Spring 5-6-2014

Transcriptome Analysis of Sea Lamprey Embryogenesis

Zakary Ilya Yermolenko
zakary.yermolenko@student.shu.edu

Transcriptome Analysis of Sea Lamprey Embryogenesis

Zakary Ilya Yermolenko

Submitted in partial fulfillment of the requirements for the degree of Master of Science in Biology from the Department of Biological Sciences of Seton Hall University

May 2014

Acknowledgements

I would like to express my greatest appreciation to the following people:

Dr. Ziping Zhang, who worked with me throughout the entire project in helping achieve a tremendous goal that I would not have been able to complete otherwise with excellence and precision. His commitment helped me become a well-balanced research analyst in the field of Molecular Genetics. Having completed the project in less than one year showed true dedication on behalf of his encouragement and trust bestowed on myself, which I represented with the utmost level of assurance. I will continue this passion in my future endeavors as it has brought me success thus far in my time with Seton Hall University.

Maria Bender and Satshil Rana for their computational analysis and expertise in assisting with gene selection and expression calculation for embryological stages within various computer programs and operations. With their help, the project was completed at an accelerated rate that would not have been accomplished otherwise.

Frank J. Zadlock for his leadership and experience in the field of Next Generation Sequence Analysis and laboratory excellence. Mr. Zadlock helped with the computer science aspect as well as wet lab research procedures. He showed how to select a small pool of genes from thousands, which would be essential in this study using the latest programs analyzing Next Generation Sequencing. Additionally, Mr. Zadlock taught all the necessary aspects of wet-lab research for RNA isolation, RT-PCR, PCR, and Gel Electrophoresis analysis that were to be used for verifying the computer science results.

Dr. Ko, for her dedication and diligence in providing an outstanding choreographed work in writing the thesis. Dr. Ko has assisted with every detail within the thesis and I owe her a great gratitude for assisting me in completing and fulfilling my requirements as a Masters of Biology student at Seton Hall University.

Dr. Klaus and Dr. Cottrell who represent the committee that analyzed the project from a professional perspective in their fields of study to help guide me in further directions for future studies. Dr. Klaus excelled my studies in both embryogenesis and microscopy and Dr. Cottrell helped me analyze the quantitative Real Time results.

The Department of Biological Sciences Faculty and Staff, who sought to excel my full potential as a Graduate student in my pursuits of academic excellence over the past two years at Seton Hall University. Additionally, I would like to thank them for honoring me a full Scholarship in representing the department as a Teaching Assistant in lecturing laboratory science classes including Microbiology, Cell Biology, and General Biology.

Table of Contents

List of Figures

List of Tables and Images

Abstract

The sea lamprey (*Petromyzon marinus*) has survived throughout evolution for hundreds of millions of years. It is considered an invasive species to the Great Lakes that has caused dramatic changes in the ecosystem for fish communities resulting in the collapse of a fishing industry that was previously valued at billions of dollars. Successful management of the sea lamprey is essential to a sustainable fishing industry and biodiversity. Therefore sea lamprey embryos were studied at various stages of development by growing them in a simulated habitat. RNAs from adult female ovaries and embryos at different time points during embryogenesis were extracted and then subjected to Next Generation Sequencing (NGS). Approximately 14,000 transcripts in total were detected. Data from each stage was analyzed using $\log_2$ values to determine significant gene expression levels on FPKM values, which are fragments of transcripts derived from cDNA. Genes with greater than threefold changes from previous adjacent stages, were analyzed using Cytoscape, an open source network program to further verify their functional roles. Genes TGFBR1 (protein kinase), SMAD2 (signal transducer), TRAF6 (signal transducer), PCNA (DNA replication), SLX4 (endonuclease), and PXN (focal adhesion) were selected for their NGS expression levels, GO annotations, and biological function significance. Their NGS expression levels were verified using Real-Time PCR (qPCR). TGFBR1 was up-regulated in the Neurula/Egg and Gall Bladder/Eye Spot stage comparisons. PXN was up-regulated in the Neurula/Egg and Hatching/Head Protrusion comparisons. SMAD2 and TRAF6 were up-regulated in the Melanophore/Hatching, PCNA in the Completion/Gall Bladder, and SLX4 in the Eye Spot/Melanophore comparisons. This transcriptome analysis serves as a starting point for determining vital genes in sea lamprey embryogenesis and controlling the sea lamprey population.

Introduction

Sea lampreys have inhabited the Great Lakes since the beginning of the 20th century and expanded their population overtime into Lake Huron, Michigan, and Superior. It was soon determined the invasive species caused a traumatic disturbance in the ecosystem within the Great Lakes. Sea lampreys feed on their host's blood that leads to scarring and death. Each lamprey can kill around 40 or more pounds of native fish in the Great Lakes. Before their presence, fisheries harvested roughly 15 million pounds of lake trout in lakes Huron and Superior annually. In the early 1960s, the harvest was only around 300,000 pounds (Great Lakes Fishery Commission, 2011). Therefore, a research program was developed to control and eliminate sea lamprey. The Great Lakes Sea Lamprey committee formed in 1946 before growing into the Great Lakes Fishery Committee (GLFC) in the late 1950's. Controlling their outgrowth included building mechanical barriers and using toxic chemicals such as lampricides to destroy larval lamprey; the sea lampreys reached their peak population around 1960's. At first, these methods proved to be successful in controlling the population as evidenced by reduced lamprey spawning and decreasing traumas on lake trout (Smith et al, 1980). Although the population of the sea lamprey had decreased by 1970, lake trout restoration was not fully competed. Other invertebrates were affected by the lampricides, which contributed negatively to the fisheries, and not all the sea lampreys were eliminated because of limitations in funding associated with these projects (Smith et al, 1974).

Investigations in the embryogenesis of the sea lamprey were studied during the same time periods. Spawning begins around spring to early summer when temperatures reach 50°F, which is when the female lamprey releases her eggs (around 68,000) where there is rapid current. Eggs are hatched in 10 to 13 days and the larvae remain nested for 18 to 21 days until they are carried

downstream until the current reduced and they burrow themselves into the soft bottom, which is composed of mud and silt (Stuaffer and Hansen, 1958).

Ammocoetes or larvae of sea lamprey spend 3 to 8 years in freshwater habitats as filter feeders, living in burrowed fine sediment until they undergo metamorphosis to migrate to the sea and start hematophagous feeding on their prey. Sea lampreys spend between 4 and 10 months without feeding during the metamorphosis stage, which is a critical stage for survival due to the fact that postmetamorphic stages require feeding on the prey's blood to recover from lost energy. For these reasons, sea lampreys benefit in moving to estuaries located downstream or in nearby rivers (Silva et al, 2012).

At the end of their metamorphose period, the sea lampreys reach their adult forms having taken on well-developed eyes, a circular sectorial mouth with teeth, and a blue and silver coloration. At this point the lampreys migrate downstream primarily in the fall and begin their feeding on the local ecosystem present in the lakes. They are capable of reaching lengths from 12 to 24 inches in 12 to 20 months. In the late fall, they migrate toward the mouths of the rivers and ascend to spawn in the spring (Johnson et al, 1969).

A growing interest in genomic expression throughout vertebrate embryogenesis has become a key concept in understanding modern science. Developing advanced technological systems to detect gene expression have become essential in fulfilling an exact analysis of a vertebrate's embryogenesis (Shendure et al, 2008). This level of study is novel and has yet to be completely understood. A number of species were analyzed for their genomic expression in embryogenesis (Shen et al, 2011). However, a full transcriptome analysis of *Petromyzon marinus* is a recent development.

To date, no research has been presented on the network analysis of the sea lamprey embryological stages from a transcriptome approach to identify gene modules and candidate genes associated with embryogenesis from a mechanistic point of view. Key genes associated with each developmental stage needs to be identified by comparing transcriptome of the developing embryos.

Over the past decade, significant progress has been made in genome wide gene expression profiling by the development of microarrays. However this technique was limited in its technological capabilities for the sought out results. For instance, this method depends on existing knowledge about genome sequences; high background levels owing to cross-hybridization; and a limited dynamic range of detection owing to both background and saturation of signals (Wang et al, 2010). Moreover, comparing expression levels across different experiments is often difficult and can require complicated normalization methods (Wang et al, 2010). In gene-expression studies, microarrays are now being replaced by Next Generation Sequencing (NGS) based methods (Metzker et al, 2010). In contrast to microarray methods, NGS provides advantages to study novel genes and revolutionizes eukaryotic transcriptome analysis based on sequence reads contributed from the product outcome (Cloonan et al, 2008; Morozova et al, 2008; Ozsolak et al, 2010; Wang et al, 2009). This project will utilize NGS to detect key genes involved in sea lamprey embryonic development.

Next Generation Sequencing provides an inexpensive, genome-wide sequence readout as an endpoint to applicants ranging from chromatin immunoprecipitation, mutation mapping and polymorphism discovery to noncoding RNA discovery (Mardis et al, 2008). The platform used in this study was SOLiD3 sequencer, which uses a unique sequencing process catalyzed by DNA

ligase. The advantages of having the ligation-based approach and the 8mer labeling are that an extra quality check of read accuracy is enabled (Mardis et al, 2008).

With the use of Next Generation Sequencing, analyzing sea lamprey and vertebrate embryogenesis has become an easier task providing a better understanding of the genetic expression levels and functional enrichments throughout the spatial stages of the organisms (Liang et al, 1992; Shen et al, 2009). A number of different analyses took place to identify genes of interest to grasp a better understanding of sea lamprey embryogenesis. Genes' expression levels were analyzed based on the $\log_2$ values of their Fragments per Kilobase of exon per Million (FPKM) values in the adjacent stages (Trapnell et al, 2010). This derivative helped isolate genes with expression folds greater than two (Toung et al, 2011). The genes were then put through an open source network platform, Cytoscape that analyzes the genes' interactions based on previous published research articles that show strong evidence supporting the physical, pathway, and predicted outcomes of neighboring genes (Saito et al, 2012).

