**Gene Expression Between *Xenopus laevis* and *Salmo salar* During Life History Transition Processes**

By: Loralei Miller, Aileen Li, Sara Ansari, Alfred Zhu, Nirvana Maleki, and Catherine Lewis

**Introduction**

Various organisms undergo life-history transition processes (e.g., metamorphosis, smoltification) that produce significant physiological, behavioral, morphological, and neural alterations. Neural plasticity is the change in neural connections during development due to environmental interactions and ontogenesis, which is the origination and development of an organism. One fascinating change that happens with these life-history transitions is the modification to the brain needed to accommodate new motor patterns, behaviors, and environments. As a result, studying gene expression across these transitions is a way to target and investigate neural plasticity.

Understanding life-history transitions in species that undergo drastic changes in a short amount of time, such as *Xenopus laevis* and *Salmo salar*, could help us identify genes associated with plasticity that apply to various organisms, including humans. Identifying novel candidate genes could provide insight into existing disorders. An example of this is Alzheimer’s disease, which is a progressive mental deterioration that can occur in middle or old age, due to generalized degeneration of the brain (Cuestas, 2020). The hypothesis of an imbalance in the cellular and molecular mechanisms of synaptic plasticity underlying this deficit is currently widely accepted (Cuestas, 2020). We will identify genes that are dynamic in expression across developmental stages for each through metamorphosis and smoltification. We will then compare the expression between the two species to find genes that are more robustly correlated with neuralplasticity. We expect *X. laevis* and *S. salar* to share genes associated with plasticity that show similar patterns of expression when comparing the earliest to latest stage. Our first aim is to perform differential analysis using DESeq2 to find how gene expression varies across developmental stages for each species. Our second aim is to find the functional categories of expressed genes using gene ontology. The third aim is to find genes that are more strongly correlated with neural plasticity by visualizing the overlaps in expression between the species.

**Materials and Methods**

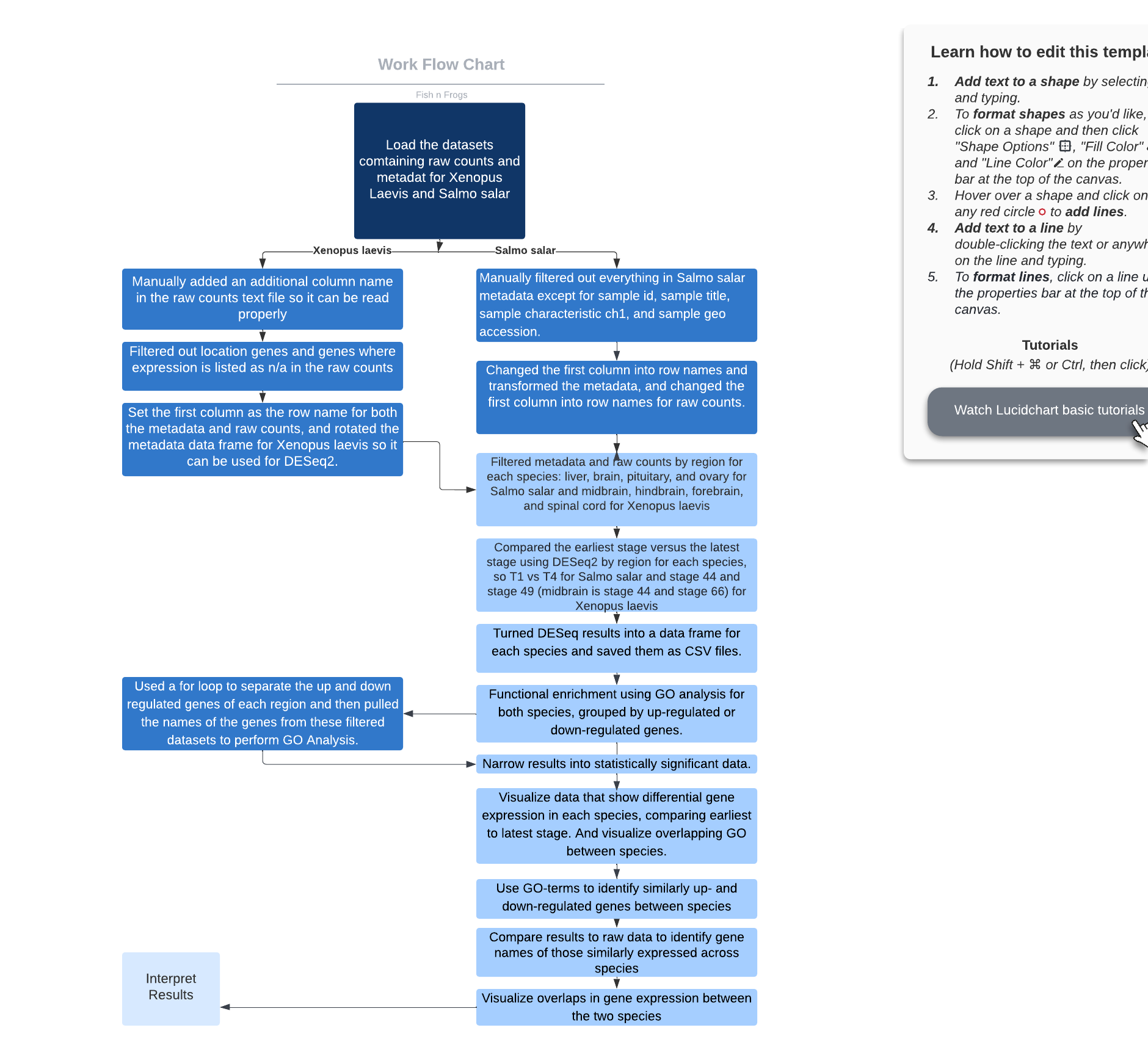
We obtained the *X. laevis* data from the research paper, “Temporal and spatial transcriptomic dynamics across brain development in *X. laevis* tadpoles” by Ta et al. (2021). The raw counts data contained observations for 49,109 genes across 31 samples. They obtained the samples from various brain regions of tadpoles (midbrain, hindbrain, forebrain, and spinal cord) in multiple stages of brain development (stages

