University of Kentucky
UKnowledge

Theses and Dissertations--Biology | Biology

2016

Characterization of Somatically-Eliminated Genes During Development of the Sea Lamprey (*Petromyzon marinus*)

Stephanie A. Bryant
University of Kentucky, sbryant1230@gmail.com
Digital Object Identifier: http://dx.doi.org/10.13023/ETD.2016.265

Right click to open a feedback form in a new tab to let us know how this document benefits you.

Recommended Citation
Bryant, Stephanie A., "Characterization of Somatically-Eliminated Genes During Development of the Sea Lamprey (*Petromyzon marinus*)" (2016). *Theses and Dissertations--Biology*. 37.
https://uknowledge.uky.edu/biology_etds/37

This Master's Thesis is brought to you for free and open access by the Biology at UKnowledge. It has been accepted for inclusion in Theses and Dissertations--Biology by an authorized administrator of UKnowledge. For more information, please contact UKnowledge@lsv.uky.edu.

STUDENT AGREEMENT:

I represent that my thesis or dissertation and abstract are my original work. Proper attribution has been given to all outside sources. I understand that I am solely responsible for obtaining any needed copyright permissions. I have obtained needed written permission statement(s) from the owner(s) of each third-party copyrighted matter to be included in my work, allowing electronic distribution (if such use is not permitted by the fair use doctrine) which will be submitted to UKnowledge as Additional File.

I hereby grant to The University of Kentucky and its agents the irrevocable, non-exclusive, and royalty-free license to archive and make accessible my work in whole or in part in all forms of media, now or hereafter known. I agree that the document mentioned above may be made available immediately for worldwide access unless an embargo applies.

I retain all other ownership rights to the copyright of my work. I also retain the right to use in future works (such as articles or books) all or part of my work. I understand that I am free to register the copyright to my work.

REVIEW, APPROVAL AND ACCEPTANCE

The document mentioned above has been reviewed and accepted by the student's advisor, on behalf of the advisory committee, and by the Director of Graduate Studies (DGS), on behalf of the program; we verify that this is the final, approved version of the student's thesis including all changes required by the advisory committee. The undersigned agree to abide by the statements above.

Stephanie A. Bryant, Student

Dr. Jeramiah Smith, Major Professor

Dr. Dave Westneat, Director of Graduate Studies

CHARACTERIZATION OF SOMATICALLY-ELIMINATED GENES DURING DEVELOPMENT OF THE SEA LAMPREY (*PETROMYZON MARINUS*)

THESIS

A thesis submitted in partial fulfillment of the
requirements for the degree of Master of Science in the
College of Arts and Sciences
at the University of Kentucky

By

Stephanie Bryant

Lexington, KY

Director: Jeramiah Smith, Professor of Biology

Lexington, Kentucky

2016

Copyright© Stephanie Bryant 2016

ABSTRACT OF THESIS

CHARACTERIZATION OF SOMATICALLY-ELIMINATED GENES DURING DEVELOPMENT OF THE SEA LAMPREY (*PETROMYZON MARINUS*)

The sea lamprey (*Petromyzon marinus*) undergoes programmed genome rearrangements (PGRs) during early development that facilitate the elimination of ~20% of the genome from the somatic cell lineage, resulting in distinct somatic and germline genomes. To improve our understanding of the evolutionary/developmental logic of PGR, we generated computational predictions to identify candidate germline-specific genes within a transcriptomic dataset derived from adult germline and the embryonic stages encompassing PGR. Validation studies identified 44 germline-specific genes and characterized patterns of transcription and DNA loss during early embryogenesis. Expression analyses reveal that several of these genes are differentially expressed during early embryogenesis and presumably function in early development of the germline. Ontology analyses indicate that many of these genes play known roles in germline development, pluripotency, and oncogenesis (when misexpressed). These studies provide support for the theory that PGR serves to segregate molecular functions related to germline development/pluripotency in order to prevent their potential misexpression in somatic cells. This larger set of eliminated genes also allows us to extend the evolutionary/developmental breadth of this theory, as some deleted genes (or their gnathostome homologs) appear to be associated with the early development of somatic lineages, perhaps through the evolution of novel functions within gnathostome lineages.

Keywords: Lamprey, Development, Genome, Rearrangement, Vertebrate

Stephanie A. Bryant

April 29th 2016

CHARACTERIZATION OF SOMATICALLY-ELIMINATED GENES
DURING DEVELOPMENT OF THE SEA LAMPREY (*PETROMYZON MARINUS*)

By

Stephanie Bryant

Dr. Jeramiah Smith
Director of Thesis

Dr. Dave Westneat
Director of Graduate Studies

April 29th 2016

ACKNOWLEDGEMENTS

I would like to thank my advisor Dr. Jeramiah Smith for his guidance throughout my graduate school career. I am grateful for his patience as he taught me the laboratory techniques and data analysis methods that made this thesis possible and for his guidance during the writing process. I would also like to thank Dr. Ashley Seifert for his unwavering confidence in me, Dr. Ann Morris for her continual encouragement and excellent Halloween parties, and Dr. Randal Voss for his professional insight and for allowing me to foray into axolotl research.

In addition, I would like to thank Melissa Keinath and Joseph Herdy for their support and friendship, and Dr. Lakshmi Pillai of the Morris Lab for her infinite wisdom and sage advice. I also thank Brett Spear and Shirley Qui for granting access to real-time PCR resources used in this project. This work was funded by grants awarded to Dr. Jeramiah Smith and Dr. Chris Amemeyia by the National Institute of Health. Finally, I thank Josh Rock for his support and encouragement throughout the duration of my graduate career.

TABLE OF CONTENTS

Acknowledgements……………………………………………………………………………..iii

List of Figures………………………………………………………………………………iv

Section 1: Introduction……………………………………………………………………1

Section 2: Background……………………………………………………………………4

Section 3: Materials and Methods…………………………………………………………7
- Animals……………………………………………………………………………7
- Lamprey Embryos…………………………………………………………………7
- RNA Sequencing……………………………………………………………………8
- Screening for Somatically-Eliminated Fragments…………………………………8
- Sequencing PCR Products…………………………………………………………9
- Nanostring Gene Expression Analysis……………………………………………10
- Real-Time PCR……………………………………………………………………11
- Differential Expression Analysis…………………………………………………12
- Gene Ontology Analysis……………………………………………………………12

Section 4: Results and Discussion…………………………………………………………14
- RNA Sequencing…………………………………………………………………… 14
- Screening for Somatically-Eliminated Fragments…………………………………15
- Real-Time PCR……………………………………………………………………18
- Nanostring Gene Expression Analysis……………………………………………19
- Gene Ontology………………………………………………………………………23

Section 5: Conclusions……………………………………………………………………27

Vita…………………………………………………………………………………………33

LIST OF FIGURES

Figure 1, Subset of validated deletions with homology information……………………..……16

Figure 2, Full set of validated eliminations (n=44) ……………………………………..…17

Figure 3, Relative abundance of a subset (n=23) of somatically-depleted genes based on qPCR in genomic DNA …………………………………………………………………19

Figure 4, Transcriptional profile and differential expression of somatically-depleted genes based on RNAseq estimates.…………………………………………………………..…21

Figure 5, Transcription and differential expression of somatically-depleted genes based on Nanostring estimates. ……………………………………………………………..…..22

SECTION 1- INTRODUCTION

The sea lamprey (*Petromyzon marinus*) is a basal vertebrate representative of a lineage that diverged from our own ~550 million years ago, and is one of few vertebrates that undergo broad-scale programmed genome rearrangement (PGR) events (Smith, et al. 2009; Smith, et al. 2012). These events occur at 2.5 days post-fertilization (dpf), coinciding approximately with the mid-blastula transition, and facilitate the deletion of approximately 20% of the germline genome from the somatic cell lineage (Smith, et al. 2009). Genome and transcriptome sequencing indicates that a large number of single-copy DNAs, gene-coding fragments, and transcribed genes are eliminated during PGR, and many of these genes appear to correspond to human homologs related to germline pluripotency (Smith, et al. 2009; Smith, et al. 2012; Smith, et al. 2013). Previously identified genes eliminated during PGR include those with roles in cell proliferation (cancer/testis antigen 68, WNT7A/B), tumor suppression (RNA Binding Motif 47), and tumor progression and metastasis (Lysophosphatidic Acid Receptor 1) (Maglott, et al. 2011; Smith, et al. 2012). Somatic misexpression of genes with pluripotency/germline-specific functions may contribute to cancers or other disease states. (Hanahan and Weinberg 2000). As such, it is theorized that PGR may act as a permanent gene silencing mechanism that serves to limit pluripotency functions to the germline to prevent their misexpresson in the soma.