In this project, we are using the next generation sequencing technique to profile gene expression patterns among embryogenesis of sea lampreys. Eight stages of embryogenesis were used in the transcriptome analysis of the sea lamprey; Egg, Neurula, Head Protrusion, Hatching, Melanophore, Eye Spot, Gall Bladder, and Completion. The goal is to identify key genes that are essential for embryonic development and that are good targets of bisazir-induced DNA damage to develop qPCR tests that can identify sterility in bisazir-treated males. We are using the next generation sequencing technique to profile gene expression patterns among embryogenesis of sea lampreys. The sea lamprey is also an ideal model organism for the investigation of early vertebrate evolution since they are phylogenetically one of the most basal animals that contain true vertebrate characteristics.

Methods and Materials

Embryogenesis

Eggs stripped from mature females were fertilized with sperm from adult males in the laboratory of Hammond Bay Biological Station (Millersburg, Michigan). Fertilized eggs and resultant embryos and larvae were maintained in aerated lake water in a constant water bath held at 18°C. Embryos were staged according to the criteria established by Tahara (Tahara, 1988).

Embryos at all major different development stages (fertilized egg, Blastula, Gastrula, Neural Plate, Head Protrusion, Hatching, Melanophore, Eye Spot, Gall bladder, and Completion of digestion tract) from at least 5 different pairs of parent animals were transferred into a labeled frozen tube and snap frozen in liquid nitrogen then stored at -80°C freezers until RNA isolation.

RNA Isolation

Total RNA was isolated according to Zhang (2013). Briefly, embryo samples were weighed under semi-frozen condition. 10 ml/g of Trizol solution (Invitrogen, Carlsbad, CA) was added to the sample, and then rapidly homogenized by a power-driven homogenizer. The cell debris and nucleic DNA was removed by centrifugation. Chloroform with 0.2 times the volume of Trizol solution was added to the solution. After vortexing and centrifugation, the aqueous layer was transferred to column of RNeasy mini RNA isolation kit (Qiagen, Valencia, CA) for further purification. The residual DNA was eliminated by performing a column DNase digestion at 37°C for 30 minutes. The integrity of RNA was determined by gel electrophoresis and its concentration was measured using a Nanodrop spectrophotometer.

Next Generation Sequencing and Mapping

Total isolated RNA was sent for sequencing using Life Technologies SOLiD3 platform. This study used Tuxedo Suite, which aligned the reads of the embryos (Chen et al. 2009). RAW data was submitted and analyzed based on FPKM, $\log_2$, and assorting various outcomes of the adjacent stages in Microsoft Excel. Over 14,000 genes were found and used for analysis. Once the desired expression levels were determined amongst the genes (Trapnell et al. 2009), Cytoscape was used in determining their interactions and importance for the different stages in embryogenesis.

Analysis of Cytoscape Network Interaction

Cytoscape 3.0.1 provided Gene Ontology (GO), Kyoto Encyclopedia of Genes and Genomes (KEGG), and GeneMANIA plug-in, which determined the biological function and significance of the genes that were inputted from Excel. Cytoscape provided visualize characteristics using hierarchical viewing of gene analysis as well as gene expression calculations brought from Excel. All interactions are predicted from previous research that was inputted into Cytoscape (Shannon et al. 2003). Each protein is represented by a circle (node) and each line (edge) represents direct interactions amongst the proteins. GeneMANIA is a plug-in within the program that identifies the genes inputted from Microsoft Excel within its own library and annotates the interactions based off of previous research. Each node and edge has various sizes and thickness, which correlates to the amount of evidence that the protein or interaction exists in today. Both Cytoscape and GeneMANIA plug-in were downloaded from www.cytoscape.org.

RT-PCR

RNA samples from different stages of Sea Lamprey Embryos were used in RT-PCR; High Capacity cDNA Reverse Transcription Kit provided by Applied Biosystems was used. The stages were F3, 9-4, 9-5, and 9-6 which are the Female eggs, Head Protrusion, Hatching, and Melanophore stages, respectively. Each embryo had its own RNA concentration that was recorded. The female egg was 107ng/μL, Head Protrusion stage was 430 ng/μL, Hatching was 933ng/μL, and Melanophore was 758ng/μL. The kit required a concentration of up to 2 μg (for a 20-μL reaction) of total RNA to cDNA. The samples were converted to 1μg to identify how many μL would be needed to use in the RT-PCR. 3.57μL of the female egg was used, 2.33μL of the Head Protrusion stage embryo was used, 1.07μL of the Hatching stage embryo was used, and 1.32μL of the Melanophore stage embryo was used.

In preparing the 2X RT master mix (20-μL reaction) for 5 samples, the volume/reaction (μL) was accounted for with the following quantifications; 10.0μL of 10X RT Buffer, 4.0μL of 25X dNTP Mix (100mM), 10.0μL of 10XRT Random Primers, 5.0μL of Multiscribe Reverse Transcriptase, and 21.0μL of Nuclease-free H_2O was used. The total volume was distributed amongst five conical tubes for each sample including the negative control to detect chances of DNA contamination. After the samples were added, an additional 10μL of ddH_2O were added to total the 20μL total volume for each conical tube required by the kit. Once the RT master mix was established with the RNA samples, the solutions were placed into Labnet International Multi Gene II PCR thermal cycler for reverse transcription. The machine was set to 25°C for 10 minutes, 37°C for 120 minutes, 85°C for 5 minutes, and 4°C for soaking.

In addition to the samples mentioned above, four other samples were tested. Sample embryos at stages, 8-9 (Completion stage), 7-9 (Completion stage from a different embryo), 8-4 (Head Protrusion), and 7-4 (Head Protrusion from a different embryo) were tested under the same circumstances as the previous sample embryos. Multiple samples from each stage were used for verification of gene expression using qPCR; located in supplemental data.

PCR

After the samples were converted to cDNA they were amplified with PCR (GoTaq® G2 DNA Polymerase). A PCR Master Mix was prepared using 62.5µL of PCR Master Mix, 5.0µL of Primer upstream, 5.0µL of Primer downstream, and 42.5µL of Nuclease Free Water. The Master Mix was divided into 5 new conical tubes evenly, and 2.0µL of the original RNA and now cDNA sample was added totaling to 25µL in each of the samples that were used. The samples were placed into the thermal cycler set at 94°C for 30 seconds, 55°C for 40 seconds, 72°C for 1 minute, all for 40 cycles, and soaked at 4°C in Completion of the cycles.

Gel Electrophoresis

Once PCR of the samples was complete, the cDNA were verified using gel electrophoresis in detection of their size and presence in the samples in addition to quality assurance with the use of the negative control sample to rule out any possibilities for contamination. The Gel was prepared at 1.5% using 50mL of TBE buffer with 0.75 grams of Ultrapure Agarose provided by Invitrogen and 2.5µL of EtBr. 1µL of a 100bp ladder was used with 4µL of loading dye. 5µL of each sample was used with 4µL of loading dye per sample were loaded into the gel and ran at 70 volts for 30 minutes.

qPCR Primer Sequences

Primer were designed and used for the selected genes. Below are the sequences of each primer used in the qPCR verification. Each gene has a forward and reverse sequence and each sequence was set with the correlating developmental stage.

Gene	Primer Sequence	Stage
TGFBR1L	TAAGGACGGTCATGTCACCA	Completion
TGFBR1R	CCAGCTCCAGGGAACTCAT	
SMAD2L	ATTCGCCTTCAACCTCAAGA	Melanophore
SMAD2R	TCACTGCTTTCCCCATCTTC	
SLX4L	GAGCGGCAGCTCCAGAAC	Eye spot
SLX4R	GAAGGGCCGGTAAAGGAG	
PXN1L	CACGGGGGATCTTTGTCTTA	Eye spot
PXN1R	ACCACCTGTCCGGCAATC	
PCNAL	TCATCTGCTGCACCAAAGAC	Neurula, Head protrusion
PCNAR	CGATGCCATATTCCACAACA	

Real-Time PCR

Genes identified from Next Generation Sequencing were verified with qPCR (Power SYBR® Master Mix). Primers were designed for the genes and replaced random primers for the qPCR mix. The original stock concentrations of oligonucleotides were diluted to 10X according to nanomolars. 100µM of standard solution were obtained for 40 oligonucleotides, 20 forward (L) and 20 reverse (R). 190µL of DNase free water were added in a new conical tube labeled appropriately for each oligonucleotide that was used, to make a concentration of 5µM. All 40 tubes were vortexed for 20 seconds and spun down at 10,000 rpm for 5 seconds. 10µL the oligonucleotides (L & R) were added to their respective tube from their 100µM concentration to make the 5µM concentrations. 90µL of DNase water and 10µL of cDNA were added into the

conical tubes; standard PCR and gel electrophoresis were examined on samples to observe if oligonucleotides annealed to the genes. Once the oligonucleotides were verified, qPCR was performed to verify the NGS analysis. Six 96-well plates were loaded with selected genes from embryogenesis stages. Each plate-well was loaded with cDNA template of the embryogenesis stages, the designed oligonucleotides for the selected genes, and qPCR master mix. Each plate row had four selected genes and three different pairs of the same stage analysis. The first three columns were assigned one primer at chronological stages descending down the columns and different variants of the same stages were used in each row for the set columns. In total, seven stages were verified and the last row was used as a reference with a mix of stages. Each well received 10μL of qPCR master mix, 2μL of primers, (L & R), and 8μL of corresponding template for the stages, totaling to 20μL. The plates were covered and inserted into the Applied Biosystems Step One Plus Real Time PCR system for analysis. Sample table may be found in Results section.