44, 46, 49, 55, 61, and 66). Although there were midbrain samples for each of these development stages, all other brain region samples were only from stages 46 and 49. We filtered unannotated genes leaving 32,913 genes for downstream analysis.

The *S. salar* data was from “Multi tissue transcriptomic profiling during the onset of salmon maturation” by Mohammad et al (2018). There were 48 samples of salmon, with tissue sampling in regions including the brain, liver, ovary, and pituitary gland. Time points were identified as T1 through T4, with T1 representing gene expression before the onset of smoltification, and T2 through T4 representing gene expression in intervals of two weeks after the onset of smoltification.

Differential gene expression analysis was performed using the DESeq2 package in R studio to determine which genes are upregulated and downregulated between the earliest and latest stage of development in each separate region for both *S.salar* and *X. laevis.* The R package Enhanced Volcano was used to generate volcano plots for differentially expressed genes. Functional enrichment analysis was performed using the gprofiler2 package in R studio to identify which functional categories of genes were overrepresented across development in *S. salar* and *X. laevis.* For each individual species, gene ontology (GO) analysis was visualized using the commands gostplot() and publish\_gosttable().

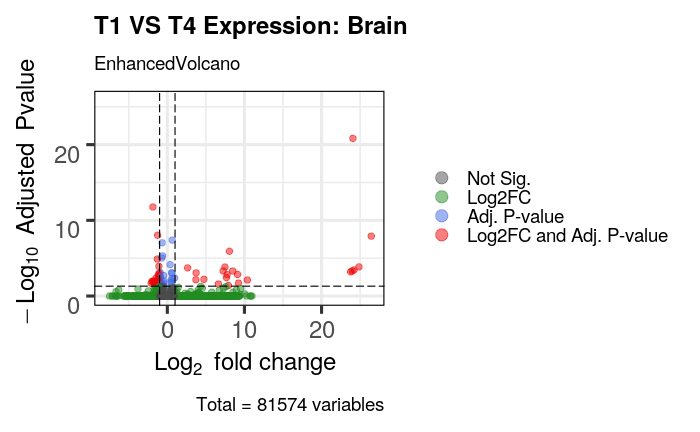
GO analysis data was used to identify up- and down-regulated pathways in the midbrain of *X. laevis*, and in the brain of *S. salar* during the first and last stages of each dataset. A data frame containing the GO-term IDs of up- and down- regulated genes from both species was created for comparison. Results were filtered to identify GO-terms that were commonly up- or down-regulated across species. A detailed workflow of data analysis is provided in Figure 1.

****

**Figure 1. Workflow Chart.** Each box is a step we took to analyze the data from respective species.**Results**

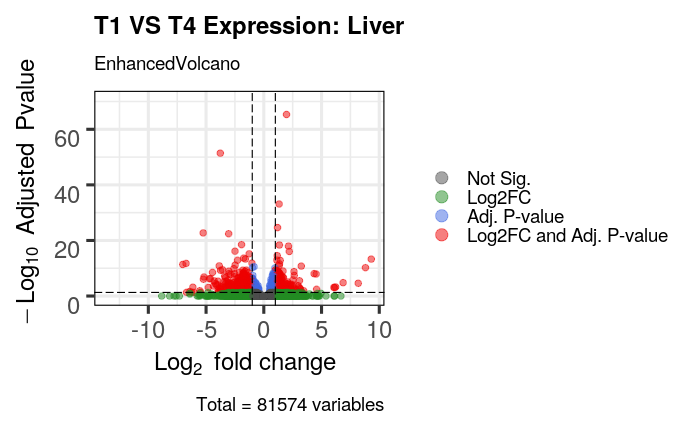
**DESeq**

In order to identify differential gene expression of the *S. salar* in each region during development, we compared the earliest stage, T1, to the latest stage, T4. For *X. laevis*, stages 46 and 49 were compared for all regions. The midbrain region was also analyzed using stages 44 and 66, the first and last stages for the midbrain data.