While many of the genes eliminated during PGR are related to germline biology, many others have yet to be characterized. Studies aimed at identifying and characterizing a larger catalog of eliminated genes may provide further insight into the biological role of

PGR. The deep vertebrate ancestry of the sea lamprey and the developmental timing of PGR also raise questions regarding the evolutionary and developmental role of PGR. That hagfish, the only other extant jawless vertebrate, undergo similar genomic rearrangement events suggests that PGR may trace its evolutionary origins to the ancestral vertebrate lineage (Goto, et al. 1998). Furthermore, genome instability is known to be a major driving factor in carcinogenesis and the development of some neurodegenerative diseases in humans (Pikor, et al. 2013). Understanding how the sea lamprey executes such a large-scale genomic rearrangement event without compromising genome integrity may have applications in studies focused on mitigating the effects of genome instability in other vertebrates.

The following studies are aimed at identifying and characterizing genes eliminated during PGR. Leveraging the use of previously generated genomic and transcriptomic resources, we screened sequences for computational evidence of somatic elimination. PCR validation of these sequences identified 44 genes that can be definitively classified as eliminated from the germline during PGR. To characterize potential gene expression, transcriptional profiles were generated for the entire set of validated genes in the developmental stages surrounding PGR and in the adult germline. Analysis of this data identified patterns of differential expression among several eliminated genes, providing additional evidence that PGR targets non-repetitive sequences that are actively transcribed throughout development. Functional information for homologs of deleted genes corroborate previous findings, indicating that many eliminated genes have functions related to pluripotency and germ cell development, but also identify several genes with homologies related to nervous system development.

Moreover, quantitative real-time PCR data for a subset of these genes reveals previously uncharacterized variation in patterns of DNA loss throughout development. These analyses provide new insight into the patterns of gene expression and DNA loss during PGR and contribute to our understanding of the biological role of DNA elimination.

SECTION 2-
BACKGROUND

It was once believed that all cell types within an organism have identical genomes, with cells of the immune system and cancer cells being the notable exceptions. Yet some organisms have been found to undergo extensive DNA elimination events, resulting in germ cells with genomes distinct from that of somatic cells. These events typically occur during early development and can result in elimination of up to ~90% of the germline DNA from somatic cell lineages in some organisms (Wang and Davis 2014). Though the mechanism and biological function of these events are still debated, over 100 organisms across a broad phylogenetic distribution are known to undergo some form of DNA elimination, including nematodes, insects, copepods, and vertebrates such as hagfish, lamprey, and finches (Bachmann-Waldmann, et al. 2004; Drouin 2006; Goday and Esteban 2001; Kubota, et al. 2001; Pigozzi and Solari 1998; Smith, et al. 2009).

Early studies on nematodes (*Ascaris suum* and *Parascaris univalens)* and hagfish (*Eptatretus cirrhatus*) suggested that eliminated DNA was highly repetitive, and more recent studies in zebra finch and copepods seem to corroborate these findings (Degtyarev, et al. 2004; Drouin 2006; Goto, et al. 1998; Itoh, et al. 2009; Muller, et al. 1982; Niedermaier and Moritz 2000). Yet the identification of highly repetitive sequences within eliminated DNA does not exclude the possibility that these events also facilitate the elimination of genes that are actively transcribed in developing or mature germ cells. In sea lamprey (*Petromyzon marinus*), studies have revealed that DNA eliminated during programmed genome rearrangement (PGR) contains large amounts of repetitive DNA (Smith, et al. 2009; Smith, et al. 2010). However, somatically-retained sequences are also

highly repetitive, and a small number of validated genes have been identified within the eliminated fraction of the lamprey genome (Smith, et al. 2013; Smith, et al. 2010). Efforts aimed at identifying eliminated genes have the capacity to provide critical insight into the evolutionary/developmental logic of DNA elimination by identifying biological functions that are withheld from somatic cell lineages but retained by the germline.

Studies with the resolution to parse relatively rare single-copy sequences from large fractions of repetitive DNA, such as those utilizing high-throughput sequencing, have already enabled deeper genomic characterization of DNA elimination events. Deep sequencing of *A. suum* germline and somatic genomes revealed that, in addition to the removal of repetitive DNA, ~12.7 Mb of single-copy sequences are eliminated from the somatic genome (Wang, et al. 2012). Likewise, analysis of low coverage genomic shotgun sequencing of the sea lamprey germline genome identified a handful of well-validated genes that are eliminated during PGR and several other candidate loci (Smith, et al. 2013). Consistent with evidence that eliminated genes in *A. suum* are expressed in the germline and during early embryogenesis, ontology analyses indicate that eliminated protein-coding sequences in lamprey likely play a role in germ cell development and in the adult germline (Smith, et al. 2012; Wang, et al. 2012).

To gain further insight into the functional role of eliminated genes, we sought to validate candidate deletions and examine their transcriptional patterns in various developmental stages, including those encompassing PGR. Here we present new data revealing patterns of gene expression, DNA loss, and comparative functional information from 44 genes that can be definitively classified as eliminated from somatic cell lineages during PGR. Analysis of transcriptional data indicates that a large fraction of eliminated

genes are differentially expressed throughout lamprey embryogenesis, particularly during the stages surrounding PGR and the maternal-to-zygotic transition (MZT). Altogether, these analyses provide insight into the patterns of DNA loss over the time course of PGR and the evolutionary/developmental logic underlying the programmatic elimination of DNA from somatic cell lineages. Moreover, these data corroborate previous studies proposing that eliminated genes have conserved roles in germline development and oncogenesis (when misexpressed), and expand these studies by identifying several new genes that likely contribute to these critical biological processes (Smith, et al. 2012).

SECTION 3-
MATERIALS AND METHODS

Animals

All animals were obtained from the Lake Michigan population via the Great Lakes Fisheries Commission (GLFC) and maintained under University of Kentucky IACUC protocol number 2011-0848. Animals were euthanized by immersion in MS-222 (150 µg/mL), dissected, and tissues were immediately snap-frozen for the isolation of DNA from adult germline and somatic tissues.

Lamprey Embryos

In vitro fertilizations were performed with sexually mature adult animals. Eggs and sperm were collected in crystallization dishes and allowed to incubate in 10% Holtfreter's solution for 10 minutes to permit fertilization (Nikitina, et al. 2009). After visually confirming activation, embryos were rinsed in distilled water to remove excess sperm and maintained in 10% Holtfreter's solution at 18° C throughout development. At days 1, 2, 2.5, 3, 4 and 5 post-fertilization, embryos were collected in 1.7 mL centrifuge tubes and snap frozen for subsequent RNA extractions. Pools of embryos used for these analyses were obtained from several independent *in vitro* fertilizations. Pools used for RNAsequencing (fertilizations generated in 2012) were also independent of those used in Nanostring and qPCR experiments (fertilizations generated in 2015).

RNA sequencing

Total RNA was isolated from lamprey embryos using Trizol extraction. RNA quality was assessed on the Agilent 2100 Bioanalyzer (Agilent Technologies) and samples with RNA Integrity Number (RIN) > 8 were sent to the HudsonAlpha Genomic Services Lab (HudsonAlpha, Huntsville AL) for paired-end sequencing on the Illumina platform. Raw reads were assembled using Trinity (trinityrnaseq_r2013-02-25) using default parameters and integrated quality clipping with Trimmomatic (Bolger, et al. 2014).

Screening for Somatically-Eliminated Fragments

To identify sequences eliminated during PGR, transcripts were filtered based on computational evidence for somatic elimination. Candidate germline-specific transcripts were identified from the transcriptome assembly by aligning contigs to shotgun sequence datasets from germline (SRX025555) and somatic (liver, AEFG01) whole genome shotgun datasets using Blast (Altschul, et al. 1990). To provide homology information, reads were aligned to human RefSeq proteins using BLASTx (Altschul, et al. 1990). Transcripts that aligned to the germline sequence dataset (>98% identical over >100 base pairs) and failed to align to the somatic dataset at these same thresholds were selected for secondary screening. Secondary screening was performed to filter transcriptome assembly artifacts based on transcript abundance estimates generated by RSEM v1.2.9 (Li and Dewey 2011). To target germline-specific transcripts with a broad range of predicted expression during development, 480 transcripts with the highest combined average expression in testes and embryos and 96 transcripts with high expression in testes and low expression in embryos were selected for PCR validation.

