Maternal-to-Zygotic Transition in Xenopus Laevis is Regulated by miRNA Family 427

Yogindra Raghav, Sarah Morgan, Peter Allen
Supervisor: Dr. Miler T. Lee

Introduction

Yogindra Raghav, Sarah Morgan, and Peter Allen of the University of Pittsburgh are investigating quantifiable aspects of the maternal-to-zygotic transition of *Xenopus laevis* using computational methods. When an animal egg is first fertilized, new mRNA are produced after a period of no mRNA production. This is an example of transcriptional silence. Before fertilization, mRNA from oogenesis is still present and referred to as the maternal contribution. The maternal-to-zygotic transition is the time from the beginning of this transcriptional silence to the end of maternal contribution degradation. (Walser & Lipshitz, 2011)

The specific stages of life that we looked at in our research are the egg stage and the Stage 10 phase of these organisms. The egg stage to Stage 10 is a period of roughly 9 hours. These are depicted in Supplementary Figure 1-2. The process that governs this transition, ultimately leading to maternal mRNA degradation, is known as maternal clearance. The question at hand is as follows: is there a significant correlation between the target sequence "GCACTT" of microRNA family 427 being present in the 3'UTR of *Xenopus laevis* genes and the maternal clearance of those genes?

Xenopus laevis is a tetraploid animal which means that for every given gene and chromosome it has two sets that have different evolutionary origins. Each gene and chromosome are therefore labeled as being "L" or "S" depending on where the gene/chromosome originates from. Our group investigated a supplementary question: for the entire genome's L and S chromosomes, are there different levels of maternal clearance observed to a significant degree between all L chromosomes and all S chromosomes?

Our group hypothesizes that maternal-to-zygotic transition in this organism is regulated by the sequence "GCACTT" which is the target seed sequence of miRNA family 427 in *Xenopus laevis*. This miRNA family promotes deadenylation of the maternal contribution (Lund et al, 2009). This target sequence is in the 3'UTR regions of this organism's exome. We obtained the FASTA file of the entire genome, the 3'UTR sequences of the exome, and the GTF (Gene Transfer Format) file for this organism from XenBase.org. We also retrieved RNA-seq data for

the specific stages from the NCBI Gene Expression Omnibus. The usage of these files and reasoning is further explained in the Methods section.

To answer our question, we used genome aligners (hisat2), read counters (featureCounts), Unix text file manipulation commands, and R to do differential gene expression analysis as well as sequence motif discovery.

We were able to isolate a subset of genes that were significantly differentially expressed, containing genes that both had and did not have the miRNA family 427 target sequence. Our definition of significantly differentially expressed is that the RNA-seq read counts at Stage 10 were less than $\log_2(4)$ to the RNA-seq read counts found for that gene at the egg stage. This is summarized in the following equation: $\log_2(\text{Stage 10}) < \log_2(\text{egg}) - 2$.

Finally, we concluded that maternal-to-zygotic transition in *Xenopus Laevis* is regulated by miRNA family 427 because of the significant connection between maternal clearance and the target seed sequence of miRNA family 427 being present. Our supplemental question yielded interesting results as well. We found that, overall, all the L chromosomes and all the S chromosomes experienced significant differences in maternal clearance by miRNA family 427 targeting.

Methods

Our group worked with *Xenopus laevis* RNA-seq data taken from the egg stage and Stage 10 of development. This data was downloaded using each set's specific SRR numbers from the Gene Expression Omnibus and then FASTQdump was used to retrieve a FASTQ file of all this data. SRR numbers and further details can be found in the Supplementary Documentation. The reads retrieved from FASTQdump were aligned using hisat2, a fast genome aligner, to the entire genome of *Xenopus laevis* which we had access to as a FASTA file. Quantification was done by using samtools and featureCounts (this requires a GTF file which was available at XenBase.org). Then, differential gene expression analysis was determined using R. Reads that are similarly expressed in both the egg and Stage 10 are possibly not maternally cleared; reads that are highly expressed in the egg stage as compared to Stage 10 are possibly maternally cleared.