Results

Embryonic Isolation and Microscopic Visualization

Microscopic Images were taken for visual representation of sea lamprey embryos at various stages of embryogenesis such as Neurula, Head Protrusion, Hatching, Melanophore, Eye Spot, Gall Bladder, and Completion as seen in Figure 1A-G. Embryos at these select stages were isolated and sacrificed and used for Transcriptome analysis.

A B

C D

E F G

Figure 1A-G. Representation of chronological stages of Sea Lamprey embryogenesis that were isolated and sacrificed at designated time intervals. A) Neurula B) Head Protrusion C) Hatching D) Melanophore E) Eye Spot F) Gall Bladder G) Completion

Transcriptome sequencing and gene coverage

To identify genes involved in sea lamprey embryogenesis, we sequenced cDNA from ovary, Neurula, Head Protrusion, Hatching, Melanophore, Eye Spot, Gall Bladder, and Completion stages. For each of these samples, transcriptome sequences were obtained by using the SOLiD3 platform. After assembly, mapping, and reads counting, Table 1 shows the amount of genes that are relevant to embryogenesis in each stage based on $\log_2$ expression from NGS that were inputted and recognized by Cytoscape.

Identified Genes from Cytoscape

Stage	Identified	Total
Ovary	5785	8442
Neurula	6532	9731
Head Protrusion	4435	6920
Hatching	6530	10039
Melanophore	6129	9439
Eye Spot	6142	9430
Gall Bladder	5502	8281
Completion	5692	8386

Table 1. Amount of genes identified in Cytoscape versus amount of genes inputted from Next Generation Sequencing Analysis.

Annotation Results

To select a prospective gene as being significant for a particular embryonic stage, the gene's expression had to be expressed at least three-fold based on a $\log_2$ calculation. Figure 2A shows the amount of transcripts expressed at the highest level for each stage. Of importance, the difference is of the amount of transcripts between the beginning of embryogenesis (Neurula) and the end of embryogenesis (Completion stage). The Neurula stage has 9,729 expressed genes while the Completion stage has 8,386 genes, a difference of 1,343 genes. Figure 2B shows the number of transcripts specifically detected at each individual stage, where the Neurula stage had more transcripts than the Completion stage. Figure 2C shows the number of transcripts differentially expressed in adjacent stages. The greatest number of transcripts was detected from the Head Protrusion/Neurula stage with 2,295 transcripts.

A

Figure 2A. Depicts the total number of transcripts expressed at the highest level for each individual embryonic stage.

B

Figure 2B. Shows the total number of transcripts expressed specifically at each stage with no expression in other stages.

Figure 2C. Depicts the total number of transcripts expressed when comparing adjacent stages based on a minimal expression fold of three.

Cytoscape

The seven different embryonic stages and female ovaries were individually analyzed using Cytoscape to display the protein-protein interactions with in a hierarchal layout. Figure 2C describes the breakdown of the amount of genes that were used to build each Cytoscape network. Each network consisted of genes that GeneMANIA used from the annotation analysis along with predicted interacting genes. The genes from the analysis correspond to nodes (represented as red circles for being up-regulated, green circles for being down-regulated, and grey circles for the predicted genes by GeneMANIA). The edges (connecting lines between the nodes) represent documented interactions. The level of significance based on previous studies is represented by the size of the nodes and thickness of the edges where the bigger the node and the thicker the edge represented greater evidence for the interactions.

Gene Ontology

Once the networks were created for the selected genes from each adjacent stage comparisons, Gene Ontologies for all genes were inputted into Microsoft Excel with analysis representing GO id description, q-value, and occurrence. Table 2 represents a total of 26 functional characteristics that were obtained for all the genes in all the networks collected as a whole. Along with Genecards, Gene Ontology evaluates the significance of the genes for the interactions and allows clustering their similarities based on functional characteristics. Description represents gene summary, q-value represents margin of error at each stage based on Occurrence, and Occurrence represents amount of genes present at each stage with relative Description (Ashburner et al. 2000).

GO id	Description	q-value (Egg)	Occurrences in Egg	q-value (Neurula)	Occurrences in Neurula	q-value (Head protrusion)	Occurrences in Head protrusion	q-value (Hatching)	Occurrences in Hatching	q-value (Melanophore)	Occurrences in Melanophore	q-value (Eye spot)	Occurrences in Eye spot	q-value (Gall Bladder)	Occurrences in Gall Bladder	q-value (Completion)	Occurrences in Completion
GO:0005925	focal adhesion	0.07803516	27	0.00529619	31	0.02607315	26	0.00887594	25	0.00410845	26	4.75715E-05	30			0.00277977	30
GO:0002764	immune response-regulating signaling pathway							0.00988137	53								
GO:0008624	induction of apoptosis by extracellular signals							0.01465755	32	0.03121668	31	0.02317604	31	0.08762073	20		
GO:0017016	Ras GTPase binding	0.07859334	22					0.09182319	18	0.0432294 36	19			1.89573E-16	66	1.41527E-16	68
GO:0044391	ribosomal subunit	6.0346E-12	63	3.5991E-17	72	6.80751E-17	66	1.9027E-21	68	1.22185E-19	66	2.84553E-20	66				
GO:0015934	large ribosomal subunit	2.2801E-07	36	5.5672E-09	39	2.08625E-08	35	2.2831E-11	37	1.6884E-10	36	1.31002E-11	37	1.06719E-08	36	3.52399E-08	36
GO:0015935	small ribosomal subunit	0.00018938	27	1.5935E-08	34	5.2907E-09	32	2.1609E-10	32	1.72737E-09	31	5.59489E-09	30	5.89426E-08	31	7.34405E-09	33
GO:0007179	transforming growth factor beta receptor signaling pathway	0.04078949	31	0.02873487	32	0.08376901	27			0.01560169	27	0.04660163	25	0.01294778	30	0.02868112	30
GO:0004674	protein serine/threonine kinase activity	1.506E-05	96	3.8725E-06	100	0.01094927 1	75	2.8094E-10	93	1.2952 9E-08	89	2.29675E-08	87	2.16807E-06	91	0.00013327	88
GO:0070412	R-SMAD binding		9	0.03880257	9			0.00468223	9	0.00496774	9			0.04818239	9		
GO:0016049	cell growth	0.07763743	68	0.01658916	73			0.06693025	55	0.00431228	61	0.00410432	60	0.0013546	70	0.00030533	72
GO:0016049	regulation of cell growth	0.03762024	56	0.0142303	59					0.01059817 2	48	0.01198761 8	47	0.00228793	56	0.00037893 4	61
GO:0007067	mitosis	2.6951E-10	92	2.8555E-09	91	1.0355E-10	85	2.8942E-15	88	5.17641E-15	88	1.39096E-13	84	4.21854E-10	85	1.56931E-07	81
GO:0045597	positive regulation of cell differentiation							0.08570364	59			0.09109812 5	58			0.07020955	70
GO:0004674	protein serine/threonine kinase activity	1.506E-05	96	3.8725E-06	100	0.01094927 1	75	2.8094E-10	93	1.2952 9E-08	89	2.29675E-08	87	2.16807E-06	91	0.00013327	88
GO:0071900	regulation of protein serine/threonine kinase activity	0.00767411	76	0.05333856	73			0.00681096	62	0.00193949 9	65	0.00778257 3	61	0.00873190 4	69	0.00775254 6	72
GO:0031396	regulation of protein ubiquitination	1.7255E-12	71	2.5433E-14	75	2.2403 9E-13	67	5.5664E-21	74	1.4447 7E-16	68	1.5557 4E-18	70	3.0413 7E-12	66	6.9889E-16	74
GO:0006281	DNA repair	3.738E-17	124	6.8879E-20	131	1.0624 7E-19	119	2.4524E-23	117	4.5477E-24	119	9.02409E-28	124	4.3191 6E-18	117	1.9908E-15	115
GO:0008629	induction of apoptosis by intracellular signals			0.01890084	30			0.06019866	23	0.00086858	28	0.00453496	26	0.02175163 5	27		
GO:0004519	endonuclease activity	0.00133108	21	0.0005101	22	0.00073276 9	20	0.00137828	18	0.00010384	20	7.5231 9E-05	20	0.08079288	16	0.09936592	16
GO:0006310	DNA recombination	6.5885E-13	69	6.2581E-12	68	8.4649 9E-11	60	2.3721E-11	57	3.0779 9E-14	62	1.7308 2E-13	60	7.7699 9E-11	61	1.9048 1E-09	60
GO:0000725	recombinational repair	0.00409133	21	0.01541112	20	0.04813206 9	17	0.03030405	16	0.00398227 3	18	0.00946148	17	0.00777221 6	19		
GO:0034061	DNA polymerase activity							0.09360962	8			0.08330621 7	8			0.09059635	9
GO:0006260	DNA replication	1.7255E-12	88	4.9058E-11	86	1.0714 7E-11	79	4.5545E-18	85	1.3418 3E-16	83	3.9447 6E-16	81	1.1496 4E-07	71	3.9473 5E-09	77
GO:0006275	regulation of DNA replication	0.03281894	26	0.04278275	26	0.02580730 1	24	0.00145589	25	0.00163508 5	25	0.00114930 3	25			0.05791053	24
GO:0008408	3'-5' exonuclease activity	3.7993E-05	17	7.7246E-06	18	0.00028094 9	15	7.5617E-08	18	8.5718 9E-07	17	3.8233 3E-05	15	0.00148426 5			

Table 2. Overall analysis from Cytoscape of genes based on (Columns) GO id, Description, q-value, and Occurrence and (Rows) GO id references in the Gene Ontology database.