PCR validation assays were performed in DNA from blood and testes in order to further evaluate predicted germline-specific regions. To assess whether patterns of deletion were consistent across somatic tissues, PCR amplifications were also performed using DNA from liver, kidney, fin and muscle. Oligonucleotide primer pairs were designed for the genomic sequences corresponding to candidate germline-specific reads using Primer3 and used to prime PCR reactions under the following amplification conditions [1 ng of DNA, 50 ng of each primer, .3 U Taq polymerase, 5X PCR buffer, and 200 mM each of dATP, dCTP, dGTP, and dTTP; oligonucleotide sequences and optimal thermal cycling conditions are provided in Appendix I] (Rozen and Skaletsky 2000). The DNAs used in these reactions were extracted from testes, liver, kidney, fin, muscle, and blood collected from two individuals (Males 12 and 13) using standard phenol/chloroform extraction (Sambrook, et al. 2006).

Sequencing PCR Products

For all 44 validated genes, PCR products corresponding to testes DNA from males 12 and 13 were sequenced. For sequences that also amplified in somatic tissues, PCR products corresponding to muscle DNA from male 12 and blood DNA from male 13 were sequenced. Muscle and blood DNA were chosen due to their consistent levels of relatively high background amplification. Samples for which there was no apparent background amplification in muscle or blood were not sequenced. In total, sequence data was generated for PCR products from 88 testes bands, 41 muscle bands, and 37 blood bands. Prior to sequencing, PCR products were purified with ExoSAP-IT (Affymatrix) according to the manufacturer's instructions. 4ul of purified template was added to 1ul of

primer mix (5pM/ul) and sequenced on the ABI3730 at the Advanced Genetic Technologies Center at the University of Kentucky. Sequence data was assembled using SeqMan Pro v13.0 (DNASTAR) with the ProAssembler algorithm. SNP Discovery parameters were as follows: Minimum Score at SNP: 20; Minimum Neighborhood Score: 0 (default); Minimum Neighborhood Window: 0 (default); Heterozygous Peak Threshold: 50%; without Strict Base Matching. These data are included in Appendix I.

Nanostring Gene Expression Analysis

The Nanostring nCounter Gene Expression Assay (Nanostring Technologies, Seattle, USA) was used to estimate expression of the 44 validated germline-specific genes throughout embryogenesis and in the adult germline. The transcript sequences to which the validated germline-specific genes aligned and the sequences for 6 control genes were submitted to Nanostring Technologies (Seattle, USA) for custom CodeSet design. Due to sequence similarity between comp266794_c0_seq1 and its variant, comp266794_c0_seq2, only a shared probe could be designed for these targets. The final codeset included 49 capture probes; 43 designed from the transcript sequences corresponding to validated germline-specific genes and 6 designed from control genes (Appendix I).

The Direct-zol RNA MiniPrep (Zymo Research) was used to isolate RNA from pools of snap-frozen embryos collected at days 1-5 post fertilization (D1, D2, D2.5, D3, D4, and D5; n=3 biological replicates each) and adult testes (n = 3 technical replicates). RNA was quantified on the Nanodrop 2000 Spectrophotometer (Thermo Scientific) and RNA integrity was assessed using the Agilent 2100 BioAnalyzer (Agilent Technologies).

RNA samples with RNA Integrity Number (RIN) ≥ 8 were used in the Nanostring nCounter Assay according to the manufacturer's protocol.

Count data for all genes were analyzed by the nSolver Analysis software v2.5 (Nanostring Technologies, Seattle, USA). Background subtraction was performed using the geometric mean of internal negative controls. Normalization was performed using the geometric mean of internal positive controls and the mean of two endogenous housekeeping genes (homologs of EF1A_A and EF1A_B), which show consistent expression in embryos and testes, respectively.

Real-Time PCR

Quantitative real-time PCR was used to measure the relative abundance of a subset (n=23) of validated germline-specific genes during embryogenesis and in germline (testes) and somatic (blood) tissues. This subset of genes consisted of 20 sequences that aligned to human RefSeq proteins and 3 other genes that exhibited patterns of dynamic expression during development based on RNAseq estimates. DNA was extracted from blood, testes, and D1, D2, D2.5, D3, D4 and D5 embryos using standard phenol-chloroform extraction (Sambrook, et al. 2006). Real-time PCR was performed on a CFX96 Touch Real-Time PCR Detection System (Bio-Rad) using Sso Advanced Universal SYBR Green Super Mix, ~1 ng DNA and 50 ng of each primer. Thermal cycling conditions were 3 min initial denaturation at 95° C, followed by 50 cycles of 95° C for 10 s and 60° C for 30 s. Relative expression estimates were calculated using the ΔΔct method of relative quantification using two control sequences shown to amplify

consistently in germline, somatic, and embryonic tissues (Appendix I). Final ΔΔct values were standardized to testes values.

Differential Expression Analysis

Differential expression analysis of RNAseq data was performed on the full RNAseq datasets for embryos and testes using EBseq v1.1.5 (Leng, et al. 2013). EBseq employs an empirical Bayes method to identify differentially expressed genes and isoforms across two or more conditions in an RNAseq experiment (Leng, et al. 2013). Fold change was calculated between testes and embryos, between D1 embryos and each subsequent embryonic stage (D1 vs. D2, D1 vs. D2.5, etc.), and between adjacent embryonic time points (D1 vs. D2, D2 vs. D2.5, etc.). False discovery rate (FDR) was controlled at 0.05. Differential expression analysis of Nanostring data was performed using the nSolver Analysis software v2.5 (Nanostring Technologies, Seattle, USA). Following background subtraction and normalization, fold change estimates were calculated between testes and embryos, between D1 embryos and each subsequent embryonic stage, and between adjacent embryonic time points. Those with p-values < 0.05 were classified as differentially expressed. Hierarchical clustering of RNAseq and Nanostring data was performed using Ward multivariate two-way clustering (SAS Institute Inc. 1989-2007).

Gene Ontology Analysis

To assess whether any functional categories were overrepresented among deleted genes, we performed a PANTHER Overrepresentation Test (release 20160321; PANTHER version 10.0 Released 2015-05-15) (Mi, et al. 2016). A non-redundant list of the human

RefSeq proteins identified for our validated genes (n=17) was compared to the list of non-redundant human RefSeq proteins (n=311) in the entire set of screened candidates using the PANTHER Pathways Annotation data set.

SECTION 4-
RESULTS AND DISCUSSION

RNA sequencing

To characterize transcripts that are expressed in germline tissues and during the time course of PGR, RNAseq data were generated for pools of embryos that were collected at D1 (day 1), D2, D2.5, D3, D4 and D5 post-fertilization. These time points correspond to the 24-32 cell stage, the blastula stage, the midblastula transition, dorsal cone formation, gastrulation, and neural groove formation, respectively, and coincide with major changes in gene expression during vertebrate development. These sequencing runs yielded ~130.4 million 100bp paired-end reads, which were relatively evenly distributed across time points (D1: 25 million, D2: 19.9 million, D2.5: 20.7 million, D3: 24 million, D4: 22.9 million, D5: 17.9 million). These reads and previously published data from testes (SRX104180) and later embryonic stages (SRX110029 - 35) were used to assemble a reference transcriptome. The final transcriptome assembly consists of 74,938 unique transcripts and 70,975 additional subsequences (i.e. isoforms or assembly variants). Estimates of transcript abundance in testes and individual embryonic stages were recalculated using RSEM v1.2.9 (Li and Dewey 2011). Alignment of assembled transcripts to human RefSeq proteins generated a total of 66,948 blast hits ($E < 1e^{-4}$), though several transcripts aligned to the same human homolog. In total, these transcripts yielded alignments to 13,159 unique human genes.

Screening for Somatically-Eliminated Fragments

To identify transcripts within eliminated regions, assembled contigs were filtered based on two criteria: 1) computational evidence for somatic elimination (i.e. transcript presence in lamprey germline assembly and absence in the somatic WGS assembly), and 2) average expression values for each transcript in testes and D1-5 embryos as estimated by RSEM (Li and Dewey 2011). These criteria identified a total of 576 transcribed sequences that were considered candidates for somatic elimination during PGR. PCR-validation of these candidate genomic intervals identified 44 genes that are present in lamprey germline DNA (testes) but absent or depleted in DNA that was extracted from a panel of somatic tissues (liver, kidney, fin, muscle, and blood) (Figures 1 and 2). The remaining genes yielded similar amplification patterns in germline and somatic tissues, and thus did not meet our operational criteria for classification as somatically-eliminated. These may represent portions of genes that are present in the soma but were not sampled by somatic genome shotgun sequencing due to random sampling processes, genes with highly-similar paralogs in germline and soma, or biases in the shotgun cloning/sequencing approach.

