeneration. Gene Q75T39 codes for a neurocan protein, a chondroitin sulfate proteoglycan considered to be related to migration and cell adhesion modulation.

Neurocan is also thought to play a role in the development of neurite growth. Furthermore, it is purely of interest to note that neurocan is also a component of the ECM in the brain.[45] Zebrafish gene Q5BJA2 exhibits no drastic differences in fold change values.

Table 6 and Figure 6 show that zebrafish gene Q566P2 exhibits small amounts of relative down-, up-, and downregulation at 1, 3, and 5 dpa, respectively, in the presence of TCDD when compared to the control organism. Zebrafish gene A2CEX8 exhibits no drastic differences in fold change values. Zebrafish gene Q502D9 shows that a noticeable upregulation of gene expression at 1 dpa is followed by eventual downregulation at 5 dpa in the presence of TCDD when compared to the control organism.

5. Summary

Various databases and online resources were accessed in order to obtain gene expression and annotation data for axolotl and the zebrafish. Scripts were written using Perl software and programming language in order to find 1:1 matches between gene sequences obtained from sample organisms during regeneration and known gene sequences from previously annotated database sources.

A table was constructed to show a Rosetta stone of 78 axolotl genes, matched to human genes, with homologous zebrafish gene information, including gene ontology annotation. Common gene ontologies from this table were counted and structured into

another table, and aspects of these two tables were discussed in detail. 20 found matches between the axolotl data and a database of human ECM genes were listed. Fold change data, both for the axolotl-zebrafish homologous genes as expressed in a previous experiment regarding gene expression at various times dpa as well as for zebrafish control-vs-TCDD gene expression data for different dpa were given in two additional other tables and discussed in some detail. Throughout, genes of interest were identified and examined.

Protein sequence topics discussed include 40S and 60S ribosomal proteins, NADH ubiquinone oxidoreductase, histones, parvalbumin, cystatin, cathepsin, squalene synthase, and neurocan. Gene ontologies and functions discussed include cell differentiation, protein synthesis via translation, mitochondrial and ribosomal operations, collagen, ECM, apoptosis, and neurite development.

6. Closing Remarks

There are many implications of the sheer breadth and depth of the research currently being devoted to the fields of genomics and bioinformatics, as well as to research projects designed to determine how altering genes of interest can be carried out in order to affect positive changes for the betterment of society. Genetically modified agriculture is currently a booming industry with the potential to help assuage food shortages crises that may be encountered in the near future, if they have not already occurred, but this means that society will soon be grappling with questions regarding the

ethical nature of altering an organisms genes. In some cases, the answer may be simple: if future biomedical and gene therapy technologies can help an individual regenerate skin tissue after a life-threatening burn from a house fire or a soldier partially regenerate an appendage that was severely damaged in combat without using an especially invasive procedure, than questions regarding these concerns about gene therapies capable of helping these people may be limited. But what if new gene therapies can help drastically extend the lives of people past our current and "normal" life expectancies? At what point would genetic manipulation cease to be about finding treatments for unfortunate events and begin to represent ways to bend the governing rules of mother nature and the human body to man's will?

The fact of the matter, however, is that the fields of genomics and bioinformatics have come a long way in the last decade, and the gene expression, annotation, and ontological data being compiled by researchers around the world are adding to the knowledge banks of the global genomics community by the day, and this information is immeasurably helpful in allowing scientists to determine what genes cause what chemical and physical changes to occur in a wide range of organisms.

And even if this research does not eventually lead to the ability of humans to alter our genetic makeup so as to enable limb regeneration, maybe it can be used to help find newer, preventative treatment options for life-altering diseases and disorders, maybe even before such maladies arise.

7. Acknowledgments

I would like to thank my advisor, Keith Hutchison for his unbelievable amounts of patience and valuable guidance, as well as my co-advisor, Benjamin King, for painstakingly helping to make sure I am able to run Perl programming scripts effectively. Both individuals have given me their time and professional knowledge, and my gratitude for all that they have done cannot be adequately expressed in words, so I sincerely hope this will suffice.

I would also like to thank Maine INBRE (IDeA Network of Biomedical Research Excellence) for awarding the INBRE Functional Genomics Thesis Fellowship for this thesis.

8. References

1 Wang, Z., Gerstein, M., & Snyder, M., "RNA-Seq: a revolutionary tool for transcriptomics." Nature Rev. Genet. 10, 57–63 (2009).