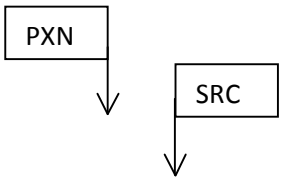
Characterization of different genes within each embryonic stage based on their expression, annotation, and network prediction

In the sea lamprey embryogenesis, eight stages including Egg, Neurula, Head Protrusion, Hatching, Melanophore, Eye Spot, Gall Bladder, and Completion were studied with seven adjacent stage comparisons for determining the significance of expression throughout embryogenesis.

Neurula/Egg Related Genes

In the Neurula/Egg stage comparison, an important gene of interest is PXN (paxillin), which is translated into a cytoskeletal protein that functions with actin-membrane attachment at sites of cell adhesion to the extracellular matrix (focal adhesion). This gene has an expression fold of -5.53, determining the gene to be less expressed in the Neurula stage than in the Egg stage as seen in Figure 3. PXN controls multiple genes and in particular interest is the SRC (v-src sarcoma (Schmidt-Ruppin A-2) viral) that functions in signaling pathways that controls various biological activities including gene transcription, immune response, cell adhesion, cell cycle progression, apoptosis, migration, and transformation and may play a role in the regulation of embryonic development (GeneCards). This gene has an expression fold of -3.47 as seen in Figure 3.

A)



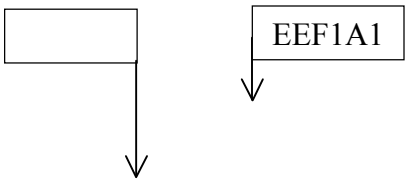
B)

C)

Figure 3. Interactome analysis of Neurula/Egg stages. Cytoscape image of PXN physical interactions, predicted interaction, pathway, and genetic interaction scaled out of 100. Expression level is strictly localized for PXN and neighboring genes. A) Represents graphic visualization. B) Expression Level of PXN interactome. C. Edges (Lines between nodes [Genes]) functions.

Another important gene in the Neurula/Egg stage is EEF1A1 (eukaryotic translation elongation factor 1 alpha 1). It is responsible for the enzymatic delivery of aminoacyl tRNAs to the ribosome. Various isoforms showed this gene to be expressed in areas such as the heart, brain, and skeletal muscle (GeneCards). The expression fold of EEF1A1 is 6.04 out of a total 13.46 for maximum expression fold in this interactome as seen in Figure 4. This gene's upper central location within the network along with its various node interactions gives it the ability to influence a lot of different downstream genes of interest. For instance, RPLP1 (ribosomal protein, large, P1), catalyzes protein synthesis consisting of 40S and 60S subunits playing an important role in the elongation step of protein synthesis (GeneCards). This gene has the maximum expression fold of 13.46, identifying it as the highest protein producing gene in this interactome.

A)



B)

C)

Figure 4. Interactome analysis of Neurula/Egg stages. Cytoscape image of EEF1A1 physical interactions, predicted interaction, pathway, and genetic interaction scaled out of 100. Expression level is strictly localized for EEF1A1 and neighboring genes. A) Visual overview of EEF1A1 and neighboring genes interactome. B) Close-up image of figure 4A. C) Expression value of EEF1A1 interactome.

Protein kinase activity is seen in the stage comparison as well where TGFBR1 (transforming growth factor, beta receptor 1) is one of those genes with this function. The protein encoded by this gene is a serine/threonine protein kinase. Activated TGFBR1 phosphorylates SMAD3 (R-SMAD), which dissociates from the receptor and interacts downstream with Smad signaling pathways. TGFBR1 gene has an expression fold of -3.92 while SMAD3 has an expression fold of -6.98 as seen in Figure 5.

A)

B)

Figure 5. Interactome analysis of Neurula/Egg stages. Cytoscape image of TGFBR1 physical interactions, predicted interaction, pathway, and genetic interaction scaled out of 100. Expression level is strictly localized for TGFBR1 and neighboring genes. A) Visual overview of TGFBR1 and neighboring genes interactome. B) Expression value of TGFBR1 interactome.

Hatching/Head Protrusion Related Genes

In the Head Protrusion/ Hatching stage, EEF1A1 is expressed and up-regulated at this comparison as well. The gene has an expression fold of 3.50 out of 9.90 as seen in Figure 6. Aside from the ribosomal proteins that EEF1A1 controls from the previous and current comparisons, it controls the gene DDX39A (DEAD [Asp-Glu-Ala-Asp] box polypeptide 39A). The gene encodes a member of the DEAD box protein families that are characterized by the conserved motif Asp-Glu-Ala-Asp (DEAD) and are putative RNA helicases (GeneCards). DDX39A has an expression fold of 4.84.

A)

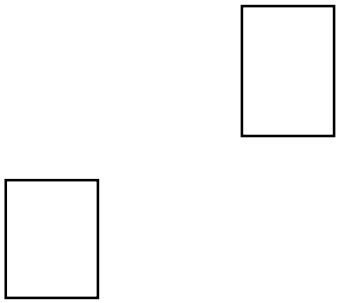
B)

C)

Figure 6. Interactome analysis of Hatching/Head Protrusion stages. Cytoscape image of EEF1A1 physical interactions, predicted interaction, pathway, and genetic interaction scaled out of 100. Expression level is strictly localized for EEF1A1 and neighboring genes. A) Visual overview of EEF1A1 and neighboring genes interactome. B) Expression value of EEF1A1 interactome. C) Edges (Lines between nodes [Genes]) functions.

Along with EEF1A1, PXN (paxillin) gene plays a critical role in the Hatching/Head Protrusion stage comparison. PXN has an expression fold of 3.42 from a total of 10.92 up-regulated expression fold as seen in Figure 7. Earlier results showed PXN to be down-regulated at the Neurula/Egg stage comparison. PXN controls PTPRG (protein tyrosine phosphatase, receptor type, G) gene. The protein encoded by this gene is a member of the protein tyrosine phosphatase (PTP) family. PTPRG is up-regulated with an expression fold of 4.00.

A)



B)

Figure 7. Interactome analysis of Hatching/Head Protrusion stages. Cytoscape image of PXN physical interactions, predicted interaction, pathway, and genetic interaction scaled out of 100. Expression level is strictly localized for PXN and neighboring genes. A) Visual overview of PXN and neighboring genes interactome. B) Expression value of PXN interactome.

Melanophore/Hatching Related Genes

In the Melanophore to Hatching comparison, TRAF6 (TNF receptor-associated factor 6, E3 ubiquitin protein ligase) gene was found to be of functional importance according to its location in the stage comparison interactome. TRAF6 protein encoded by this gene is a member of the TNF receptor associated factor (TRAF) protein family and is associated with, mediating signal transduction from, members of the TNF receptor superfamily as well as the Toll/IL-1 family (Cao, 1996). TRAF6 is down-regulated with an expression fold of -8.93 as seen in Figure 8. One of the genes TRAF6 controls is RPS27A (ribosomal protein S27a) gene better known as ubiquitin, is a highly conserved protein that functions by targeting cellular proteins for degradation by the 26S proteasome. RPS27A has an expression of -6.65, one of the lowest expressed genes in the interactome.

A)

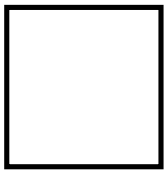
B)

C)

Figure 8. Interactome analysis of Melanophore/Hatching stages. Cytoscape image of TRAF6 physical interactions, predicted interaction, pathway, and genetic interaction scaled out of 100. Expression level is strictly localized for TRAF6 and neighboring genes. A) Visual overview of TRAF6 and neighboring genes interactome. B) Expression value of TRAF6 interactome. C) Edges (Lines between nodes [Genes]) functions.

Alongside TRAF6, SMAD2 gene (SMAD family member 2) is another gene of importance for this comparison. The protein encoded by this gene regulates cellular processes, such as cell proliferation, apoptosis, and differentiation. SMAD2 is up-regulated in the Melanophore/Hatching comparison with an expression fold of 8.69, which is the highest expression fold in this interactome as seen in Figure 9.

A)



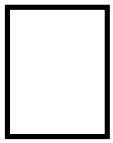
B)

Figure 9. Interactome analysis of Melanophore/Hatching stages. Cytoscape image of SMAD2 physical interactions, predicted interaction, pathway, and genetic interaction scaled out of 100. Expression level is strictly localized for SMAD2 and neighboring genes. A) Visual overview of SMAD2 and neighboring genes interactome. B) Expression value of SMAD2 interactome.

Eye Spot/Melanophore Related Genes

In the Eye Spot to Melanophore comparison, SLX4 (SLX4 structure-specific endonuclease subunit homolog [*S. cerevisiae*]) gene was one of the significant genes. SLX4 gene is involved with endonuclease activity. It is required for recovery from DNA-damage and is involved with resolution of DNA double-strand breaks (GeneCards). This gene has an expression fold of 4.42 out of a max 14.40 as seen in Figure 10.

A)



B)

C)

Figure 10. Interactome analysis of Eye Spot/Melanophore stages. Cytoscape image of SLX4 physical interactions, predicted interaction, pathway, and genetic interaction scaled out of 100. Expression level is strictly localized for SLX4 and neighboring genes. A) Visual overview of SLX4 and neighboring genes interactome. B) Expression value of SLX4 interactome. C) Edges (Lines between nodes [Genes]) functions.