While levels of amplification between germline and somatic tissues as a whole are strikingly different, background amplification was observed for some primers and tissues. In particular, several faint bands were observed in amplified samples from male 12 muscle DNA and male 13 blood DNA. To test whether these background bands represent amplification of the targeted germline candidate region, paralogous sequence, or amplification artifacts, we sequenced PCR products corresponding to all background bands from these two samples and their matched germline amplicon. Of the 78 amplicons

targeted for sequencing, 25 yielded sequence data of sufficient quality to permit comparison to their corresponding germline sequence. For all of these samples, the sequences for blood and muscle were identical to their corresponding germline bands (Appendix I). This observation suggests that background amplicons are likely derived from rare cells that retain some fraction of the germline genome yet reside in somatic tissues, rather than representing inefficient amplification of paralogous sequences or amplification artifacts. While characterization of these specific cell types is likely to be challenging, it is interesting to speculate that these might represent cells that are capable of adopting a wider range of cell fates later in development (i.e. resident stem cells). Alternately, amplification of germline-specific sequences might reflect the presence of cell-free DNA that traces its origin to the differentiated germline.

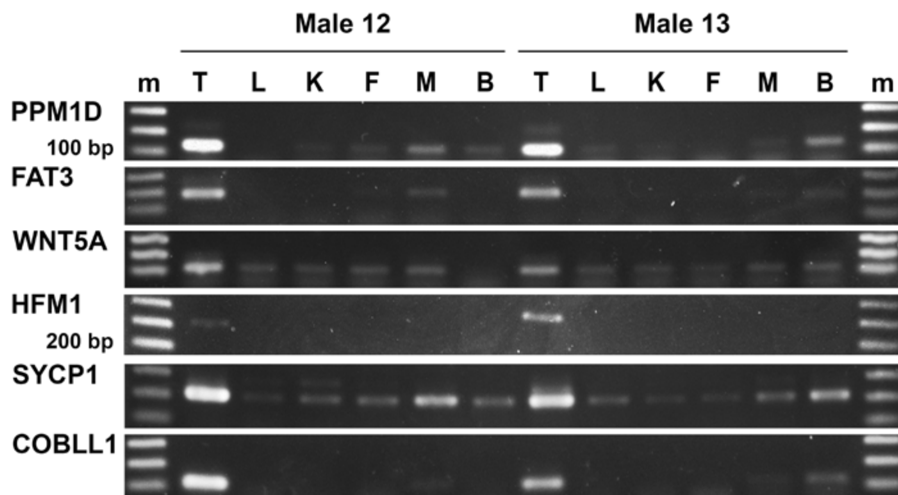

Figure 1 – Subset of validated deletions with homology information. Validated sequences show amplification in testes DNA but no or reduced amplification in DNA from a panel of somatic tissues (liver, kidney, fin, muscle, blood) from two animals (Male 12 and 13). T = Testes, L = Liver, K = Kidney, F = Fin, M = Muscle, B = Blood, m = 100 bp DNA ladder. The lowest band under "m" corresponds to the 100bp band of the DNA ladder in all panels except for that of HFM1, which corresponds to the 200bp band. Amplified loci are labeled with their corresponding human homolog.

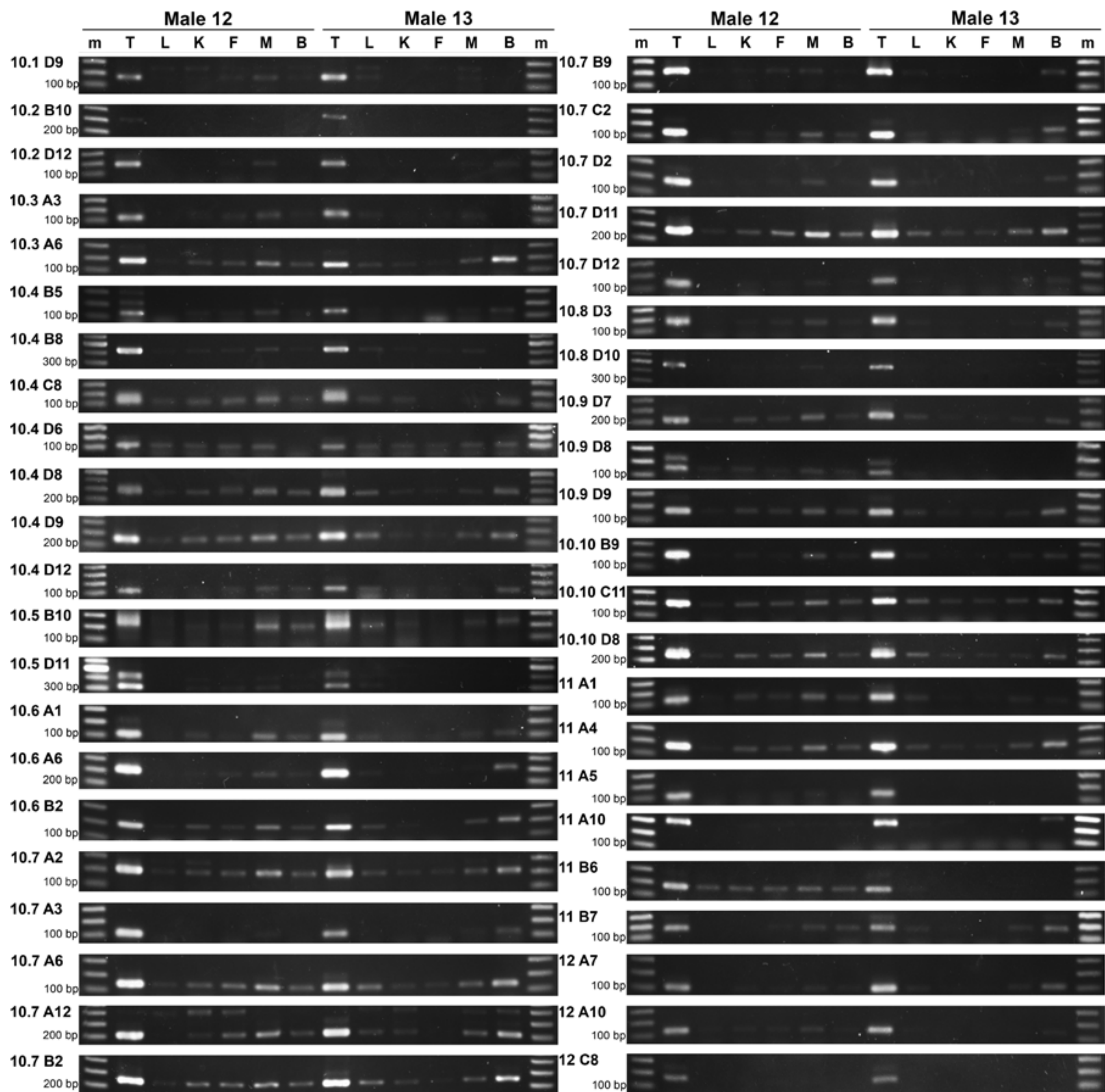

Figure 2: Full set of validated eliminations (n=44). Validated sequences show amplification in testes DNA but no or reduced amplification in DNA from a panel of somatic tissues (liver, kidney, fin, muscle, blood) from two animals (Male 12 and 13). T = Testes, L = Liver, K = Kidney, F = Fin, M = Muscle, B = Blood, m = 100 bp DNA ladder. Alphanumeric labels correspond to sequence identifiers given in Appendix I.

Real-Time PCR

Of our 44 validated germline-specific sequences, 20 aligned to human RefSeq proteins. These 20 genes, in addition to 3 other genes that exhibited dynamic patterns of expression during development, were chosen for qPCR analysis to characterize the pattern of DNA elimination and the time course of PGR. The abundance of germline DNA sequences relative to retained sequences was estimated by quantitative real-time PCR in genomic DNA from embryos (D1-5), testes, and blood. Results of these analyses reveal a relatively uniform pattern of elimination across tissues, consistent with previous findings that indicate that DNA loss coincides with the MZT (Figure 3) (Smith, et al. 2009). At D1, the majority of surveyed sequences are at their highest relative abundance, closely resembling estimates in the germline (testes). By D2.5, the approximate timing of the MZT, the abundance of all germline-specific DNAs has substantially decreased, and by D5 most germline-specific DNAs are essentially absent, similar to estimates in somatic tissue (blood).

While this pattern of deletion largely correlates with that of the previously described germline-limited fragment *Germ1,* the presence of a slight excess of residual DNA between D2.5 and D5 suggests the existence of a transitional period between the PGR event at D2.5 and the physical removal of targeted DNA apparent by D5 (Smith, et al. 2012). This observation is particularly interesting in light of previous studies suggesting the presence of abundant DNA breaks in post-gastrulation embryonic stages (Smith, et al. 2009).