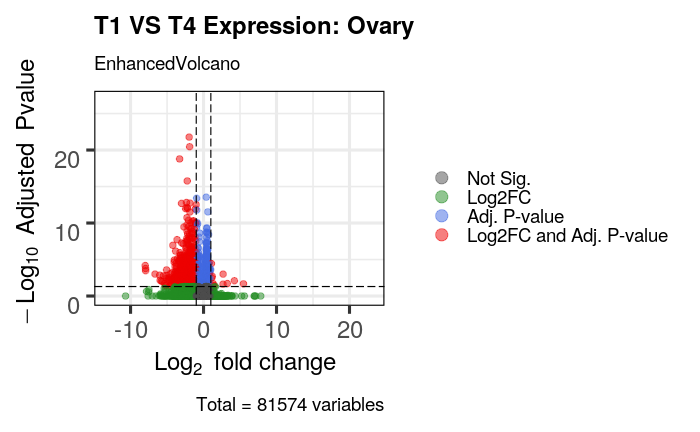
**Figure 2. Volcano plot comparing T1 and T4 for the brain region.** T1 means time point 1, which is the stage before *S. salar* maturation. T4 means time point 4, which is six weeks after smoltification began. Gray means a gene has an insignificant adjusted p-value. Green means a gene has an insignificant adjusted p-value but an absolute log2 fold change greater than one. Blue means a gene is expressed insignificantly with an absolute LogFC less than one. Red is statistically significant with adjusted p-value < 0.05 and with an absolute LogFc greater than or equal to one.

There are 42 genes with adjusted p-values less than 0.05, half of which are more significantly expressed in T1 than T4 and vice versa (Fig. 2). Majority of the 21 upregulated genes have a log2 fold change less than 10 and adjusted p-value less than eight. However, there are five genes outside of these ranges, one of which has a highly significant adjusted p-value. For the 21 down-regulated genes, all of them have a log2 fold change greater than -2.5. There are two genes with a noticeably higher adjusted p-value. One has a -log10 adjusted p-value greater than 10 and another has a -log10 adjusted p-value greater than five.



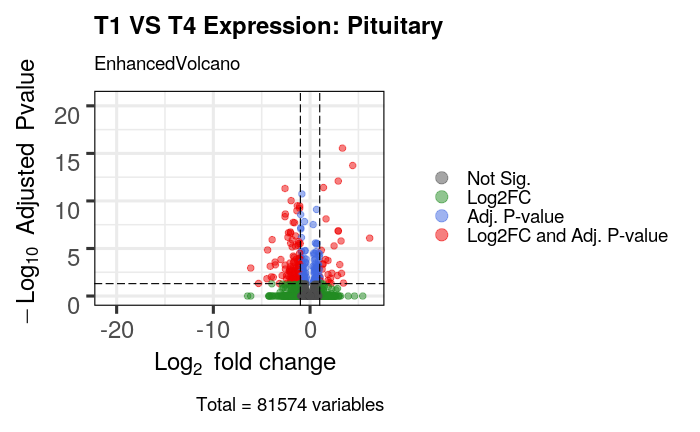
**Figure 3. Volcano plot comparing T1 and T4 for the liver region.** T1 means time point 1, which is the stage before *S. salar* maturation. T4 means time point 4, which is six weeks after smoltification began. Gray means a gene has an insignificant adjusted p-value. Green means a gene has an insignificant adjusted p-value but an absolute log2 fold change greater than one. Blue means a gene is expressed insignificantly with an absolute LogFC less than one. Red is statistically significant with adjusted p-value < 0.05 and with an absolute LogFc greater than or equal to one.

We found 770 genes differentially expressed in the liver (at adjusted p-value < 0.05; Fig. 3). There are 423 significant genes with a log2 fold change less than -1 and 347 significant genes with a log2 fold change greater than one. Both up-regulated and down-regulated genes are more clustered as the -log10 adjusted p-value decreases and as the absolute value of log2 fold change approaches 1. There are two genes with greater -log10 adjusted p-values, one up-regulated and one down-regulated, that have a -log10 adjusted p-value greater than 60 and 50, respectively.



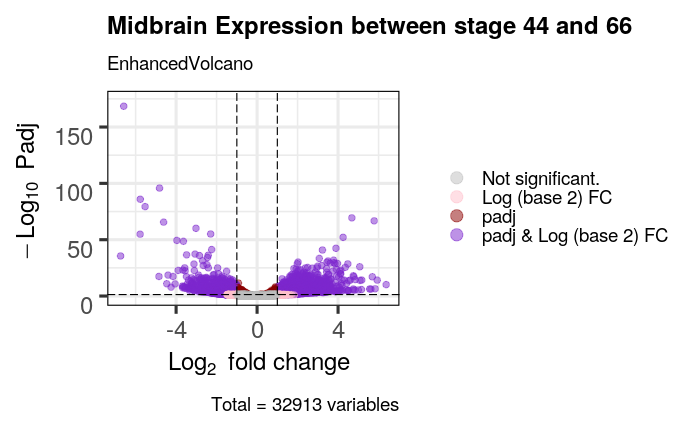
**Figure 4. Volcano plot comparing T1 and T4 for the ovary region.** T1 means time point 1, which is the stage before *S. salar* maturation. T4 means time point 4, which is six weeks after smoltification began. Gray means a gene has an insignificant adjusted p-value. Green means a gene has an insignificant adjusted p-value but an absolute log2 fold change greater than one. Blue means a gene is expressed insignificantly with an absolute LogFC less than one. Red is statistically significant with adjusted p-value < 0.05 and with an absolute LogFc greater than or equal to one.