Then, the file of all 3'UTR regions for all genes was searched for the target sequence "GCACTT" using tools in Unix. Cross-referencing the subset of differentially expressed genes with all genes containing the 3'UTR target sequence gave us a set of genes that we hypothesized

to have undergone maternal clearance. To statistically validate our results, we ran a Fisher's significance test which can be found in Supplementary Figure 1-5.

To determine whether all L chromosomes and all S chromosomes experience significant difference in maternal clearance, we used another Fisher's significance test. Refer to Supplementary Figure 1-6 to look at the specific table and its setup.

Results

Supplementary Figure 1-3 is a differential expression biplot we made using the smoothed and logged RPM versions of our data. We used a log-difference threshold of 2 to determine which genes were significantly more expressed in the egg dataset than in the stage 10 dataset, which is the subset of maternally-cleared genes. This subset is shown in red in the graph.

1,583 of the 29,508 genes (5.36%) analyzed displayed a significant decrease between the egg stage and Stage 10. 550 genes of the 1,583 decreased genes (34.74%) contained the target sequence. The connection between decreased expression and containing the target sequence was shown to be statistically significant using a Fisher test. So, we conclude that the target sequence for miRNA family 427 regulates the processes of maternal clearance.

Additionally, 316 of the 550 genes (57.45%) were on the L chromosomes, and 234 were on the S (42.55%). By performing another Fisher test, we could conclude that there was a significant difference in regulation between the L and S chromosomes, with more regulation happening in genes on L chromosomes.

Discussion

We found that there was a statistically significant correlation between a gene containing the miRNA family 427 target sequence "GCACTT" in its 3'UTR region and a gene showing a decrease in expression from the egg stage to Stage 10 of development, which points to maternal clearance controlling the expression of that gene. In addition, we found that there was a significantly larger amount of regulated L genes than there were S genes, which leads us to believe that for some reason maternal clearance happens more on genes on the L chromosomes than on the genes on S chromosomes.

Our main limitation from this project is that we only used one experiment's dataset of reads from the egg Stage and Stage 10 of *Xenopus laevis* development. We have to be careful

before declaring that more genes on L chromosomes of *Xenopus laevis* are regulated than those on S chromosomes based off of only one experiment's data. In the future we could run the same methods we showed above on additional datasets to reinforce our initial findings. From there we could investigate why there might be this difference, such as looking for additional known target sequences that are present in the 3' UTR of genes on the L chromosomes and absent in those on the S chromosomes.