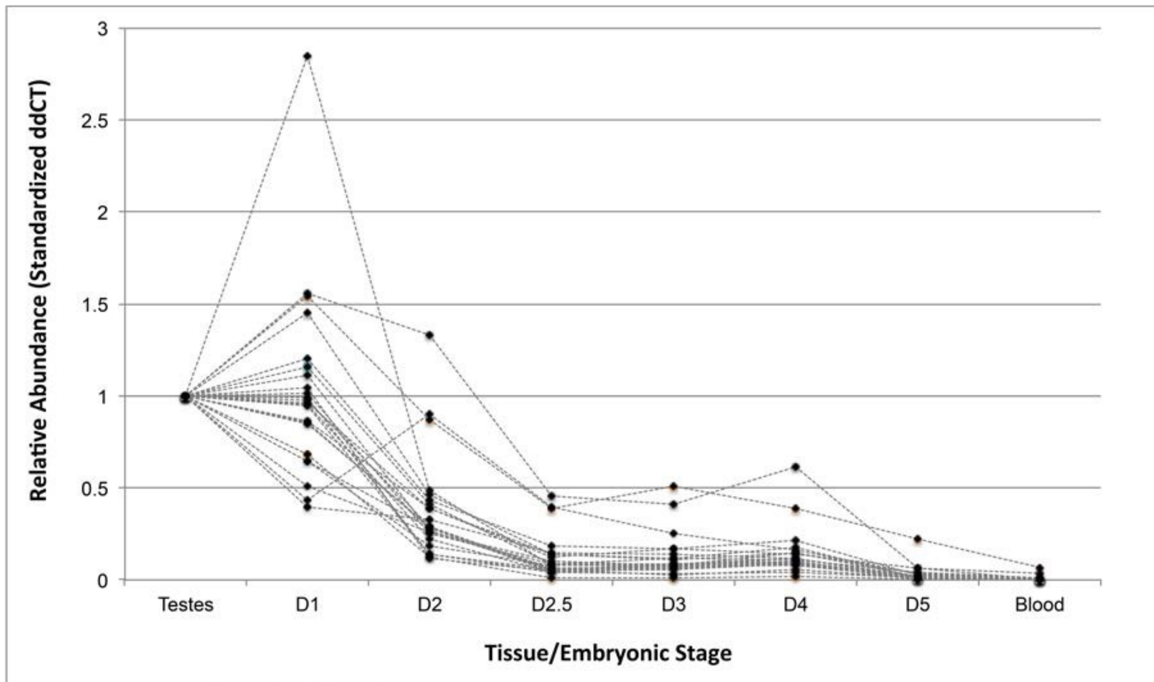

Figure 3 - Relative abundance of a subset (n=23) of somatically-depleted genes based on qPCR in genomic DNA. Relative abundance estimates were calculated using the ΔΔct method of relative quantification using two sequences that amplify consistently in germline, somatic, and embryonic tissues as controls (Appendix I, comp4257946_c0_seq1 and comp4249214_c0_seq2). The ΔΔct values for each gene are standardized to ΔΔct values from testes. D1, D2, D2.5, D3, D4, and D5 correspond to days 1-5 post-fertilization.

Nanostring Gene Expression Analysis

Closer examination of our RNAseq analysis revealed that the effective read counts of the majority of our validated genes is ~1 across most embryonic time points, presumably due to low transcription levels in early embryonic stages. However, patterns of gene expression appeared to be generally consistent with the timing of the activation of zygotic transcription, with the number of transcriptionally active genes increasing after D2.5 (Figure 4). Differential expression analysis identified 33 genes within the total validated subset that show significant differential expression between tissues (Figure 4). Because eliminated genes were generally characterized by low effective read counts in

our RNAseq datasets, we cross-validated our analyses using the Nanostring nCounter Gene Expression Assay (Nanostring Technologies, Seattle, USA).

Nanostring probes were designed to target transcripts with coding sequences that overlap validated germline genomic intervals (Appendix I). Because two germline reads aligned to variants of the same transcript, for which only one unique probe could be designed, only 43 unique transcript sequences were used for probe design. Nanostring abundance estimates generated average normalized transcript counts of ~87 across embryonic samples. Temporal clustering of transcript counts reveals that most genes exhibit an abrupt increase in transcription after D2 (Figure 5). This broad-scale pattern presumably reflects major changes in gene expression that accompany the onset of zygotic gene expression at the MZT (Yartseva and Giraldez 2015). Fold change analyses identified 42 of 43 genes that display significant ($p < 0.05$) differential expression between tissues. Of these, 42 are differentially expressed relative to testes, 25 are differentially expressed between D1 and later embryonic time points (D2-5), and 3 are differentially expressed between adjacent embryonic time points (Figure 5). Together, results of both RNAseq and Nanostring experiments reveal that eliminated genes are differentially transcribed during lamprey development, suggesting that many of the genes targeted for elimination by PGR may play a functional role in embryogenesis.

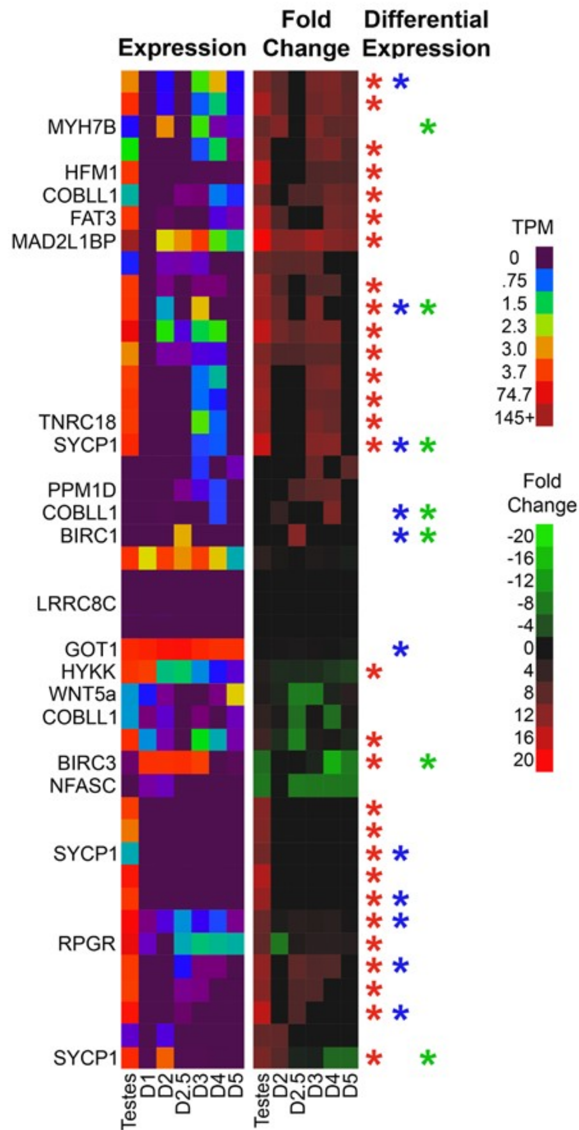

Figure 4 - Transcriptional profile and differential expression of somatically-depleted genes based on RNAseq estimates.
(Left) Transcript abundance as estimated by RNAseq. Data are presented using a spectral color theme to highlight the broad range of expression values across genes, tissues, and embryonic time points. Purple indicates low expression and red indicates high expression. Transcripts are labeled with their corresponding human homolog. (Center) Log_2 fold changes in expression relative to D1. (Right) Red asterisks indicate genes that are differentially expressed between testes and any embryonic time point (D1-5), blue asterisks indicate genes that are differentially expressed between D1 and any subsequent time point (D1 vs. D2, D1 vs. D2.5, etc.), and green asterisks indicate genes that are differentially expressed between adjacent time points (D1 vs. D2, D2 vs. D2.5, etc.). D1, D2, D2.5, D3, D4, and D5 correspond to days 1-5 post-fertilization.