There are 954 statistically significant down-regulated genes and 11 up-regulated genes in the ovary region (Fig. 4). There are more down-regulated genes with significant adjusted p-values compared to the 11 up-regulated genes. Five of the down-regulated genes are more statistically significant than the rest due to their -log10 adjusted p-value being greater than 15, two of which being even greater than 20. 939 of the down-regulated genes are found to have a log2 fold change greater than negative five. All the up-regulated genes have a -log10 adjusted p-value less than five.



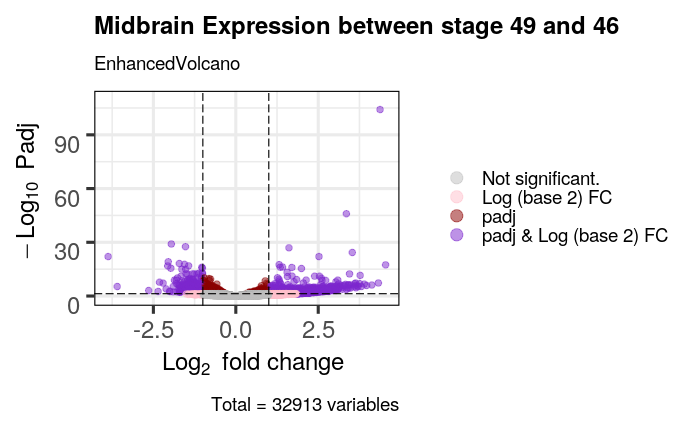
**Figure 5. Volcano plot comparing T1 and T4 for the pituitary region.** T1 means time point 1, which is the stage before *S. salar* maturation. T4 means time point 4, which is six weeks after smoltification began. Gray means a gene has an insignificant adjusted p-value. Green means a gene has an insignificant adjusted p-value but an absolute log2 fold change greater than one. Blue means a gene is expressed insignificantly with an absolute LogFC less than one. Red is statistically significant with adjusted p-value < 0.05 and with an absolute LogFc greater than or equal to one.

There are 166 differentially expressed genes in the pituitary region (Fig. 5). 136 genes are down-regulated. 28 genes are up-regulated. 134 the down-regulated genes have a log2 fold change greater than negative five. The up-regulated genes are more vertically dispersed than the down-regulated genes with the greatest -log10 adjusted p-value around 15, while the most statistically significant down-regulated gene’s -log10 adjusted p-value is less than 12.5. All up-regulated genes of statistical significance except for one have a log2 fold change less than five.



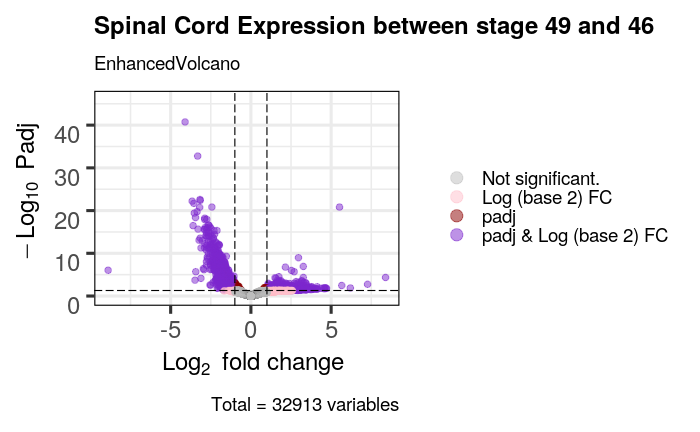
**Figure 6. Volcano plot comparing stages 44 and 66 in *X. laevis* Midbrain region.** Stage 44 is the earliest stage of metamorphosis available for *X. laevis*. Stage 66 is the latest stage of metamorphosis available for *X. laevis*. Gray means insignificant adjusted p-value. Red means significant with an absolute LogFC less than one. Purple means significant adjusted p-value with an absolute LogFc greater than or equal to one.

There are 590 genes that are more significantly expressed in stage 44 than in stage 66 in the midbrain region. Most of the genes have an absolute log2 fold change value less than four. There are 13 downregulated genes that have an absolute log2 fold change value greater than four. There are 44 upregulated genes that have an absolute log2 fold change value greater than four.



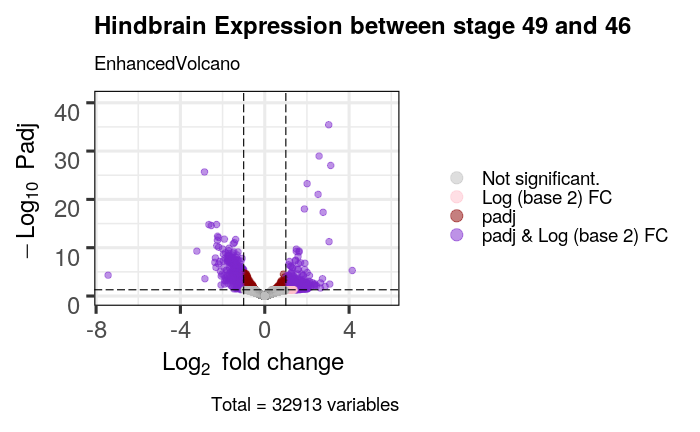
**Figure 7. Volcano plot comparing stages 46 and 49 in *X. laevis* Midbrain region.** Stage 46 and stage 49 are different stages of metamorphosis for *X. laevis*. Gray means insignificant adjusted p-value. Red means significant with an absolute LogFC less than one. Purple means significant adjusted p-value with an absolute LogFc greater than or equal to one.