2 Michalopoulos, George K., and Marie C. DeFrances. "Liver regeneration." *Science* 276.5309 (1997): 60-66.

3 Levesque, M., Villiard, E., Roy, S., 2010. "Skin wound healing in axolotls: a scarless process." J. Exp. Zool. (Mol. Dev. Evol.) 314B:684-697.

4 Tal, T.L., Franzosa, J.A., & Tanguay, R.L. "Molecular signaling networks that choreograph epimorphic fin regeneration in zebrafish – a mini-review." Gerontology 2010:56:231-240.

5 Majchrzak, A. "Ambystoma mexicanum." Animal Diversity Web. University of Michigan, Museum of Zoology, 2004. Web. Accessed 2 April 2013 at http://animaldiversity.ummz.umich.edu/accounts/Ambystoma_mexicanum/

6 Markowski, D. "Danio rerio." Animal Diversity Web. University of Michigan, Museum of Zoology, 2011. Web. Accessed 2 April 2013 at http://animaldiversity.ummz.umich.edu/accounts/Danio_rerio/

7 Monaghan, J., Walker, J., Page, R., Putta, S., Beachy, C., Voss, S., 2007. "Early gene expression during natural spinal cord regeneration in the salamander Ambystoma mexicanum." Journal of Neurochemistry, v. 101, p.27-40.

8 "A Basic Introduction to the Science Underlying NCBI Resources." NCBI. U.S. National Library of Medicine, n.d. 10 Aug. 2012. Web. Accessed 7 April 2013 at http://www.ncbi.nlm.nih.gov/About/primer/microarrays.html

9 Prasch, A. L.; Tanguay, R. L.; Mehta, V.; Heideman, w.; Peterson, R. E., "Identification of Zebrafish ARNT1 homologs: 2,3,7,8-tetrachlorodibenzo-p-dioxin toxicity in the developing zebrafish requires ARNT1." The American Society for Pharmacology and Experimental Therapeutics: Molecular Pharmacology, 2006; Vol. 69, pp 776-787.

10 Prasch, A. L.; Heideman, W.; Peterson, R. E., "ARNT2 is not required for TCDD Developmental Toxicity in Zebrafish." Society of Toxicology: Toxicological Sciences, 2004; Vol. 82, pp 250-258.

11 Zodrow, Jeanmarie M., and Robert L. Tanguay. "2, 3, 7, 8-tetrachlorodibenzo-p-dioxin inhibits zebrafish caudal fin regeneration." Toxicological Sciences 76.1 (2003): 151-161.

12 Matthew, L. K.; Sengupta, S.; Kawakami, A.; Andreasen, E. A.; Löhr, C. V.; Loynes, C. A.; Renshaw, S. A.; Peterson, R. T.; Tanguay, R. T., "Unraveling Tissue Regeneration Pathways Using Chemical Genetics." Journal of Biological Chemistry, 2007; Vol. 282, pp 35202-35210.

13 "Inflammation." Medical Dictionary. MediLexicon, 2006. Web. Accessed 12 April 2013 at http://www.medilexicon.com/medicaldictionary.php

14 Gurtner, Geoffrey C., et al. "Wound repair and regeneration." *Nature* 453.7193 (2008): 314-321.

15 Brockes JP, Kumar A: "Plasticity and reprogramming of differentiated cells in amphibian regeneration." Nat Rev Mol Cell Biol 2002, 3:566-574.

16 Rao, Nandini, et al. "Proteomic analysis of blastema formation in regenerating axolotl limbs." BMC biology 7.1 (2009): 83.

17 Alberts, B.; Johnson, A.; Lewis, J.; et al. "The Extracellular Matrix of Animals." Molecular Biology of the Cell. 4th edition. New York: Garland Science; 2002. Web. Accessed 12 April 2013 at http://www.ncbi.nlm.nih.gov/books/NBK26810/

18 King, M. "Extracellular matrix." The Medical Biochemistry Page, 8 February 2013. Web. Accessed 12 April 2013 at http://themedicalbiochemistrypage.org/extracellularmatrix.php

19 "How Affymetrix GeneChip DNA Microarrays Work". Affymetrix, 2013. Web. Accessed 12 April 2013 at http://public.tgen.org/tgen.org/downloads/autism/Genotypingessentials.pdf

20 Sadava, et al., "High-Throughput Sequencing. Life: The Science of Biology, Ninth Edition. 2010. Web. Accessed 13 April 2013 at http://www.sumanasinc.com/webcontent/animations/content/highthroughput.html

21 Bentley, D. R.; Balasubramanian, S.; Swerdlow, H. P.; Smith, G. P.; Milton, J.; Brown, C. G.; Hall, K. P.; Evers, D. J. et al. (2008). "Accurate whole human

genome sequencing using reversible terminator chemistry". Nature 456 (7218): 53–59.