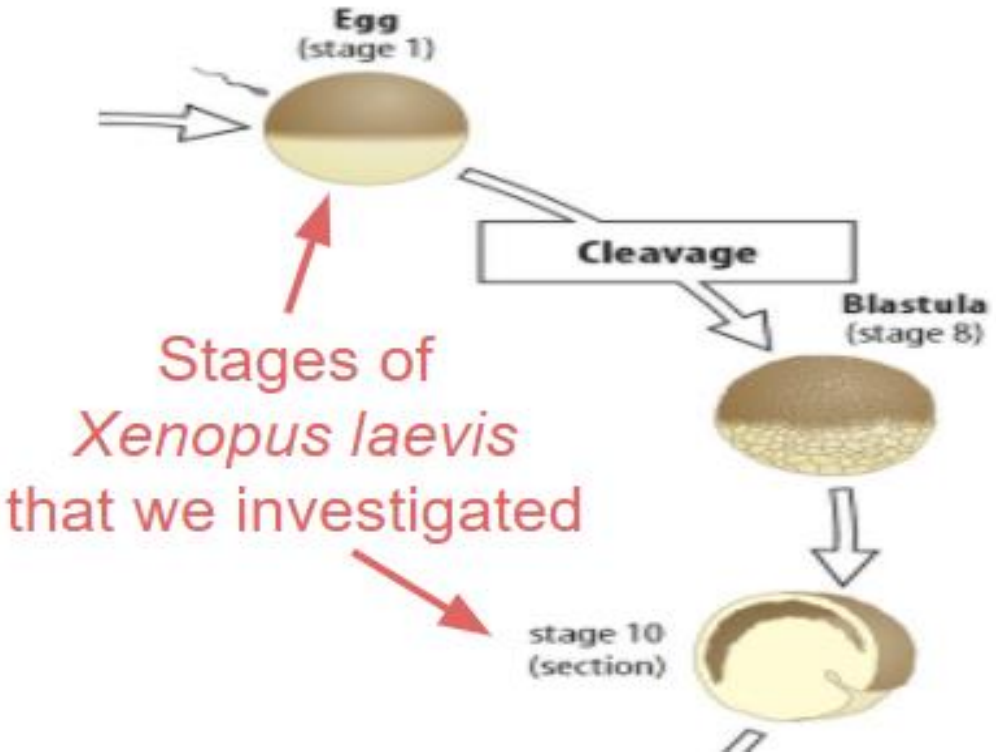
Supplementary Figures

Supplementary Figure 1-1. Entry counts.

8754	3’UTR file entries containing the target sequence GCACTT
175849	3’UTR identifiers matching only chromosomal reads from xbGene.9.1
13056	GTF entries that are determined to have significantly decreased expression by R and also have the 3’UTR sequence
7587	GTF entries that are significantly decreased, have the 3’UTR, and are on the L chromosomes
5469	GTF entries that are significantly decreased, have the 3’UTR, and are on the S chromosomes

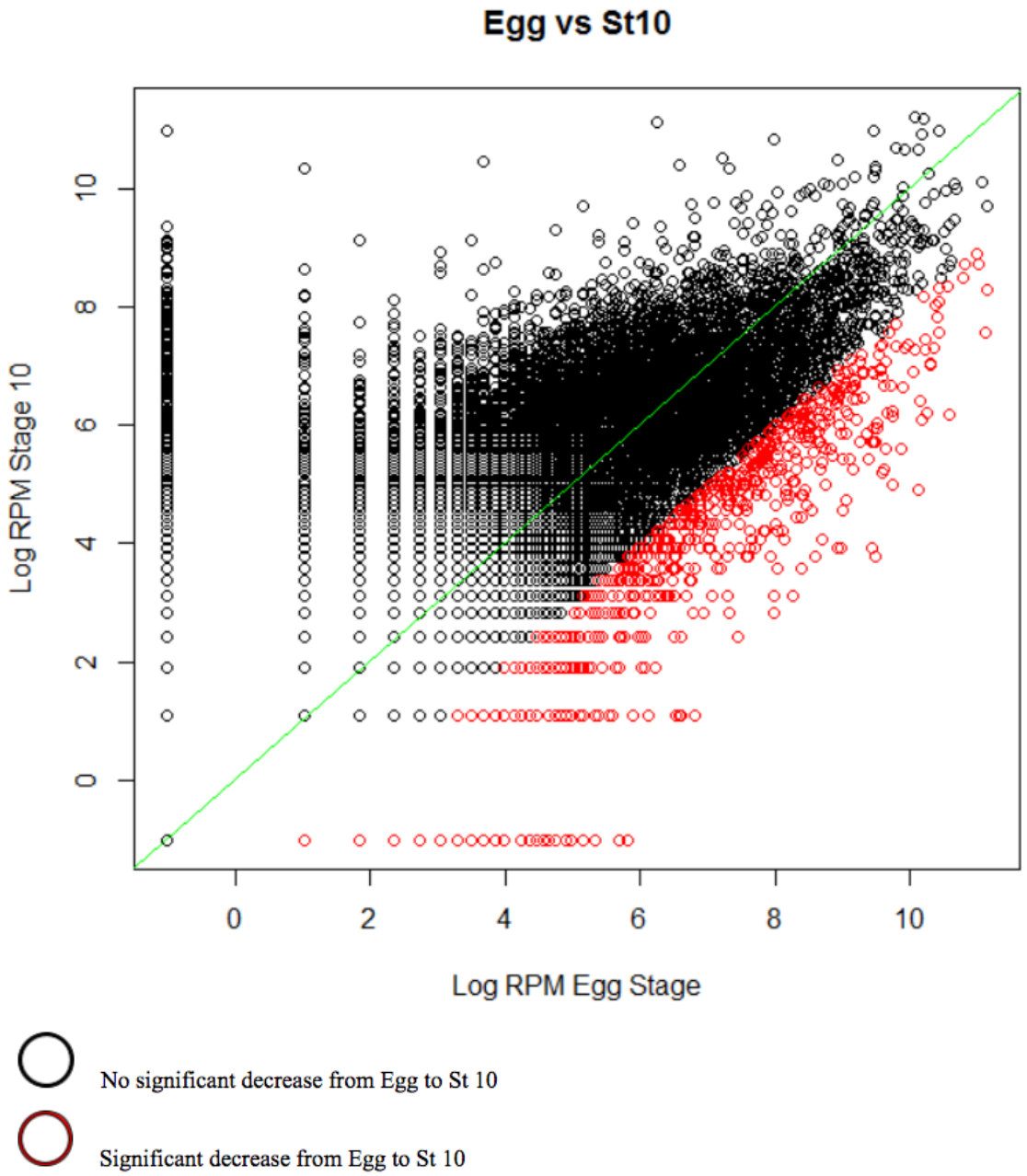
Basic information about the numbers found from our files using procedures explained in the Supplementary Documentation. These numbers are high due to duplicate entries in the GTF file. Extracting the geneID and removing duplicate entries gives the final count. The GTF file was used to match 3’UTR labels to geneID labels, as the files used had only one identifier or the other.