There are 1,802 more significantly expressed genes in stage 49 than in stage 46 in the midbrain region. The upregulated genes also include more genes that have a log2 fold change value greater than 2.5 while most downregulated genes have an absolute log2 fold change value less than 2.5. There are 177 upregulated genes that have an absolute log2 fold change value greater than 2.5 while downregulated genes have three.



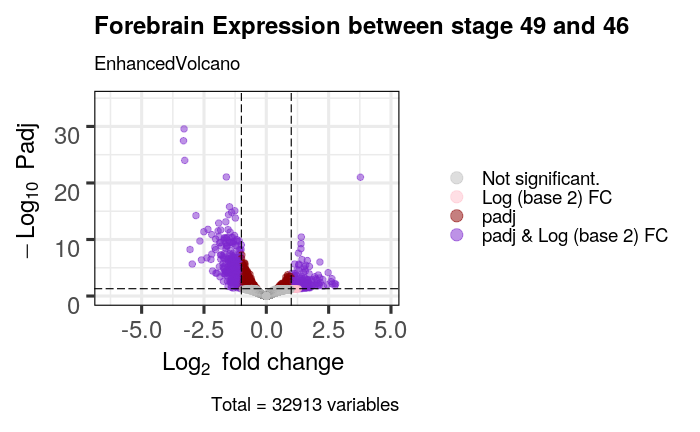
**Figure 8. Volcano plot comparing stages 46 and 49 in *X. laevis* Midbrain region.** Stage 46 and stage 49 are different stages of metamorphosis for *X. laevis*. Gray means insignificant adjusted p-value. Red means significant with an absolute LogFC less than one. Purple means significant adjusted p-value with an absolute LogFc greater than or equal to one.

There are 1,027 more significantly expressed genes in stage 46 than in stage 49. Both the upregulated and the downregulated genes mostly have an absolute log2 fold change value less than five. There are only five upregulated genes that have an absolute log2 fold change value greater than five while there is only one downregulated gene with an absolute log2 fold change value greater than five.



**Figure 9. Volcano plot comparing stages 46 and 49 in *X. laevis* Hindbrain region.** Stage 46 and stage 49 are different stages of metamorphosis for *X. laevis*. Gray means insignificant adjusted p-value. Red means significant with an absolute LogFC less than one. Purple means significant adjusted p-value with an absolute LogFc greater than or equal to one.

There are 186 more significantly expressed genes in stage 46 than stage 49. Most upregulated and downregulated genes have an absolute log2 fold change value less than four with only one gene for both having an absolute log2 fold change value greater than four. It is noteworthy that the downregulated gene has an absolute log2 fold change value of 7.4.



**Figure 10. Volcano plot comparing stages 46 and 49 in *X. laevis* Forebrain region.** Stage 46 and stage 49 are different stages of metamorphosis for *X. laevis*. Gray means insignificant adjusted p-value. Red means significant with an absolute LogFC less than 1. Purple means significant adjusted p-value with an absolute LogFc greater than or equal to one.

There are 225 more significantly expressed genes in stage 46 than in stage 49. Both the upregulated and the downregulated genes mostly have an absolute log2 fold change value less than 2.5. There are seven upregulated genes that have an absolute log2 fold change value greater than 2.5 while there are nine downregulated genes with an absolute log2 fold change value greater than 2.5.

**Gene Ontology (GO) Analysis**

|  |  |  |  |
| --- | --- | --- | --- |
| **Salmon GO Terms- Brain Timepoints T1 and T4** | | | |
| **Term\_ID** | **Source** | **Term\_Name** | **Regulation** |
| GO:0045098 | GO:CC | Type III intermediate filament | UP |
| KEGG:00100 | KEGG | Steroid Biosynthesis | UP |
| GO:0008250 | GO:CC | oligosaccharyltransferase complex | DOWN |