22 "Clonal amplification." Second-generation sequencing - Amplification. Molbiol.ru, 12 December 2008. Web. Accessed 13 April 2013 at http://seq.molbiol.ru/sch_clon_ampl.html

23 Pearson, William R., and David J. Lipman. "Improved tools for biological sequence comparison." Proceedings of the National Academy of Sciences 85.8 (1988): 2444-2448.

24 "Pairwise local/global alignment - Introduction." 2Can Support Portal - Protein and Proteomic Analysis. European Molecular Biology Laboratory-EBI, 2013. Web. Accessed 13 April 2013 at http://www.ebi.ac.uk/2can/tutorials/protein/align.html

25 "BLAST Software." Basic Local Alignment Search Tool. 2013. Web. Accessed October 2012 at http://blast.ncbi.nlm.nih.gov/

26 "Dataset Homepage." BioMart. Ensembl Genome Browser. 2013. Web. Accessed Dec 2012 at http://www.ensembl.org/biomart/martview

27 "Homepage." Gene Expression Omnibus (GEO) DataSet Brower. NCBI, 2013. Web. Accessed Dec 2012 at http://www.ncbi.nlm.nih.gov/sites/GDSbrowser/

28 "ActivePerl." ActiveState. ActiveState Software Inc., 2013. Web. Accessed Oct 2012 at http://www.activestate.com/activeperl

29 "Query Page." UniProtKB. UniProt, 2013. Web. Accessed Nov 2012 at http://www.uniprot.org/help/uniprotkb

30 "The Hierarchical Catalog of Orthologs." OrthoDB. Université de Genève, Department of Microbiology and Molecular Medicine, 2012. Web. Accessed April 7 2012 at http://cegg.unige.ch/orthodb6

31 Pestova, Tatyana V., et al. “Molecular mechanisms of translation initiation in eukaryotes.” Proceedings of the National Academy of Sciences 98.13 (2001): 7029-7036.

32 Twiss, Jeffery L., et al. “Translational control of ribosomal protein L4 mRNA is required for rapid neurite regeneration.” Neurobiology of disease 7.4 (2000): 416-428.

33 Nakamaru-Ogiso, Eiko, et al. “The ND2 subunit is labeled by a photoaffinity analogue of asimicin, a potent complex I inhibitor.” FEBS letters 584.5 (2010): 883-888.

34 “Pumping protons: a complex problem.” Protein Data Bank. European Molecular Biology Laboratory-EBI, 2013. Web. Accessed April 14 2013 at http://www.ebi.ac.uk/pdbe-apps/quips?story=CXI

35 Chomova, M., and P. Racay. “Mitochondrial complex I in the network of known and unknown facts.” General physiology and biophysics 29.1 (2010): 3.

36 Brockes, Jeremy P. “Amphibian limb regeneration: rebuilding a complex structure.” Science 276.5309 (1997): 81-87.

37 Xu, Jun, et al. “Extracellular histones are major mediators of death in sepsis.” Nature medicine 15.11 (2009): 1318-1321.

38 Cates, M. Susan, et al. “Metal-ion affinity and specificity in EF-hand proteins: coordination geometry and domain plasticity in parvalbumin.” *Structure* 7.10 (1999): 1269-1278.

39 Zhou, Xiang, et al. “Interferon Induced IFIT Family Genes in Host Antiviral Defense.” International journal of biological sciences 9.2 (2013): 200.

40 Chwieralski, C. E., T. Welte, and F. Bühling. “Cathepsin-regulated apoptosis.” Apoptosis 11.2 (2006): 143-149.

41 Berrier, Allison L., and Kenneth M. Yamada. "Cell–matrix adhesion." Journal of cellular physiology 213.3 (2007): 565-573.

42 Dabney, A.; Storey, J. D., “q-value: Q-value estimation for false discovery rate control.” Warnes, G. R., Ed. 2011; Vol. R package version 1.26.0.

43 "Adjusted P values as part of multiple comparisons." GraphPad Software, Inc.: 2002-2012.