Another gene with significance in the Eye Spot to Melanophore comparison is the down-regulated TGFBR1, which was previously observed in the Neurula/Egg stage comparison. TGFBR1 is down regulated with an expression fold of -3.85 from the lowest expression fold of -18.56 as detected in Figure 11. TGFBR1 controls ENG (endoglin) gene that is involved in forming a heterodimeric complex with TGFBR1. ENG is down-regulated with an expression fold of -3.34.

A)

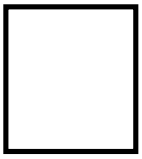
B)

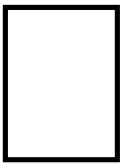
Figure 11. Interactome analysis of Eye Spot/Melanophore. Cytoscape image of TGFBR1 physical interactions, predicted interaction, pathway, and genetic interaction scaled out of 100. Expression level is strictly localized for TGFBR1 and neighboring genes. A) Visual overview of TGFBR1 and neighboring genes interactome. B) Expression value of TGFBR1 interactome.

Gall Bladder/Eye Spot Related Genes

TGFBR1 is also present in the Gall Bladder/Eye Spot stage comparison with an expression fold of 5.44 out of 6.17 as shown in Figure 12. ENG is also up-regulated with expression fold of 5.92.

A)





B)

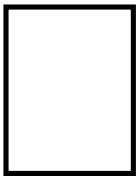
C)

Figure 12. Interactome analysis of Gall Bladder/Eye Spot stages. Cytoscape image of TGFBR1 physical interactions, predicted interaction, pathway, and genetic interaction scaled out of 100. Expression level is strictly localized for TGFBR1 and neighboring genes. A) Visual overview of TGFBR1 and neighboring genes interactome. B) Expression value of TGFBR1 interactome. C) Edges (Lines between nodes [Genes]) functions.

Completion/Gall Bladder Related Genes

In the final stage comparison of Completion/Gall Bladder, two genes showed significant expression level and biological properties. One of the genes of interests is the PCNA (proliferating cell nuclear antigen) gene. PCNA helps increase the processivity of leading strand synthesis during DNA replication. The protein can be found in the nucleus and is a cofactor of DNA polymerase delta. It has an expression fold of 3.98 out of 8.02 as seen from Figure 13.

A)



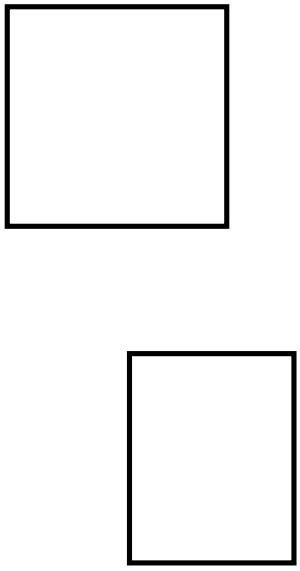
B)

C)

Figure 13. Interactome analysis of Completion/Gall Bladder stages. Cytoscape image of PCNA physical interactions, predicted interaction, pathway, and genetic interaction scaled out of 100. Expression level is strictly localized for PCNA and neighboring genes. A) Visual overview of PCNA and neighboring genes interactome. B) Expression value of PCNA interactome. C) Edges (Lines between nodes [Genes]) functions.

EXOSC10 (exosome component 10) is the second significant gene in the Completion/ Gall Bladder comparison. It functions as a putative catalytic component of the RNA exosome complex which has 3'→5' exoribonuclease activity and participates in a multitude of cellular RNA processing and degradation events (Lejeune, 2003; West, 2006; Mullen, 2008). Its expression fold is the highest at 5.16, as seen in Figure 14. EXOSC10 controls the downstream C1D (Nuclear Receptor Corepressor) gene. The protein encoded by this gene is a DNA binding and apoptosis-inducing protein that is localized in the nucleus (GeneCards). C1D has an expression fold of 3.92.

A)



B)

Figure 14. Interactome analysis of Completion/Gall Bladder stages. Cytoscape image of EXOSC10 physical interactions, predicted interaction, pathway, and genetic interaction scaled out of 100. Expression level is strictly localized for EXOSC10 and neighboring genes. A) Visual overview of EXOSC10 and neighboring genes interactome. B) Expression value of EXOSC10 interactome.

Preliminary quantitative Real Time PCR

qPCR was utilized to verify the data from NGS. TGFBR1, SMAD2, PCNA, PXN, TRAF6 and SLX4 were selected for qPCR verification. The 18S ribosomal gene was used as the house-keeping gene to determine relative quantification for gene expression. Adjacent stage ΔC_T values were analyzed for gene expression. In this novel experiment, evidence of gene expression was verified with TGFBR1. Using NGS and qPCR will help determine a vast amount of gene interactions in a given transcriptome. This technique may initiate the control of sea lamprey embryogenesis in the Great Lakes and restore the natural fish ecosystem.

Figure 15. qPCR verification of TGFBR1 expression levels in the Eye Spot/Melanophore comparison. Values are the inverse of ΔC_T to represent expression from readings.

Figure 16. qPCR verification of TGFBR1 expression levels in the Gall Bladder/Eye Spot comparison. Values are the inverse of ΔC_T to represent expression from readings.

Gel Electrophoresis

Gel Electrophoresis was utilized as an additional verification tool for qPCR results to illustrate the presence or absence of genes in the samples. It should be noted that bands represent an unbiased gene expression. Images 8A-D show conclusive and inconclusive gene expression verification observed in qPCR data analysis.

A

B

C

D

Image 8. Gel electrophoresis of genes selected for qPCR verification. Lane 1 represents 100bp DNA Ladder. A) All genes from all plates in sample embryos B) PCNA and PXN1 genes from Plates 2 and 3 C) All genes but SLX4 from plate 5 D) SLX4 gene in Egg stage from plate 6. Additional bands represent other genes in other stages.

Discussion

Sea lampreys that arose from the jawless fish, which first appeared around 500 million years ago, are able to remodel their genomes during embryonic development. Soon after the fertilized egg divides into several cells, the embryo discards millions of units of DNA. Only slight modifications for proper immune response were believed to occur in the vertebrate genome as opposed to broad-scale rearrangements seen in invertebrates. However, the DNA in the early embryonic cells had countless breaks that resembled dying cells but the cells were not dying. The embryonic cells had significantly fewer repeat DNA sequences than did the sperm cells and their precursors. The reconstruction occurs throughout embryogenesis where deletions along the strands of DNA are thought to move certain regulatory switches in the genome closer to previously distant segments. Although uncertain of actual mechanism, it is possible that the extra genetic material may be utilized for proliferation of precursor cells for sperm and eggs, and in early embryonic development. Once it is no longer needed, the genetic material is discarded to prevent abnormal growth. The tightly regulated structural changes have been compared to DNA errors that give rise to cancers or other genomic disorders seen in higher vertebrates. Studying the mechanism of how sea lamprey DNA rearrangements are regulated may provide information on what stabilizes or changes the genome in addition to the role of restructuring in helping different types of cells such as fin, muscle, or liver cells. During early embryogenesis, roughly 20 percent of their genome disappears. The germline (precursor cells for sperm and eggs) is a continuous lineage through time and is set apart during early embryogenesis. The germline genome never changes. Genetic material is assumed to be lost only in early embryonic cells destined to give rise to somatic tissue and not in cells that will give rise to the next generation. The embryonic cells play a role where they must be able to undergo meiotic recombination and

the capacity to differentiate into every cell type. These inevitable tasks that the cells face are set at odds with somatic tissues, because factors promoting recombination and pluripotency may disrupt genome integrity or specification of cell fate if they are misexpressed in somatic cell lineages leading to oncogenesis. Somatically deleted genes function in both adult (meiotic) germline and in the developing germline during embryogenesis. They are inherited evenly across all somatic tissues and result in deletions. Examples of genes include APOBEC-$_1$, RNA Binding Motif 46 (cancer/testis antigen 68, which have functional roles in cell fate, cell proliferation and oncogenesis/tumorigenesis (Smith et al, 2009; Smith et al, 2010; Smith et al, 2012).

In this study, Figure 2C represents a decrease of gene production ranging from the Head Protrusion/Neurula stage with 2,295 genes detected compared to Completion/Gall Bladder stage with 446 genes. Additionally, a couple of the selected genes are involved with similar functional ontologies; SMAD2 (cell proliferation, apoptosis, and differentiation) and SLX4 (primordial germ cell proliferation and induced apoptosis). All of the selected genes were analyzed based on their functional annotation derived from Gene Ontology.

Neurula/Egg Related Genes

Table 2 shows occurrence of focal adhesion more in the Egg than in the Neurula. PXN may be one of the genes that are included in the description even though its expression level at this stage of vertebrate development is down-regulated. There may be other genes more important for focal adhesion at this stage where PXN would not be needed (Mazaki et al, 1997). Table 2 also represents immune response-regulating signaling pathway genes to only have occurrence in the Hatching stage. When cells adhere via focal adhesions to the extracellular matrix, signals are transmitted by integrins into the cell resulting in tyrosine phosphorylation of a number of focal adhesion proteins, including PTK2/FAK1 and paxillin (PXN). Recent studies

have shown PXN to be a signal transduction agent to SRC providing the protein to be active (Schaller et al, 2001). Since PXN cannot function properly with the extracellular matrix, SRC will consequently not function properly as well. Table 2 depicts induction of apoptosis by extracellular signals, such as those by PXN, only showing occurrence in later stages. SRC is also involved in the RAS pathway (Rosen et al, 1986; David-Pfeuty et al, 1995; Giglione et al, 2001; Wang et al, 2011). In Table 2, Ras GTPase binding has no occurrence in the Neurula stage but does have occurrence in the Egg, Hatching, Melanophore, and Gall Bladder stages. Figure 4 is the prime example of the relationship seen between the two genes; both genes are down-regulated, representing little to no activity when observing the Neurula/Egg comparison.