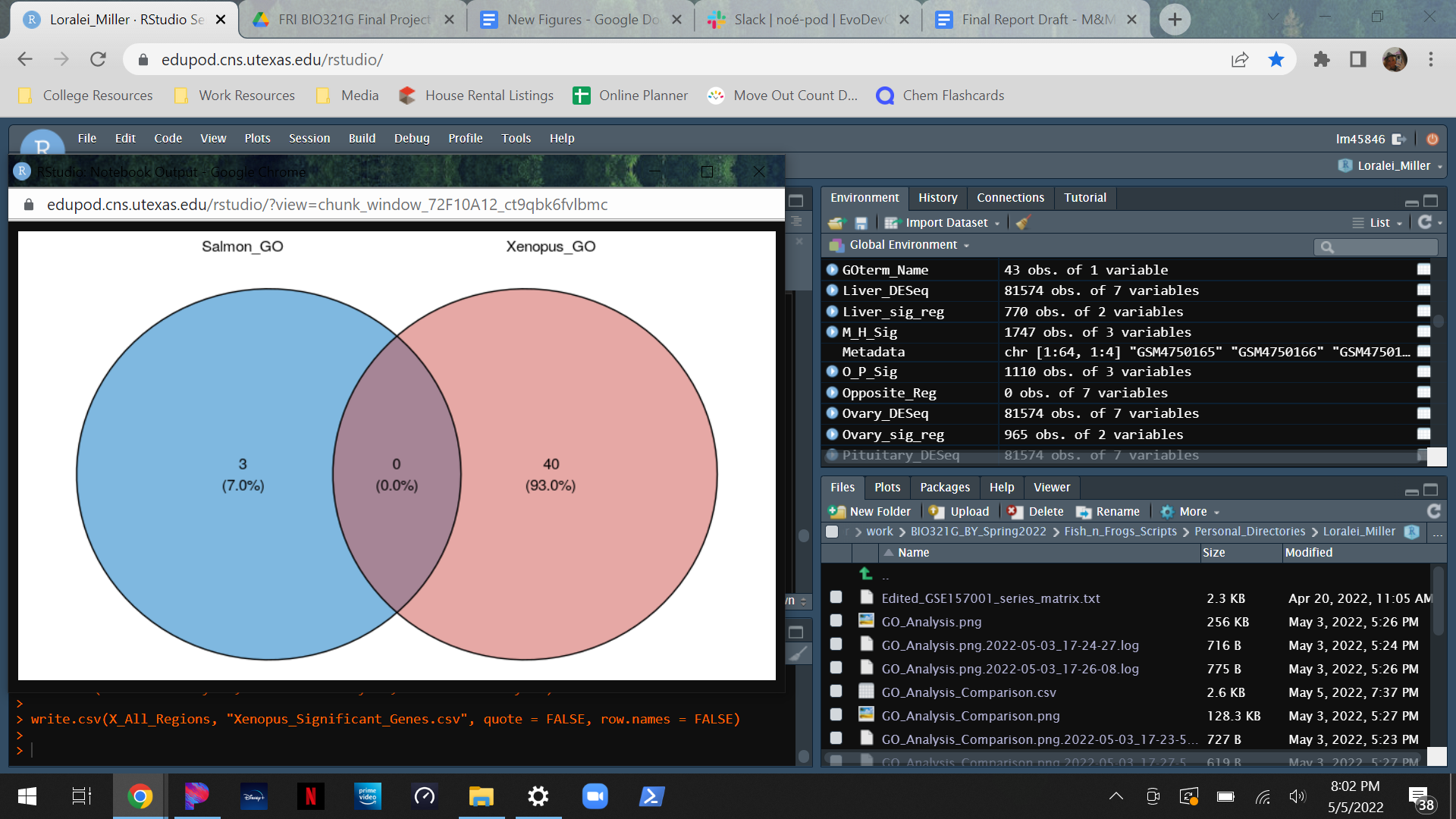
**Figure 11. Salmon GO terms that are significantly up- or down- regulated in the brain.** Source refers to the database from which the data was collected. Term name provides a brief description of the function of each GO term. Regulation identifies whether the GO term was up- or down-regulated in this species.

GO analysis was performed using the *S. salar* gene ontology data in order to find the function of significantly expressed genes within this species. There are 3 GO terms associated with the statistically significant genes in this species, with two upregulated (GO:0045098, a cellular component term, and KEGG:00100) and one downregulated (GO:008250, also a cellular component term).

|  |  |  |  |
| --- | --- | --- | --- |
| **Xenopus GO Terms- Midbrain Stage 44 and 66** | | | |
| **Term\_ID** | **Source** | **Term\_Name** | **Regulation** |
| GO:0003774 | GO:MF | cytoskeletal motor activity | UP |
| GO:0031531 | GO:MF | Thyrotropin-  releasing hormone receptor binding | UP |
| GO:0140296 | GO:MF | general transcription initiation factor binding | UP |
| GO:0051428 | GO:MF | peptide hormone receptor binding | UP |
| GO:0051427 | GO:MF | hormone receptor binding | UP |
| GO:0005524 | GO:MF | ATP binding | UP |
| GO:0036402 | GO:MF | Proteasome-  activating activity | UP |
| GO:0030554 | GO:MF | adenyl nucleotide binding | UP |
| GO:0032559 | GO:MF | adenyl ribonucleotide binding | UP |
| GO:0000166 | GO:MF | nucleotide binding | UP |
| GO:1901265 | GO:MF | nucleoside phosphate binding | UP |
| GO:0097367 | GO:MF | carbohydrate derivative binding | UP |
| GO:0043167 | GO:MF | ion binding | UP |
| GO:0035639 | GO:MF | purine ribonucleoside triphosphate binding | UP |
| GO:0032555 | GO:MF | purine ribonucleotide binding | UP |
| GO:0032553 | GO:MF | ribonucleotide binding | UP |
| GO:0017076 | GO:MF | purine nucleotide binding | UP |
| GO:0008092 | GO:MF | cytoskeletal protein binding | UP |
| GO:0004683 | GO:MF | Calmodulin-  dependent protein kinase activity | UP |
| GO:0043168 | GO:MF | anion binding | UP |
| GO:0036094 | GO:MF | small molecule binding | UP |
| GO:0008134 | GO:MF | transcription factor binding | UP |
| GO:0004879 | GO:MF | nuclear receptor activity | UP |
| GO:0098531 | GO:MF | ligand-activated transcription factor activity | UP |
| GO:0097159 | GO:MF | organic cyclic compound binding | UP |
| GO:1901363 | GO:MF | heterocyclic compound binding | UP |
| GO:0005516 | GO:MF | calmodulin binding | UP |
| GO:0003777 | GO:MF | microtubule motor activity | UP |
| GO:0001664 | GO:MF | G protein-coupled receptor binding | UP |
| KEGG:03050 | KEGG | Proteasome | UP |
| GO:0004129 | GO:MF | cytochrome-c oxidase activity | DOWN |
| GO:0005042 | GO:MF | netrin receptor activity | DOWN |
| GO:0016675 | GO:MF | oxidoreductase activity, acting on a heme group of donors | DOWN |
| GO:0005507 | GO:MF | copper ion binding | DOWN |
| GO:0015453 | GO:MF | Oxidoreduction-  driven active transmembrane transporter activity | DOWN |
| GO:0009055 | GO:MF | electron transfer activity | DOWN |
| KEGG:00190 | KEGG | Oxidative phosphorylation | DOWN |
| KEGG:04260 | KEGG | Cardiac muscle contraction | DOWN |
| REAC:R-XTR-6809371 | REAC | Formation of the cornified envelope | DOWN |
| REAC:R-XTR-6805567 | REAC | Keratinization | DOWN |