Supplementary Figure 1-2. Depiction of developmental stages of fertilized *X. laevis* egg.



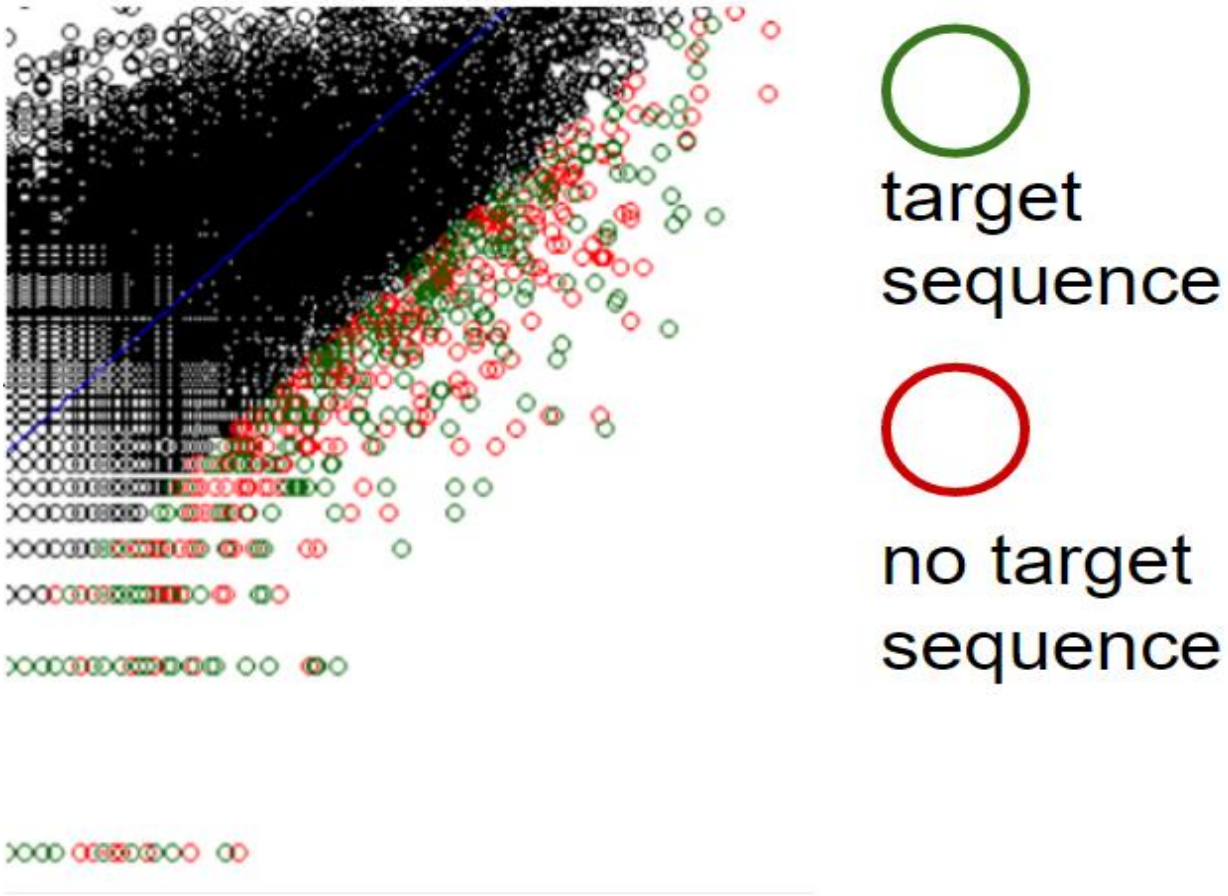
This image displays the difference between the egg stage and stage 10. Stage 10 is after many cellular cleavages and the formation of the blastula.