Table 2 shows highest occurrence of ribosomal subunits, large and small, in the Neurula stage. The data supports an earlier study of an elongation factor gene interacting with ribosomal genes at early stages of embryogenesis (Bozinovic et al, 2011). EEF1A1 is used as internal controls in published qPCR experiments where it is a determined and qualified reference gene for distinct developmental periods from early embryonic to end of metamorphosis, but mostly used during early embryogenesis (Dhorne-Pollet et al, 2013).

Table 2 shows a higher occurrence in the Neurula than Egg stage for transforming growth factor beta receptor signaling pathway functions and protein serine/threonine kinase activity. Although TGFBR1 is less expressed in the Neurula stage, it may still play an important role as it is still expressed, albeit not as significant as in the Egg stage. Mutations of signaling cascades produced by TGFBR1 can cause multiple craniofacial malformations as it is involved in palate formation and odontogenesis during embryogenesis (Xu et al, 2007).

Hatching/Head Protrusion Related Genes

EEF1A1's functional characteristics show most levels of occurrence in the Head Protrusion stage according to Table 2. As stated, EEF1A1 controls DDX39A in this comparison. DEAD box genes are involved in a number of cellular processes involving altering the RNA secondary structure. Some members of the DEAD box protein family may be involved in embryogenesis, spermatogenesis, and cellular growth and division (Pryor et al, 2004). Table 2 verifies the occurrence of DDX39A for cell growth at the Hatching stage and no occurrence at the Head Protrusion stage. Since EEF1A1 controls DDX39A, and both genes' function are occurring in the Hatching stage, this annotation verifies the expression level of EEF1A1 and DDX39A to be up-regulated and serving a purpose.

Table 2 shows occurrence of focal adhesion at the Hatching stage, which is a function of PXN. Studies showed that when phosphotyrosine-containing proteins were immunoprecipitated from embryonic chicken tissues extracts and major proteins of 110, 70, and 50 kD were observed, specifically at 70 kD antibodies adhered to paxillin identifying paxillin as a major tyrosine kinase substrate during chick embryonic development. Approximately, 20% of paxillin was phosphorylated on tyrosine during early embryogenesis but undetectable in the adult. Similar studies were observed in rat embryos suggesting that phosphorylation on paxillin may be a critical role controlling cell and tissue formation rearrangement during vertebrate development (Turner et al, 1991). PXN had little to no expression in the earlier embryonic development whereas now in the Hatching/Head Protrusion stages PXN is expressed with regulating protein activity. As mentioned, PXN controls PTPRG. PTPs are known to be signaling molecules that regulate cellular processes including cell growth, differentiation, and mitotic cycle (Barr et al, 2009). In Table 2, occurrence of cell growth, positive regulation of cell differentiation, and

mitosis is seen at the Hatching stage. There is a correlation seen between PXN and genes related to phosphorylation, as depicted by the GO annotations.

Melanophore/Hatching Genes

TRAF6 functions as an E3 ubiquitin ligase that, together with UBE2N and UBE2V1, synthesizes 'Lys-63'-linked-polyubiquitin chains conjugated to proteins, such as IKBKG, AKT1 and AKT2 (Wang et al, 2006; Lamothe et al, 2007; Xia et al, 2009; Yang et al, 2009). TRAF6 was shown to be down-regulated for this comparison and there is less occurrence in the Melanophore stage than in the Hatching stage for regulation of ubiquitin-protein ligase activity. TRAF6 protein has been shown to be an important signal transduction protein in the innate immune system of vertebrates and involved in regulation of a large gene set during early embryogenesis (Stockhammer et al, 2010). One of the genes TRAF6 controls, RPS27A is made of either polyubiquitin chains or a single ubiquitin fused covalently to an unrelated protein. When covalently bound, it can conjugate to target proteins via an isopeptide bond as a polymer linked via different Lys residues of the ubiquitin (polyubiquitin chains) with the help of TRAF6. RPS27A can function differently depending on the Lys residue of the protein (Huang et al, 2006; Komander et al, 2009). Table 2 shows less occurrence of regulation of protein ubiquitination, supporting the down-regulated expression level of RPS27A for the two stage comparisons. Since RPS27A relies on TRAF6, its expression and functional significance would be limited because TRAF6 is down-regulated.

SMAD2 is part of the R-SMAD family and Table 2 shows occurrence in the Melanophore stage for R-SMAD genes. It is recruited to the TGF-beta receptors through its interaction with the SMAD anchor for receptor activation (SARA) protein. TGF-beta receptors

phosphorylate and induce the dissociation of this protein with SARA and the association with SMAD4. This new conjugation is important for the translocation into the nucleus, where the protein can bind to target promoters and form a transcription repressor complex with other cofactors (Lebrun et al, 1999; Lin et al, 2006; Dai et al, 2009). Its expression has been seen in various organs during embryogenesis as well as the nervous system (Dick et al, 1998).

Eye Spot/Melanophore Related Genes

SLX4 functions to protect genome stability by repairing damaged secondary DNA structures that may arise from replication and recombination (Munoz et al, 2009). Table 2 shows occurrence of endonuclease activity, DNA repair, and DNA recombination, and recombinational repair at the Eye Spot stage. Loss of this gene during embryogenesis may cause impaired primordial germ cell proliferation and increased apoptosis. This mutation would reduce spermatogonial pool in the early postnatal testis (Holloway et al, 2011).

GO annotation in Table 2 on TGFBR1 show a less occurrence in the Eye Spot stage compared to the Melanophore stage. This correlates with the expression level of TGFBR1. Although there is no GO annotation for ENG and its function, which is the regulation of angiogenesis (Castonguay et al, 2011) studies have shown that mutations in the gene can cause death in embryogenesis with numerous lesions in the cardiovascular tree, vessel dilation, hemorrhage and abnormal cardiac morphogenesis (Nomura-Kitabayashi et al, 2009). One correlation that is seen is the expression level from NGS and the functional significance of both genes. Both TGFBR1 and ENG are down-regulated, which support their biological role in their physical interaction as a complex protein.

Gall Bladder/Eye Spot Related Genes

TGFBR1 was the gene of interest for this stage comparison as it showed functional value based on its upstream location in the hierarchical view of the interactome for these stage comparisons. Having lower expression present in the previous adjacent stage comparison, TGBR1 is up-regulated and Table 2 shows more occurrence of its functions within the Gall Bladder stage than the Eye Spot stage. ENG follows the same expression pattern and its functional description has more occurrence in the Gall Bladder stage then Eye Spot stage as well.

Completion.Gall Bladder Related Genes

Table 2 shows occurrence of DNA polymerase activity only at three stages; one of the stages is the Completion stage. There is more occurrence of DNA replication in the Completion stage than in the Gall Bladder stage. There is also occurrence of regulation of DNA replication present in the Completion stage but not in the Gall Bladder stage as seen in Table 2. In response to DNA damage, this protein is ubiquitinated and is involved in the RAD6-dependent DNA repair pathway. Occurrence of ubiquitination is present at the Completion stage as seen in Table 2. PCNA recruits DNA damage response proteins that complete DNA replication after DNA damage: 'Lys-63'-linked polyubiquitination of PCNA helps with the error-free pathway and uses recombination mechanisms to synthesize across the damaged area (Burkovics et al, 2009). Its major function of replication shows most occurrence in the Egg stage. The transcripts in the early embryonic stages are contributed mainly by maternal gene expression (Yamaguchi et al, 1990). Mutants of the strains can show pleiotropy of temperature-sensitive lethality, hypersensitivity to DNA-damaging agents, and female sterility (Henderson et al, 1990).

The final gene of interest is the EXOSC10. Table 2 shows occurrence of 3'→ 5' exonuclease activity at the Completion stage. EXOSC10 is required for nucleolar localization of C1D and may mediate the association of C1D and other genes with the RNA exosome involved in the maturation of 5.8S rRNA. C1D recruits RNA exosome complex to pre-rRNA to mediate 3'-5' end processing of the 5.8S rRNA (Schilders et al, 2007). Figure 14 show that both genes share the same characteristics based on expression level where both genes are up-regulated.

Novel/Putative Genes

All the genes represented along with the Table analysis provide insight to novel studies performed on sea lamprey embryogenesis through Next Generation Sequencing. Further studies may develop as there are thousands of genes detected from NGS that have yet to be analyzed for their functional roles in vertebrate embryogenesis. Some genes are significantly expressed; yet have no biological history developed for their roles thus far.

Having identified key genes in sea lamprey embryogenesis may exploit capabilities in controlling their population within the Great Lakes and restore the ecosystem. Synthesizing chemicals to directly target specific sequences belonging only to sea lampreys may develop a concise mechanism of action to prevent collateral damage in the environment that has once been observed with lampricides. The purpose of this study was to target key genes of sea lampreys to include genes present during embryogenesis that may not have been accounted for previously.

Conclusion

14,476 *Petromyzon marinus* genes during all 8 stages of vertebrate embryogenesis highlight the temporal dynamics of developmental gene expression. The expression intensity of each gene and stage was normalized using the intensity-based log ratio median method (Yang et al, 2003). This procedure helped identify 6,774 genes that showed significant levels of differential expression. Hierarchical clustering of 6,774 genes demonstrated diverse temporal expression profiles of sea lamprey embryogenesis. This subset of developmentally regulated genes was used for further analysis of gene physical, predicted, and pathway interactions based on functional enrichment. The GO annotations verified gene function and role for each of the stages in terms of their occurrence and q-values. Analyzing the differences in gene expression of adjacent stages and identifying initial peak gene activity illustrate the importance of determining stages during embryogenesis correctly (Bozinovic et al, 2011).