**Figure 12. Xenopus GO terms that are significantly up- or down- regulated in the midbrain.** Source refers to the database from which the data was collected. Term name provides a brief description of the function of each GO term. Regulation identifies whether the GO term was up- or down- regulated in this species

GO analysis was performed using *X. tropicalis* gene ontology data in order to find the function of significantly expressed genes within this species. There are 40 terms associated with the statistically significant genes in *X. laevis*, with 30 upregulated (Including mostly GO terms related to Molecular Functions, and one KEGG term) and 10 downregulated ((Including mostly GO terms related to Molecular Functions, two KEGG terms, and two REAC terms).



**Figure 13. Venn diagram comparing GO Analysis term overlap between Salmon brain region and Xenopus midbrain region.** *S. salar* data, including number of GO terms and percentage of data, is on the left in blue, *X. laevis* data, including the same measurement mentioned above, is on the right in red. Shared GO terms would be in the center in purple.

Using the GO terms identified from each species, a venn diagram was created to compare the function of significantly expressed genes between the two species. Between the three GO terms found in *S. salar,* and the 40 GO terms found in *X. laevis*, there are zero significantly expressed GO terms shared between the two species.

|  |  |  |
| --- | --- | --- |
| **Xenopus Significant Gene Expression Across Regions- Stages 46 and 49** | | |
| **Gene Name** | **Present in All Brain Regions** | **Presence in Spinal Cord** |
| hbd.L | Yes | No |
| hba2.L | Yes | No |
| enpp2.L | Yes | No |
| snrpg.S | Yes | No |
| pmp2.S | Yes | Yes |
| ptgds.L | Yes | No |
| prdx1.S | Yes | No |
| got2.L | Yes | Yes |
| hadhb.L | Yes | No |
| snrpf.L | Yes | No |
| eif5a.L | Yes | No |
| sncb.S | Yes | Yes |
| atp6v1h.S | Yes | Yes |
| snrpe.S | Yes | No |
| igf2bp3.S | Yes | Yes |
| plscr1.S | Yes | Yes |
| tsn.L | Yes | No |

**Figure 14. Xenopus genes similarly expressed across brain regions and spinal cord.** Gene names include gene symbol IDs of significantly expressed genes. In columns describing presence of genes in various regions, “Yes” signifies presence in the region and “No” signifies absence in the region.

There are 17 significantly expressed genes that are shared across all brain regions of the *X. laevis* data, including Midbrain, Forebrain, and Hindbrain. 6 of these genes also show significant expression in the Spinal Cord region.

|  |  |  |  |
| --- | --- | --- | --- |
| **Salmon Significant Gene Expression Across Regions- Timepoints T1 and T4** | | | |
| **Gene Name** | **Present in All Brain Regions** | **Present in Ovary** | **Present in Liver** |
| gene48968:106587736 | Yes | Yes | No |
| gene16920:106610036 | Yes | No | No |
| gene10660:106604089 | Yes | No | Yes |

**Figure 15. Salmon genes that are similarly expressed across brain regions, ovary, and liver.** Gene names include ENTREZ IDs of significantly expressed genes. In columns describing presence of genes in various regions, “Yes” signifies presence in the region and “No” signifies absence in the region.

There are 3 significantly expressed genes that are shared across all brain regions of the *S. salar* data, including brain and pituitary gland. One of these genes (gene48968:106587736) was found to have significant expression in the ovary in addition to the brain regions. Another gene (gene10660:106604089) was also found to have expression in the liver in addition to the brain regions.