44 Tansey, Terese R., and Ishaiahu Shechter. "Structure and regulation of mammalian squalene synthase." Biochimica et biophysica acta 1529.1-3 (2000): 49.

45 Rauch, Uwe, et al. "Cloning and primary structure of neurocan, a developmentally regulated, aggregating chondroitin sulfate proteoglycan of brain." Journal of Biological Chemistry 267.27 (1992): 19536-19547.

9. Appendix A – Perl Scripts

Perl Script – “blast_parser_combo3b”

“

```
#!/usr/bin/perl
print " \n";
print "% IDENTITY? (input 0-100) : ";
$PERCENT_ID=<STDIN>;
print "% LENGTH? (input 0-100) : ";
$COVERAGE=<STDIN> / 100;
print "\n";


$target_lengths_fh = open(TLENGTHS, "<Amby_001a520165F_targets.fa_lengths.txt");
$assembly_lengths_fh = open(ALENGTHS, "<axolotl_assembly.fa_lengths.txt");
$annotate1_fh = open(ANNOTATE1, "<axolotl_blastx_1e-.txt");
$output = open(OUTPUT, ">Table_2_output.txt");


$input_fh1 = open(INPUT1,"<targets_vs_assembly_blastn.txt");

while (<ANNOTATE1>) {
        $line1 = $_;
        chomp($line1);

        @fields = split("\t",$line1);
        $a1_contig = $fields[0];
```

```
        $a1_locus = $fields[1];
        $a1_hit = $fields[2];
        $a1_pvalue = $fields[3];
        $a1_match = $fields[5];

        $proteinnames{"Consensusfrom" . $a1_contig} = $a1_locus;
        $proteinnames2{"Consensusfrom" . $a1_contig} = $a1_match;

            #print $line1,"\t";
}


while (<TLENGTHS>) {
        $line1 = $_;
        chomp($line1); # removes new line '\n' character at end of symbol

        @fields = split("\t",$line1);
        $id1 = $fields[0];
        $length1 = $fields[1];

        $target_lengths{$id1} = $length1;

        #print $id1,"\t",$length1,"\n";

}
```

```
while (<INPUT1>) {

        $line1 = $_;
        chomp($line1);

        @fields = split("\t",$line1);

        $query_id1 = $fields[0];
        $subject_id1 = $fields[1];
        $percent_id1 = $fields[2];
        $align_length1 = $fields[3];
        $mismatches1 = $fields[4];
        $gaps1 = $fields[5];

        $identity_i1{$query_id1} = $percent_id1;
        $length_i1a{$query_id1} = $align_length1;

        #print $query_id1,"\t",$percent_id1,"\n";


        if ($target_lengths{$query_id1} > 0) {
              if ($percent_id1 >= $PERCENT_ID) {
                    if (($align_length1/$target_lengths{$query_id1}) >=
$COVERAGE) {
                          if ($query_to_subject1{$query_id1}) {
                          $query_to_subject1{$query_id1} =
$query_to_subject1{$query_id1} . "," . $subject_id;
                          }
                          else {
```

```
                                $query_to_subject1{$query_id1} = $subject_id1;
                                }
                        }
                }
        }

  $gene_symbols_to_chr1{$symbol1} = $chr1;

}


$num_targets_w_mult_hits = 0;
$num_targets_w_one_hit = 0;

foreach $i1 (keys %query_to_subject1) {

        #print $i1," hits = ",$query_to_subject1{$i1},"\n";
        if ($query_to_subject1{$i1} =~/,/) {
           $num_targets_w_mult_hits1 = $num_targets_w_mult_hits1 + 1;
        }
        else {
              $num_targets_w_one_hit1 = $num_targets_w_one_hit1 + 1;
        }
}


$input_fh2 = open(INPUT2,"<assembly_vs_targets_blastn.txt");
```

```
while (<ALENGTHS>) {
        $line2 = $_;
        chomp($line2); # removes new line '\n' character at end of symbol

        @fields = split("\t",$line2);
        $id2 = $fields[0];
        $length2 = $fields[1];

        $assembly_lengths{$id2} = $length2;
}
while (<INPUT2>) {

        $line2 = $_;
        chomp($line2);

        @fields = split("\t",$line2);

        $query_id2 = $fields[0];
        $subject_id2 = $fields[1];
        $percent_id2 = $fields[2];
        $align_length2 = $fields[3];
        $mismatches2 = $fields[4];
        $gaps2 = $fields[5];

        $identity_i2{$query_id2} = $percent_id2;

        if ($assembly_lengths{$query_id2} > 0) {
                if ($percent_id2 >= $PERCENT_ID) {
```

```
                        if (($align_length2/$assembly_lengths{$query_id2}) >=