Preliminary qPCR and gel electrophoresis helped verify gene expression of the selected genes. Utilizing NGS, qPCR, and gel electrophoresis helped determine presence of key genes in sea lamprey embryogenesis.

Literature Cited

Abdollah S, Macias-Silva M, Tsukazaki T, Hayashi H, Attisano L, Wrana JL. 1997. TβRI phosphorylation of Smad2 on Ser465 and Ser467 is required for Smad2-Smad4 complex formation and signaling. J Biol Chem. 272(44): 27678–27685. doi: 10.1074/jbc.272.44.27678.

Altschul SF, Gish W, Miller W, Myers E.W, Lipman DJ. 1990. Basic local alignment search tool. J. Mol. Biol. 215: 403–410.

Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP, Dolinski K, Dwight SS, Eppig JT, Harris MA, Hill DP, Issel-Tarver L, Kasarskis A, Lewis S, Matese J, Richardson J, Ringwald M, Rubin GM, Sherlock G. 2000. Gene Ontology: tool for the unification of biology. The Gene Ontology Consortium. Nat Genet 25: 25–29.

Barr AJ, Ugochukwu E, Lee WH, King ON, Filippakopoulos P, Alfano I, Savitsky P, Burgess-Brown NA, Müller S, Knapp S. (2009). Large-scale structural analysis of the classical human protein tyrosine phosphatome. Cell. 136: 352–363.

Bothos J, Summers MK, Venere M, Scolnick DM, 2003. Halazonetis T.D. The Chfr mitotic checkpoint protein functions with Ubc13-Mms2 to form Lys63-linked polyubiquitin chains. Oncogene. 22(46): 7101–7107.

Bozinovic G, Sit TL, Hinton DE, Oleksiak MF. 2011. Gene expression throughout a vertebrate's embryogenesis. BMC Genomics. 12: 132.

Burkovics P, Szukacsov V, Unk I, Haracska L. 2006. Human Ape2 protein has a 3'-5' exonuclease activity that acts preferentially on mismatched base pairs. Nucleic Acids Research. Vol. 34(no. 9): 2508–2515.

Castonguay R, Werner ED, Matthews RG, Presman E, Mulivor AW, Solban N, Sako D, Pearsall RS, Underwood KW, Seehra J, Kumar R, Grinberg AV. 2011. Soluble endoglin specifically binds bone morphogenetic proteins 9 and 10 via its orphan domain, inhibits blood vessel formation, and suppresses tumor growth. J Biol Chem. 286: 30034–30046.

Cao Z, Xiong J, Takeuchi M, Kurama T, Goeddel DV. 1996. TRAF6 is a signal transducer for interleukin-1. Nature. 383: 443–446.

Chen Y, Souaiaia T, Chen T. 2009. PerM: efficient mapping of short sequencing reads with periodic full sensitive spaced seeds. Bioinformatics. 25: 2514–2521.

Cloonan N, Grimmond SM. 2008. Transcriptome content and dynamics at single-nucleotide resolution. Genome biology. 9: 234.

Dai F, Lin X, Chang C, Feng XH. 2009. Nuclear export of Smad2 and Smad3 by RanBP3 facilitates termination of TGF-beta signaling. Dev Cell. 16: 345–357. 10.1016/j.devcel.2009.01.022 PubMed: 19289081.

David-Pfeuty T, Nouvian-Dooghe Y. 1995. Highly specific antibody to Rous sarcoma virus src gene product recognizes nuclear and nucleolar antigens in human cells. Journal of Virology. 69(3): 1699–1713.

Dhorne-Pollet S, Thélie A, Pollet N. 2013. Validation of novel reference genes for RT-qPCR studies of gene expression in Xenopus tropicalis during embryonic and post-embryonic development. Dev. Dyn. 242: 709–717. doi: 10.1002/dvdy.23972

Dick A, Risau W, Drexler H. 1998. Expression of Smad1 and Smad2 during embryogenesis suggests a role in organ development. Dev Dyn. 211(4): 293–305.

Giglione C, Gonfloni S, Parmeggiani A. 2001. Differential actions of p60c-Src and Lck kinases on the Ras regulators p120-GAP and GDP/GTP exchange factor CDC25Mm. Eur J Biochem. 268: 3275–83.

Great Lakes Fishery Commision. December 2011. Strategic Vision of the Great Lakes Fishery Commission 2011-2020. Great Lakes Fishery Commission.

Gu MX, York JD, Warshawsky I, Majerus PW. 1991. Identification, cloning, and expression of a cytosolic megakaryocyte protein-tyrosine-phosphatase with sequence homology to cytoskeletal protein 4.1. Proc Natl Acad Sci U S A. 88: 5867–71.

Guo RT, Chong YE, Guo M, Yang XL. 2009. Crystal structures and biochemical analyses suggest a unique mechanism and role for human glycyl-tRNA synthetase in Ap4A homeostasis. J. Biol. Chem. 284: 28968–28976.

Henderson DS, Banga SS, Grigliatti TA, Boyd JB. 1994. Mutagen sensitivity and suppression of position-effect variegation result from mutations in mus209, the Drosophila gene encoding PCNA. EMBO J. 13: 1450–9.

Hofmann RM, Pickart CM. 1999. Noncanonical MMS2-encoded ubiquitin- conjugating enzyme functions in assembly of novel polyubiquitin chains for DNA repair. Cell. 96: 645–653.

Holloway AJ, van Laar RK, Tothill RW, Bowtell DD. 2002. Options available—from start to finish—for obtaining data from DNA microarrays II. Nat. Genet. 32 Suppl. 481–489.

Holloway JK, Mohan S, Balmus G, Sun X, Modzelewski A, Borst PL, Freire R, Weiss RS, Cohen PE. 2011. Mammalian BTBD12 (SLX4) protects against genomic instability during mammalian spermatogenesis. PLoS Genet. 7:e1002094.

Huang F, Kirkpatrick D, Jiang X, Gygi S, Sorkin A. 2006. Differential regulation of EGF receptor internalization and degradation by multiubiquitination within the kinase domain. Mol Cell. 21: 737–748.

Herbert-Johnson BG. 1969. Some statistics of the populations of parasitic phase sea lampreys in Canadian waters of the Great Lakes. Proc. Conf. Great Lakes Res. 12: 45–52.

Komander D, Rape M. 2012. The ubiquitin code. Annu. Rev. Biochem. 81: 203–229.

Kusano S, Raab-Traub N. 2002. I-mfa domain proteins interact with Axin and affect its regulation of the Wnt and c-Jun N-terminal kinase signaling pathways. Mol Cell Biol. 22: 6393–6405. doi: 10.1128/MCB.22.18.6393-6405.2002.

Lamothe B, Besse A, Campos AD, Webster WK, Wu H, Darnay BG. 2007. Site-specific Lys-63-linked tumor necrosis factor receptor-associated factor 6 auto-ubiquitination is a critical determinant of I kappa B kinase activation. J Biol Chem. 282: 4102–4112.

Lebrun JJ, Takabe K, Chen Y, Vale WW. 1999. Roles of pathway-specific and inhibitory Smads in activin receptor signaling. Mol. Endocrinol. 13: 15–23.

Lee NY, Blobe GC. 2007. The interaction of endoglin with beta-arrestin2 regulates transforming growth factor-beta-mediated ERK activation and migration in endothelial cells. J Biol Chem. 282: 21507–21517.

Lejeune F, Li X, Maquat LE. 2003. Nonsense-mediated mRNA decay in mammalian cells involves decapping, deadenylating, and exonucleolytic activities. Mol Cell 12: 675–687.

Liang P, Pardee AB. 1992. Differential display of eukaryotic messenger RNA by means of the polymerase chain reaction. Science. 257: 967–971.

Lin X, Duan X, Liang YY, Su Y, Wrighton KH, Long J, Hu M, Davis CM, Wang J, Brunicardi FC, Shi Y, Chen YG, Meng A, Feng XH. 2006. PPM1A functions as a Smad Phosphatase to Terminate TGF-beta Signaling. Cell. 125: 915–928.

Liu W, Rui H, Wang J, Lin S, He Y, Chen M, Li Q, Ye Z, Zhang S, Chan SC, Chen YG, Han J, Lin SC. 2006. Axin is a scaffold protein in TGF-beta signaling that promotes degradation of Smad7 by Arkadia. EMBO J. 25: 1646–58. doi: 10.1038/sj.emboj.7601057.

Mardis ER. 2008. The impact of next-generation sequencing technology on genetics. Trends Genet. 24: 133–141.

Mathavan S, Lee SGP, Mak A, Miller LD, Murthy KRK, Govindarajan KR, Tong Y, Wu YL, Lam SH, Yang H, Ruan Y, Korzh V, Gong Z, Liu ET, Lufkin T. 2005.Transcriptome analysis of zebrafish embryogenesis using microarrays. PLoS Genet 1(2): e29.

Mazaki Y, Hashimoto S, Sabe H. 1997. Monocyte cells and cancer cells express novel paxillin isoforms with different binding properties to focal adhesion proteins. J Biol Chem. 272(11): 7437–44.

Meng Q, Lux A, Holloschi A, Li J, Hughes JM et al. 2006. Identification of Tctex2beta, a novel dynein light chain family member that interacts with different transforming growth factor-beta receptors. J Biol Chem. 281: 37069-37080.10.1074/jbc.M608614200 PubMed: 16982625.