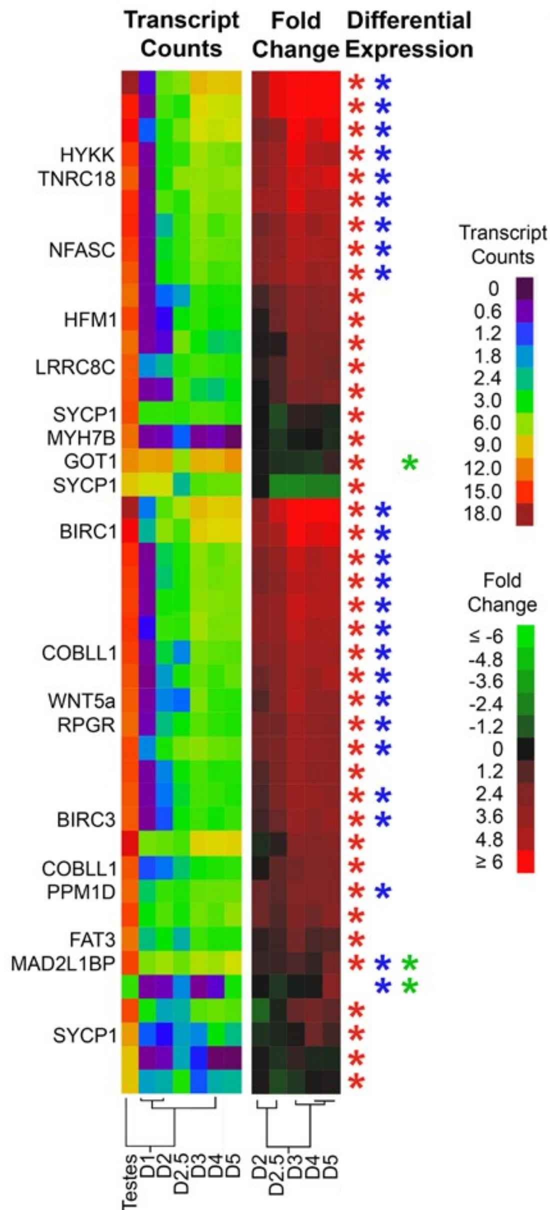

Figure 5 - Transcription and differential expression of somatically-depleted genes based on Nanostring estimates. (Left) Transcript abundance as estimated by Nanostring count values. Data are presented using a spectral color theme to highlight the broad range of expression values across genes, tissues, and embryonic time points. Purple indicates low expression and red indicates high expression. Transcripts are labeled with their corresponding human homolog. (Center) Log_2 fold changes in expression relative to D1. (Right) Red asterisks indicate genes that are differentially expressed between testes and any embryonic time point (D1-5), blue asterisks indicate genes that are differentially expressed between D1 and any subsequent time point (D1 vs. D2, D1 vs. D2.5, etc.), and green asterisks indicate genes that are differentially expressed between adjacent time points (D1 vs. D2, D2 vs. D2.5, etc.). D1, D2, D2.5, D3, D4, and D5 correspond to days 1-5 post-fertilization.

Gene Ontology

To shed light on the possible function of eliminated sequences, we performed a statistical overrepresentation test using PANTHER (Huang da, et al. 2009a, b; Maglott, et al. 2011; Mi, et al. 2016). Within the full set of genes screened for germline-specificity (n=576), 311 had non-redundant hits to the human RefSeq protein database and were used as the reference list in our overrepresentation test. Within the validated subset of germline-specific genes (n=44), 20 had hits to the human RefSeq protein database, 17 of which were non-redundant and were used as the analyzed list in our overrepresentation test. 294 protein IDs from the reference list and 15 protein IDs from the analyzed list were mapped to the PANTHER database.

Analysis of these lists using the PANTHER Pathways annotation data set indicates that genes in the Wnt and Cadherin signaling pathways are the most overrepresented in our validated subset of genes, accounting for 20% and 13.33% of our validated genes, respectively. However, the majority of the genes in each list were unclassified (n=244 reference genes and n=9 validated genes). As such, overrepresentation estimates are based on only 6 unique protein IDs from the analyzed list, and several of these appear in multiple categories. For instance, the two genes that are assigned to the Cadherin signaling pathway are also assigned to the Wnt signaling pathway, resulting in only 3 unique genes falling under these annotation categories (WNT-5A, Myosin-7B, and Protocadherin Fat 3). Other pathways that this analysis indicates are overrepresented among our validated genes include the p53 pathway and p53 pathway feedback loop 2, inflammation mediated by chemokine and cytokine signaling, asparagine and aspartate biosynthesis, and nicotinic acetylcholine receptor

signaling. However, these overrepresentation values are based on the presence of one gene in each category, many of which occur in multiple categories.

To gain further insight on the function of eliminated genes, including those that were unclassified in our PANTHER Pathways analysis, we manually examined gene-specific datasets curated at *NCBI Gene* to gain further perspective on the known functions of their human homologs (Maglott, et al. 2011). Consistent with the results of our statistical overrepresentation test, this broad-based functional information suggests that several of our validated germline-specific genes are involved in Wnt and Cadherin signaling pathways, both of which are strongly implicated in oncogenesis (Nelson and Nusse 2004; Polakis 2000). Among our validated deletions are sequences with homology to HFM1 ATP-dependent DNA helicase homolog (HFM1), which is expressed primarily in germ cells, and human cancer-testis antigen synaptonemal complex protein 1 (SYCP1) (Tanaka, et al. 2006; Yi, et al. 2007). Notably, expression of cancer-testis antigens is normally limited to the germline, and is only observed within somatic tissues in the context of oncogenesis (Fratta, et al. 2011). Moreover, previous experiments have shown that ectopic expression of germline-limited genes contributes to tumorigenesis in other species (Drosophila and Hydractinia) (Fratta, et al. 2011; Janic, et al. 2010; Millane, et al. 2011). Other deleted genes include protein phosphatase Mg2+/Mn2+ dependent 1D (PPM1D), hydroxylysine kinase (HYKK), wingless-type MMTV integration site family member 5A (WNT-5a), and MAD2L1 binding protein (MAD2L1BP), all of which are implicated in oncogenesis or tumorigenesis either in their normal state or when misexpressed (Date, et al. 2013; Lin, et al. 2014; Wang and Liu 2014; Zhang, et al. 2014). These results are consistent with homology information for previously validated

deletions, which include sequences homologous to cancer-testis antigen 68 and WNT7a/b (Smith, et al. 2012).

Other genes in our validated subset tend to fall into categories related to apoptosis and development. In this context, it seems likely that apoptosis regulators such as baculoviral IAP repeat containing 1 (BIRC1) and baculoviral IAP repeat containing 3 (BIRC3) may be targeted for deletion to prevent possible misregulation of apoptosis and subsequent tumorigenesis in somatic cells (Allam, et al. 2015; Bai, et al. 2014; Davoodi, et al. 2010). Yet a small number of deleted genes have no overt annotations related to oncogenesis or other disease states. These include retinitis pigmentosa GTPase regulator (RPGR: eye photoreceptor development), myosin, heavy chain 7B, cardiac muscle, beta (MYHB: skeletal and cardiac muscle development), glutamic-oxaloacetic transaminase 1 (GOT1: amino acid metabolism), and leucine-rich repeat containing 8 family member C (LRRC8C: adipocyte and immune cell differentiation) (Esposito, et al. 2013; Gakovic, et al. 2011; Murga-Zamalloa, et al. 2010; Shen, et al. 2011; Tominaga, et al. 2004) It seems plausible that this relatively small fraction of genes might have evolved germline-specific functions in the lamprey lineage or that gnathostome genomes have evolved to deploy this subset of genes during the development and maintenance of somatic cell lineages. In this regard, it is notable that some eliminated genes have human homologs that play roles in nervous system development, such as FAT atypical cadherin 3 (FAT3: similar to the Drosophila tumor suppressor gene FAT), cordon-bleu WH2 repeat protein like 1 (COBLL1), neurofascin (NFASC), RPGR, and trinucleotide repeat containing 18 (TNRC18) (Buttermore, et al. 2012; Carroll, et al. 2003; Gakovic, et al. 2011; Koticha, et al. 2005; Margolis, et al. 1997; Mitsui, et al. 2002; Murga-Zamalloa, et al. 2010). As

these genes are known to be involved neurogenesis in many other vertebrate and non-vertebrate species, it is unclear how genes with these functions may have become dispensable in the lamprey somatic genome.

While these analyses provide insight into possible functions and expression patterns of germline-specific genes over early embryonic development, it is important to note that transcription at the whole-embryo level is a function of both transcriptional regulation and DNA elimination. We reasoned that comparing the pattern of genomic DNA elimination to our transcriptional data might provide greater insight into the relationship between gene abundance and transcription throughout development. While our qPCR analyses show that the predicted pattern of elimination in genomic DNA holds true for most surveyed genes, a small number of sequences show greater variation in abundance across embryogenesis. These include genes with homology to FAT3, BIRC1, and BIRC3. The elimination of these genes seems to occur over a more protracted period than the majority of other eliminated genes. Comparing this pattern of elimination to our estimates of gene expression, we find that their transcript abundance is highest at D3-5. We speculate that the delayed elimination of these genes might reflect their function, perhaps indicating that they play some role in the early differentiation of somatic lineages. As such, it may be more constructive to conceptualize eliminated genes as somatically-depleted, rather than "germline-specific" per se.