$COVERAGE) {
                                if ($query_to_subject2{$query_id2}) {
                                $query_to_subject2{$query_id2} =
$query_to_subject2{$query_id2} . "," . $subject_id;
                                }
                                else {
                                $query_to_subject2{$query_id2} = $subject_id2;
                                }
                        }
                }
        }

$gene_symbols_to_chr2{$symbol2} = $chr2;

}


$num_contigs_w_mult_hits = 0;
$num_contigs_w_one_hit = 0;

foreach $i (keys %query_to_subject2) {

        if ($query_to_subject2{$i} =~/,/) {
           $num_contigs_w_mult_hits2 = $num_contigs_w_mult_hits2 + 1;
        }
        else {
              $num_contigs_w_one_hit2 = $num_contigs_w_one_hit2 + 1;
        }
```

```
}


$num_1to1 = 0;


foreach $target (keys %query_to_subject1) {
        if ($query_to_subject1{$target} !~ /,/) {
                foreach $subject (keys %query_to_subject2) {
                        if ($query_to_subject2{$subject} !~ /,/) {
                                if (($query_to_subject1{$target} eq $subject) &&
($query_to_subject2{$subject} eq $target)) {


                print "-> Found 1:1 mapping:
",$target,"\t",$query_to_subject1{$target},"\n",$proteinnames{$query_to_subject1{$targ
et}},"\t",$proteinnames2{$query_to_subject1{$target}},"\t",$identity_i1{$target},"\t",$i
dentity_i2{$subject},"\t","\n";
                print OUTPUT "-> Found 1:1 mapping:
",$target,"\t",$query_to_subject1{$target},"\n",$proteinnames{$query_to_subject1{$targ
et}},"\t",$proteinnames2{$query_to_subject1{$target}},"\t",$identity_i1{$target},"\t",$i
dentity_i2{$subject},"\t","\n";


                $num_1to1 = $num_1to1 + 1;
                                }
                        }
                }
        }
}


print " \n";
print " For % IDENTITY = ",$PERCENT_ID," and % LENGTH   =
",$COVERAGE*100," ...","\n";
print " \n";
```

```
print "Number of targets that align to >= % identity and >= % alignment length","\n";
print "specified above...","\n";
print " \n";
print ">=1 contig = ",scalar(keys(%query_to_subject1)),"\n";
print "=1 contig = ",$num_targets_w_one_hit1,"\n";
print ">1 contig = ",$num_targets_w_mult_hits1,"\n";
print " \n";

print "Number of contigs that align to >= % identity and >= % alignment length","\n";
print "specified above...","\n";
print " \n";
print ">=1 target = ",scalar(keys(%query_to_subject2)),"\n";
print "=1 target = ",$num_contigs_w_one_hit2,"\n";
print ">1 target = ",$num_contigs_w_mult_hits2,"\n";
print " \n";

print "-> Number of 1:1 mappings = ",$num_1to1,"\n";
print " \n";
”
```

10. Author's Biography

Justin Paul Bolinger was born on January 1st, 1991, in Huntingdon, England. He was raised primarily in Gorham, Maine, where he attended Gorham High School and experienced many athletic and academic achievements. He accepted a University of Maine Pulp and Paper Foundation Scholarship and attended the University of Maine at Orono in the fall of 2009 as a major in Chemical Engineering. Justin is graduating in May 2013 with a Bachelor's of Science degree in Chemical Engineering and Dean's List recognition throughout his entire career at the University of Maine.

Outside of academia, Justin's hobbies include mathematics, golfing, reading, enjoying the outdoors, and being a devoted Philadelphia Eagles fan. His immediate family includes his mother and father, Martha Dow Bolinger and John William Bolinger, and his twin sister, Siobhan Lynn Bolinger, who also attends the University of Maine at Orono and is graduating in December 2013 with a Bachelor's of Science degree in Animal and Veterinary Sciences with a Pre-Vet concentration. Upon graduation, he intends to pursue a job in the fields of pulp and paper, chemicals, polymer science, genomics, or a field in which he can otherwise put his self-proclaimed "Renaissance man" skills to use to make the world a better place for those who inhabit it.