Metzker ML. 2010. Sequencing technologies—The next generation. Nature Reviews Genetics 11: 31–46.

Morozova O, Marra MA, 2008. Applications of next-generation sequencing technologies in functional genomics. Genomics. 92: 255–264.

Mullen TE, Marzluff WF. 2008. Degradation of histone mRNA requires oligouridylation followed by decapping and simultaneous degradation of the mRNA both 5′ to 3′ and 3' to 5′. Genes Dev. 22: 50–65.

Muñoz IM, Hain K, Déclais AC, Gardiner M, Toh GW, Sanchez-Pulido L, Heuckmann JM, Toth R, Macartney T, Eppink B. 2009. Coordination of structure-specific nucleases by human SLX4/BTBD12 is required for DNA repair. Mol. Cell. 35: 116–127.

Nomura-Kitabayashi A, Anderson GA, Sleep G, Mena J, Karabegovic A, Karamath S, Letarte M, Puri M. 2009. Endoglin is dispensable for angiogenesis, but required for endocardial cushion formation in the midgestation mouse embryo. Developmental Biology. Vol 335. I 1: 66–77.

Ozsolak F, Milos, PM. 2010. RNA sequencing: advances, challenges and opportunities. Nature Reviews Genetics. 12: 87–98.

Politz JC, Yarovoi S, Kilroy SM, Gowda K, Zwieb C, Pederson T. 2000.Signal recognition particle components in the nucleolus. Proc. Natl. Acad. Sci. U.S.A. 97: 55–60. doi: 10.1073/pnas.97.1.55.

Pollet N, Muncke N, Verbeek B, Li Y, Fenger U, Delius H, Niehrs C. 2005. An atlas of differential gene expression during early Xenopus embryogenesis. Mech Dev. 122: 365–439. doi: 10.1016/j.mod.2004.11.009.

Pryor A, Tung L, Yang Z, Kapadia F, Chang TH, Johnson LF. 2004. Growth-regulated expression and G0-specific turnover of the mRNA that encodes URH49, a mammalian DExH/D box protein that is highly related to the mRNA export protein UAP56. Nucleic Acids Res. 32: 1857–1865.

Rosen N, Bolen JB, Schwartz AM, Cohen P, DeSeau V, Israel MA. 1986. Analysis of pp60c-src protein kinase activity in human tumor cell lines and tissues. J. Biol. Chem. 261: 13754–13759.

Safran M, Dalah I, Alexander J, Rosen N, Iny Stein T, Shmoish M, Nativ N, Bahir I, Doniger T, Krug H, Sirota-Madi A, Olender T, Golan Y, Stelzer G, Harel A and Lancet D. 2010. GeneCards Version 3: the human gene integrator Database. doi: 10.1093/database/baq020 [PDF].

Saito R, Smoot ME, Ono K, Ruscheinski J, Wang PL, Lotia S, Pico AR, Bader GD, Ideker T. 2012. A travel guide to Cytoscape plugins. Nat Methods. 14(11): 1069–1076. doi: 10.1038/nmeth.2212.

Schaller MD. 2001. Paxillin: a focal adhesion-associated adaptor protein. Oncogene. 20: 6459–6472.

Schilders G, van Dijk E, Pruijn JM. 2007. C1D and hMtr4p associate with the human exosome subunit PM/Scl-100 and are involved in pre-rRNA processing. Nucleic Acids Res 35: 2564–2572.

Shannon P, Markiel A, Ozier O, Baliga N, Wang J, Ramage D, Amin N, Schwikowski B, Ideker T. 2003. Cytoscape: A Software Environment for Integrated Models of Biomolecular Interaction Networks. Genome Research 13: 2498–2504.

Shen B, Zhang Z, Wang Y, Wang G, Chen Y, Lin P, Wang S, Zou Z. 2009. Differential expression of ubiquitin-conjugating enzyme E2r in the developing ovary and testis of penaeid shrimp Marsupenaeus japonicus. Molecular biology reports. 36: 1149–1157.

Shen Y, Garcia T, Walter R. 2011. Gene Expression Analysis Using RNA-Seq from Organisms Lacking Substantial Genomic Resources, Systems and Computational Biology - Molecular and Cellular Experimental Systems, Prof. Ning-Sun Yang (Ed.). ISBN: 978-953-307-280-7.

Shendure J, Ji H. 2008. Next-generation DNA sequencing. Nature biotechnology. 26: 1135–1145.

Silva S, Servia MJ, Vieira-Lanero R, Cobo F. 2012. Downstream migration and hematophagous feeding of newly metamorphosed sea lamprey (*Petromyzon marinus* Linnaeus, 1758). Hydrobiologia. 700: 277–286.

Smith BR, Tibbles JJ. 1980. Sea Lamprey (*Petromyzon marinus*) in Lakes, Huron, Michigan, and Superior: History of Invasion and Control, 1936. Canadian Journal of Fisheries and Aquatic Sciences. 37(11): 1780–1801, 10. 1139/f80-222.

Smith BR, Tibbles JJ, Johnson BGH. 1974. Control of the Sea Lamprey in Lake Superior, 1953–70. Great Lake Fisheries Commission.

Smith JJ, Antonacci F, Eichler EE, Amemiya CT. 2009. Programmed loss of millions of base pairs from a vertebrate genome. Proc. Natl. Acad. Sci. USA. 106: 11212–11217.

Smith JJ, Baker C, Eichler EE, Amemiya CT. 2012. Genetic Consequences of Programmed Genome Rearrangement. *Current Biology.* 22 (16): 1524–1529. DOI: 10.1016/j.cub.2012.06.028.

Smith JJ, Stuart AB, Sauka-Spengler T, Clifton SW, Amemiya CT. 2010. Development and analysis of a germline BAC resource for the sea lamprey, a vertebrate that undergoes substantial chromatin diminution. Chromosoma. 119: 381–389.

Stockhammer O, Rauwerda H, Jonker MJ, Wittink FR, Breit TM, Meijer AM, Spaink HP. 2010. Transcriptome analysis of Traf6 function in early zebrafish embryogenesis. Mol Immunol. 48(1-3): 179–90.

Stuaffer TM, Hansen MJ. 1958. Distribution of sea lamprey ammocetes in Michigan tributaries of Lake Superior. Mich. Dep. Conserv. Inst. Fish. Misc. Publ. 11: 25 p.

Toung JM, Morley M, Li M, Cheung VG. 2011. RNA-sequence analysis of human B-cells. Genome Res. 21: 991–998.

Trapnell C, Williams BA, Pertea G, Mortazavi A, Kwan G, van Baren MJ, Salzberg SL, Wold BJ, Pachter L. 2010. Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. Nat Biotechnol. 28: 511–515.

Turner CE. 1991. Paxillin is a major phosphotyrosine-containing protein during embryonic development. J. Cell Biol. 115: 201–207.

Velculescu VE, Zhang L, Vogelstein B, Kinzler KW. 1995. Serial analysis of gene expression. Science. 270 (5235): 484–487.

Wang XW, Luan JB, Li JM, Bao YY, Zhang CX, Liu SS. 2010. De novo characterization of a whitefly transcriptome and analysis of its gene expression during development. BMC genomics. 11: 400.

Wang Y, Cao H, Chen J, McNiven MA. 2011. A direct interaction between the large GTPase dynamin-2 and FAK regulates focal adhesion dynamics in response to active Src. Mol Biol Cell. 22: 1529 – 1538.

Wang Y, Tang Y, Teng L, Wu Y, Zhao X, Pei G. 2006. Association of β-arrestin and TRAF6 negatively regulates Toll-like receptor-interleukin 1 receptor signaling. Nat Immunol. 7: 139–147.

Wang Z, Gerstein M, Snyder M. 2009. RNA-Seq: a revolutionary tool for transcriptomics. Nature Reviews Genetics. 10: 57–63.

Wen X, Fuhrman S, Michaels GS, Carr DB, Smith S, Barker JL, Somogyi R. 1998. Large-scale temporal gene expression mapping of central nervous system development. Proc. Natl Acad. Sci. USA. 95: 334–339.

West S, Gromak N, Norbury CJ, Proudfoot NJ. 2006. Adenylation and exosome-mediated degradation of cotranscriptionally cleaved pre-messenger RNA in human cells. Mol Cell. 21: 437– 443. doi:10.1016/j.molcel.2005.12.008.

Xia ZP, et al. 2009. Direct activation of protein kinases by unanchored polyubiquitin chains. Nature. 461(7260): 114–119.

Xu, X. 2007. The Function of TGF-beta Signaling in Palatal and Dental Epithelium During Embryogenesis. ProQuest.

Yamaguchi M, Nishida Y, Moriuchi T, Hirose F, Hui CC, Suzuki Y, Matsukage A. 1990. Drosophila proliferating cell nuclear antigen (cyclin) gene: structure, expression during development, and specific binding of homeodomain proteins to its 5'-flanking region. Mol. Cell. Biol. 10: 872–9.

Yang H. 2003. A segmental nearest neighbor normalization and gene identification method gives superior results for DNA-array analysis. Proc Natl Acad Sci. 7(3): 1122–1127.

Yang W L, Wang J, Chan CH, Lee SW, Campos AD, Lamothe B, Hur L, Grabiner BC, Lin X, Darnay BG, Lin HK. 2009. The E3 ligase TRAF6 regulates Akt ubiquitination and activation. Science. 325: 1134–1138.10.1126/science.1175065.

Zhang Z, Wang Y, Wang S, Liu J, Warren W, Mitreva M, Walter RB. 2011. Transcriptome Analysis of Female and Male *Xiphophorus maculatus* Jp 163 A. PLOS ONE. DOI: 10. 1371/journal.pone.0018379.