Supplementary Figure 1-3. Differential-gene-expression analysis visualization graph.



A visual representation of entries that have a log fold difference of 2 and showed a significant decrease in expression in stage 10 as compared to the egg stage. These decreased expression entries are in red; all others are in black.

Supplementary Figure 1-4.



The region of the graph from Supplementary Figure 1-3 that was marked in red has been zoomed in here. This is to better visualize which genes that were significantly differentially expressed in a lower fashion from egg to Stage 10 have the miRNA family 427 target seed sequence and which ones did not not contain it.

Supplementary Figure 1-5. Fisher test to quantify the connection between decreasing expression and having the target sequence.

Results			
	has seq	not seq	*Marginal Row Totals*
is decrease	550	1033	1583
no decrease	8204	19721	27925
Marginal Column Totals	8754	20754	29508 (Grand Total)

The Fisher exact test statistic value is 8E-06. The result is significant at p < .05.

P = 8E-06, so it is significant at p < 0.05. This means that there is a significant connection between having the target sequence and decreased expression.

Supplementary Figure 1-6. Fisher test to quantify the connection between gene location on L or S and being possibly maternally cleared.

Results			
	L	S	*Marginal Row Totals*
dec/seq	316	234	550
no dec/seq	14438	14520	28958
Marginal Column Totals	14754	14754	29508 (Grand Total)

The Fisher exact test statistic value is 0.000479. The result is significant at p < .05.

The location of a gene on L or S chromosomes was shown to be significant, with a p value of 0.000479. So, there is a difference in regulation between L and S chromosomes, with L having a significantly higher amount of regulated genes.

Works Cited

Walser, C. Lipshitz, H. (2011.) “Transcript clearance during the maternal-to-zygotic transition.” *Current Opinion in Genetics and Development,* 21(4). http://www.sciencedirect.com/science/article/pii/S0959437X11000621

Lund, E. et al. (2009.) “Deadenylation of maternal mRNAs mediated by miR-427 in *Xenopus laevis* embryos.” *RNA Journal*, 15. http://rnajournal.cshlp.org/content/15/12/2351.full

Easy Fisher Test Calculator. http://www.socscistatistics.com/tests/fisher/Default2.aspx