SECTION 5-
CONCLUSION

In depth characterization of lamprey embryonic transcriptomes has permitted the identification of a larger catalog of somatically-depleted genes and provides critical evidence supporting the implicit hypothesis that eliminated genes are biologically functional (i.e. they are transcribed). Somatically-depleted genes identified here may be useful markers of eliminated material in future experiments designed to further characterize PGR. These data corroborate previous findings in studies of the nematode *A. suum* that identify single-copy transcribed genes related to germline maintenance and development in eliminated DNA, providing further evidence that PGR may be a bona fide mechanism of germline gene regulation. However, these studies also identify several eliminated genes with homologs that have no known germline or pluripotent functions, perhaps suggesting additional/alternative roles of PGR. One such non-exclusive role may be the reduction of genome complexity, where repetitive DNA and/or genes that are non-essential in somatic tissues are eliminated from the soma but retained in the germline for reproductive purposes. Studies aimed at identifying more contiguous eliminated chromosomal fragments and their sequence content will likely provide greater perspective on the content of eliminated DNA and the biological function of PGR. Altogether, these studies contribute to a growing body of evidence that suggests PGR serves as a permanent gene silencing mechanism that prevents misexpression of several genes related to germline development, pluripotency, and oncogenesis, including genes for which such functions have yet to be defined in human or other mammalian models.

REFERENCES

Allam R, Maillard MH, Tardivel A, Chennupati V, Bega H, Yu CW, Velin D, Schneider P, Maslowski KM 2015. Epithelial NAIPs protect against colonic tumorigenesis. J Exp Med 212: 369-383. doi: 10.1084/jem.20140474

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ 1990. Basic local alignment search tool. J Mol Biol 215: 403-410. doi: 10.1016/S0022-2836(05)80360-2

Bachmann-Waldmann C, Jentsch S, Tobler H, Muller F 2004. Chromatin diminution leads to rapid evolutionary changes in the organization of the germ line genomes of the parasitic nematodes A. suum and P. univalens. Mol Biochem Parasitol 134: 53-64.

Bai L, Smith DC, Wang S 2014. Small-molecule SMAC mimetics as new cancer therapeutics. Pharmacol Ther 144: 82-95. doi: 10.1016/j.pharmthera.2014.05.007

Bolger AM, Lohse M, Usadel B 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 30: 2114-2120. doi: 10.1093/bioinformatics/btu170

Buttermore ED, Piochon C, Wallace ML, Philpot BD, Hansel C, Bhat MA 2012. Pinceau organization in the cerebellum requires distinct functions of neurofascin in Purkinje and basket neurons during postnatal development. J Neurosci 32: 4724-4742. doi: 10.1523/JNEUROSCI.5602-11.2012

Carroll EA, Gerrelli D, Gasca S, Berg E, Beier DR, Copp AJ, Klingensmith J 2003. Cordon-bleu is a conserved gene involved in neural tube formation. Dev Biol 262: 16-31.

Date DA, Burrows AC, Venere M, Jackson MW, Summers MK 2013. Coordinated regulation of p31(Comet) and Mad2 expression is required for cellular proliferation. Cell Cycle 12: 3824-3832. doi: 10.4161/cc.26811

Davoodi J, Ghahremani MH, Es-Haghi A, Mohammad-Gholi A, Mackenzie A 2010. Neuronal apoptosis inhibitory protein, NAIP, is an inhibitor of procaspase-9. Int J Biochem Cell Biol 42: 958-964. doi: 10.1016/j.biocel.2010.02.008

Degtyarev S, Boykova T, Grishanin A, Belyakin S, Rubtsov N, Karamysheva T, Makarevich G, Akifyev A, Zhimulev I 2004. The molecular structure of the DNA fragments eliminated during chromatin diminution in Cyclops kolensis. Genome Res 14: 2287-2294. doi: 10.1101/gr.2794604

Drouin G 2006. Chromatin diminution in the copepod Mesocyclops edax: diminution of tandemly repeated DNA families from somatic cells. Genome 49: 657-665. doi: 10.1139/g06-022

Esposito T, Sampaolo S, Limongelli G, Varone A, Formicola D, Diodato D, Farina O, Napolitano F, Pacileo G, Gianfrancesco F, Di Iorio G 2013. Digenic mutational inheritance of the integrin alpha 7 and the myosin heavy chain 7B genes causes congenital myopathy with left ventricular non-compact cardiomyopathy. Orphanet J Rare Dis 8: 91. doi: 10.1186/1750-1172-8-91

Gakovic M, Shu X, Kasioulis I, Carpanini S, Moraga I, Wright AF 2011. The role of RPGR in cilia formation and actin stability. Hum Mol Genet 20: 4840-4850. doi: 10.1093/hmg/ddr423

Goday C, Esteban MR 2001. Chromosome elimination in sciarid flies. Bioessays 23: 242-250. doi: 10.1002/1521-1878(200103)23:3<242::AID-BIES1034>3.0.CO;2-P

Goto Y, Kubota S, Kohno S 1998. Highly repetitive DNA sequences that are restricted to the germ line in the hagfish Eptatretus cirrhatus: a mosaic of eliminated elements. Chromosoma 107: 17-32.

Hanahan D, Weinberg RA 2000. The hallmarks of cancer. Cell 100: 57-70.

Huang da W, Sherman BT, Lempicki RA 2009a. Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. Nucleic Acids Res 37: 1-13. doi: 10.1093/nar/gkn923

Huang da W, Sherman BT, Lempicki RA 2009b. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nat Protoc 4: 44-57. doi: 10.1038/nprot.2008.211

Itoh Y, Kampf K, Pigozzi MI, Arnold AP 2009. Molecular cloning and characterization of the germline-restricted chromosome sequence in the zebra finch. Chromosoma 118: 527-536. doi: 10.1007/s00412-009-0216-6

Koticha D, Babiarz J, Kane-Goldsmith N, Jacob J, Raju K, Grumet M 2005. Cell adhesion and neurite outgrowth are promoted by neurofascin NF155 and inhibited by NF186. Mol Cell Neurosci 30: 137-148. doi: 10.1016/j.mcn.2005.06.007

Kubota S, Takano J, Tsuneishi R, Kobayakawa S, Fujikawa N, Nabeyama M, Kohno S 2001. Highly repetitive DNA families restricted to germ cells in a Japanese hagfish (Eptatretus burgeri): a hierarchical and mosaic structure in eliminated chromosomes. Genetica 111: 319-328.

Leng N, Dawson JA, Thomson JA, Ruotti V, Rissman AI, Smits BM, Haag JD, Gould MN, Stewart RM, Kendziorski C 2013. EBSeq: an empirical Bayes hierarchical model for inference in RNA-seq experiments. Bioinformatics 29: 1035-1043. doi: 10.1093/bioinformatics/btt087

Li B, Dewey CN 2011. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. BMC Bioinformatics 12: 323. doi: 10.1186/1471-2105-12-323

Lin L, Liu Y, Zhao W, Sun B, Chen Q 2014. Wnt5A expression is associated with the tumor metastasis and clinical survival in cervical cancer. Int J Clin Exp Pathol 7: 6072-6078.

Maglott D, Ostell J, Pruitt KD, Tatusova T 2011. Entrez Gene: gene-centered information at NCBI. Nucleic Acids Res 39: D52-57. doi: 10.1093/nar/gkq1237

Margolis RL, Abraham MR, Gatchell SB, Li SH, Kidwai AS, Breschel TS, Stine OC, Callahan C, McInnis MG, Ross CA 1997. cDNAs with long CAG trinucleotide repeats from human brain. Hum Genet 100: 114-122.

Mitsui K, Nakajima D, Ohara O, Nakayama M 2002. Mammalian fat3: a large protein that contains multiple cadherin and EGF-like motifs. Biochem Biophys Res Commun 290: 1260-1266. doi: 10.1006/bbrc.2002.6338

Muller F, Walker P, Aeby P, Neuhaus H, Felder H, Back E, Tobler H 1982. Nucleotide sequence of satellite DNA contained in the eliminated genome of Ascaris lumbricoides. Nucleic Acids Res 10: 7493-7510.

Murga-Zamalloa C, Swaroop A, Khanna H 2010. Multiprotein complexes of Retinitis Pigmentosa GTPase regulator (RPGR), a ciliary protein mutated in X-linked Retinitis Pigmentosa (XLRP). Adv Exp Med Biol 664: 105-114. doi: 10.1007/978-1-4419-1399-9_13

Nelson WJ, Nusse R 2004. Convergence of Wnt, beta-catenin, and cadherin pathways. Science 303: 1483-1487. doi: 10.1126/science.1094291

Niedermaier J, Moritz KB 2000. Organization and dynamics of satellite and telomere DNAs in Ascaris: implications for formation and programmed breakdown of compound chromosomes. Chromosoma 109: 439-452.

Nikitina N, Bronner-Fraser M, Sauka-Spengler T 2009. Culturing lamprey embryos. Cold Spring Harb Protoc 2009: pdb prot5122. doi: 10.1101/pdb.prot5122

Pigozzi MI, Solari AJ 1998. Germ cell restriction and regular transmission of an accessory chromosome that mimics a sex body in the zebra finch, Taeniopygia guttata. Chromosome Res 6: 105-113.

Pikor L, Thu K, Vucic E, Lam W 2013. The detection and implication of genome instability in cancer. Cancer Metastasis Rev 32: 341-352. doi: 10.1007/s10555-013-9429-5

Polakis P 2000. Wnt signaling and cancer. Genes Dev 14: 1837-1851.

Rozen S, Skaletsky H 2000. Primer3 on the WWW for general users and for biologist programmers. Methods Mol Biol 132: 365-386.

Sambrook J, Russell DW, Sambrook J. 2006. The condensed protocols from Molecular cloning : a laboratory manual. Cold Spring Harbor, N.Y.: Cold Spring Harbor Laboratory Press.

SAS Institute Inc. C, NC. 1989-2007. JMP®. Version 11.

Shen H, Damcott C, Shuldiner SR, Chai S, Yang R, Hu H, Gibson Q, Ryan KA, Mitchell BD, Gong DW 2011. Genome-wide association study identifies genetic variants in GOT1 determining serum aspartate aminotransferase levels. J Hum Genet 56: 801-805. doi: 10.1038/jhg.2011.105

Smith JJ, Antonacci F, Eichler EE, Amemiya CT 2009. Programmed loss of millions of base pairs from a vertebrate genome. Proc Natl Acad Sci U S A 106: 11212-11217. doi: 10.1073/pnas.0902358106

Smith JJ, Baker C, Eichler EE, Amemiya CT 2012. Genetic consequences of programmed genome rearrangement. Curr Biol 22: 1524-1529. doi: 10.1016/j.cub.2012.06.028

Smith JJ, Kuraku S, Holt C, Sauka-Spengler T, Jiang N, Campbell MS, Yandell MD, Manousaki T, Meyer A, Bloom OE, Morgan JR, Buxbaum JD, Sachidanandam R, Sims C, Garruss AS, Cook M, Krumlauf R, Wiedemann LM, Sower SA, Decatur WA, Hall JA, Amemiya CT, Saha NR, Buckley KM, Rast JP, Das S, Hirano M, McCurley N, Guo P, Rohner N, Tabin CJ, Piccinelli P, Elgar G, Ruffier M, Aken BL, Searle SM, Muffato M, Pignatelli M, Herrero J, Jones M, Brown CT, Chung-Davidson YW, Nanlohy KG, Libants SV, Yeh CY, McCauley DW, Langeland JA, Pancer Z, Fritzsch B, de Jong PJ, Zhu B, Fulton LL, Theising B, Flicek P, Bronner ME, Warren WC, Clifton SW, Wilson RK, Li W 2013. Sequencing of the sea lamprey (Petromyzon marinus) genome provides insights into vertebrate evolution. Nat Genet 45: 415-421, 421e411-412. doi: 10.1038/ng.2568

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. doi: 10.1007/s00412-010-0263-z

Tanaka K, Miyamoto N, Shouguchi-Miyata J, Ikeda JE 2006. HFM1, the human homologue of yeast Mer3, encodes a putative DNA helicase expressed specifically in germ-line cells. DNA Seq 17: 242-246.

Thomas PD, Campbell MJ, Kejariwal A, Mi H, Karlak B, Daverman R, Diemer K, Muruganujan A, Narechania A 2003. PANTHER: a library of protein families and subfamilies indexed by function. Genome Res 13: 2129-2141. doi: 10.1101/gr.772403

Tominaga K, Kondo C, Kagata T, Hishida T, Nishizuka M, Imagawa M 2004. The novel gene fad158, having a transmembrane domain and leucine-rich repeat, stimulates adipocyte differentiation. J Biol Chem 279: 34840-34848. doi: 10.1074/jbc.M312927200

Wang H, Liu ZD 2014. Genetic association between AGPHD1 variant and lung cancer risk. Cell Biochem Biophys 70: 1963-1968. doi: 10.1007/s12013-014-0157-4

Wang J, Davis RE 2014. Programmed DNA elimination in multicellular organisms. Curr Opin Genet Dev 27: 26-34. doi: 10.1016/j.gde.2014.03.012

Wang J, Mitreva M, Berriman M, Thorne A, Magrini V, Koutsovoulos G, Kumar S, Blaxter ML, Davis RE 2012. Silencing of germline-expressed genes by DNA elimination in somatic cells. Dev Cell 23: 1072-1080. doi: 10.1016/j.devcel.2012.09.020

Yartseva V, Giraldez AJ 2015. The Maternal-to-Zygotic Transition During Vertebrate Development: A Model for Reprogramming. Curr Top Dev Biol 113: 191-232. doi: 10.1016/bs.ctdb.2015.07.020

Yi M, Liao H, Huang S, Xie X, Luo G 2007. [Expression of cancer-testis antigen SCP-1 mRNA in human nasopharyngeal carcinoma]. Lin Chung Er Bi Yan Hou Tou Jing Wai Ke Za Zhi 21: 343-345.

Zhang C, Chen Y, Wang M, Chen X, Li Y, Song E, Liu X, Kim S, Peng H 2014. PPM1D silencing by RNA interference inhibits the proliferation of lung cancer cells. World J Surg Oncol 12: 258. doi: 10.1186/1477-7819-12-258

VITA

Stephanie A. Bryant

Education

Master of Science in Biology, University of Kentucky	2016
Bachelor of Science in Biology, Western Kentucky University	2011

Professional Positions

Research Analyst, Markey Cancer Center, University of Kentucky	2016
Graduate Teaching Assistant, University of Kentucky	2015-2016
Graduate Research Assistant, University of Kentucky	2013-2015
Undergraduate Research Assistant, Western Kentucky University	2010-2011

Fellowships and Awards

American Society for Microbiology Teaching Fellow	2016
Travel and Conference Fellowship Award, Okinawa Institute of Science And Technology Graduate University	2015
University of Kentucky Graduate School Travel Award	2014-2015
Getrude Flora Ribble Travel Award	2014-2015
University of Kentucky Graduate School Travel Award	2014
University of Kentucky Teaching Assistantship	2013-2016
University of Kentucky Research Assistantship	2013-2015

Professional Publications

Bryant S., Herdy J., Amemiya C., Smith J. Characterization of Somatically-Eliminated Genes During Development of the Sea Lamprey (*Petromyzon marinus*). Molecular Biology and Evolution. In press. (2016)

Wigginton S.K., J. Edwards, N. Roff, H. Steinhaus, S. Milesko, J. Webb, S. Bryant, J. Binion, E. Smith, B. Tinsley, M. Penick, S. Grubbs, A. Meier. An assessment of the influence of *Podostemum ceratophyllum* on nutrient and sestonic algal levels in the Green River. Proceedings of the 14th Symposium on the Natural History of Lower Tennessee and Cumberland River Valleys. Center of Excellence for Field Biology, Austin Peay State Univ., Clarksville, Tenn. (2012)

Published Abstracts and Presentations

Molecular Correlates of Programmed DNA Elimination
Stephanie A. Bryant, Joseph R. Herdy and Jeramiah J. Smith
Plant and Animal Genome XXIII Conference, San Diego, CA (2015)

Characterization of Somatically-Eliminated Genes in the Sea Lamprey
Stephanie A. Bryant, Vladimir Timoshevskiy, Jeramiah J. Smith
Plant and Animal Genome XXIII Conference, San Diego, CA (2015)

Analysis of Pluripotency Genes in Spinal Cord Regeneration of the Sea Lamprey (*Petromyzon marinus*)
Stephanie A. Bryant, Jeramiah J. Smith
Plant and Animal Genome XXII Conference, San Diego, CA (2014)

A Study of the Aquatic Macrophyte *Podostemum ceratophyllum* in the Green River, Kentucky
Sara K. Wigginton, Joseph N. Edwards, Nikki H. Roff, Heidi Steinhaus, Steven Milesko, Joseph Webb, Stephanie Bryant, Jenna Binion, Eric Smith, Marie Tinsley, Mary Penick, Scott Grubbs, Albert Meier
Kentucky Academy of Science Meeting, Bowling Green, KY (2010)

Effects of Climate Change on Marine Chemistry and Biodiversity
Stephanie A. Bryant, William G. Hess, Margaret M. Crowder
Campus Community Partnership for Sustainability Conference, Bowling Green, KY (2009)