**Discussion**

DESeq:

The volcano plots of the *S. salar* have an overall greater down regulation of genes than up-regulation, meaning that there is more gene expression in the latest stage (T4) than in the earliest stage (T1). As smoltification occurs, the *S. salar* not only transitions from freshwater to saltwater, but also transitions from a fry, which is the juvenile stage, to an adult fish (Wolf, 2005). This could explain why this trend is most prominent in the ovary and pituitary regions, since the pituitary gland is an important endocrine gland in the brain that is responsible for controlling growth, development, and the functioning of other endocrine glands, such as the ovaries (Khairnar, 2018). Therefore, there is greater gene expression near the end of smoltification than before smoltification in the pituitary and ovary regions because these regions play a major role in the development of *S. salar* becoming an adult. The ovary region of course would only apply to female *S. salar*. While it would’ve been interesting to compare the magnitude of gene expression for regions specific to male *S. salar*, there was no data found for the *S. salar* from the research paper pertaining to males only.

While the liver region is more down-regulated, it is not majorly down-regulated. Since the liver performs many of the same metabolic functions throughout the life history transition, it was interesting to see there was a large number of up- and down-regulated genes. We suppose that this could be due to the liver performing specific metabolic functions prior to smoltification and during smoltification, since it is a major transformation that would require a large amount of energy and by extension, nutrients the liver metabolizes. This could be looked into further with GO analysis.

The exception to the downward regulation trend is the volcano plot of the brain region that has an equal number of genes. We did not expect there to be so few significant genes in the brain region, however, the genes that were significant were distributed interestingly.

For *X. laevis* there is an overall down regulation of genes when comparing the earliest and latest stage in all regions as the earlier stages have more gene expression. This trend of down regulation is most apparent in the midbrain and spinal cord regions where the number of differential genes is above 500 with the spinal cord region above 1000. This might be because of our limited range of stages for the metamorphoses of *X. laevis*. There may be more overall gene expression if a comparison is made between stage 66 and an earlier stage of *X. laevis* before stage 44. Another potential reason for the overall downregulation of genes could be because certain genes essential for metamorphosis are more important in the earlier stages than the later stages. These potential genes may include genes to start mitosis in specific regions for limb development as well as pathways to start lung formation. The hindbrain and forebrain have the least amount of differential expression with 186 and 225 genes differentially expressed respectively. This may indicate that these regions may not be as involved in the metamorphosis process when compared with the midbrain and spinal cord regions.

The only region where genes are more expressed in the later stage is in the comparison between stage 46 and stage 49 of the midbrain region. It has more differential genes than the other comparisons between regions as well as the comparison between stage 44 and stage 66 in the midbrain. This might be the case due to the role of the midbrain in the regulation of breathing (Gargaglioni, 2007). *X. laevis* may need to regulate its breathing due to it needing energy to develop and requiring more oxygen to perform respiration.

GO Analysis:

In *S. salar*, there are two up-regulated GO terms, meaning they are more significantly expressed in time point T1 than T4. The first is the type III intermediate filament. There are four proteins associated with this filament including desmin, glial fibrillary acidic protein, peripherin, and vimentin. Desmin is expressed in all muscle types, GFAP is expressed in non-neuronal nervous system cells, and peripherin is expressed in peripheral neurons. (Elly and Yassemi, 2017) These protein types are all important in nervous system and muscular function, which plays a large part in the development of new motor functions in *S. salar.* The second up-regulated GO term regarding steroid horomones plays a large role in regulating salinity, metabolism and stress response, and in initiating and maintaining sexual differentiation and reproduction. (Schiffer, et al., 2019) The ability to regulate salt concentration in the blood is extremely important as the salmon make the transition from freshwater to saltwater. The decreased expression of these functions entering smoltification is contradictory to what would be expected and may suggest some error in our results.

The downregulated GO term in *S. salar*, the oligosaccharyltransferase complex, is responsible for the addition of glycans to Asparagine amino acids. This process plays an important role in protein quality control and trafficking, signal transduction, and cell communication that maintain cell homeostasis (Harada et al, 2019) . The downregulation of this function means that it is more significantly expressed during time point T4 than time point T1. Cell homeostasis is very important for the organism to survive its transition from freshwater to saltwater. Higher expression of these functions could help to ensure that the fish can adapt to its new environment despite the drastic changes to water composition.

In *X. laevis* the down regulated genes relate to cellular respiration and the production of ATP. These genes are more expressed in stage 66 rather than in stage 44. The process of metamorphosis involves rapid cellular division in order for *X. laevis* to grow limbs and in size. This process requires large amounts of energy, and as a result, gene expression for cellular respiration to produce ATP increases as *X. laevis* progresses through metamorphosis. *X. laevis* becomes larger as it progresses through metamorphosis which results in more sites of cell division which will require more energy.

The upregulated genes in *X. laevis* vary in function. Some of these genes include those that involve signal transduction pathways which are vital for cell communication and gene expression. When *X. laevis* undergoes metamorphosis, its brain region must communicate with other body cells in order to coordinate and control gene expression so the organism can develop properly. There are also genes that are involved in DNA transcription processes. These processes are vital to gene expression as DNA transcription produces mRNA which will then be translated to proteins. Signal transduction pathways control DNA transcription of certain genes so it makes sense that both are upregulated. These upregulated genes are essential for communication, coordination, and gene expression between cells which is necessary for metamorphosis where cells need to coordinate in order for the organism to develop properly. This also lines up with our DESeq results where there is more expression in the earlier stage than the later stage as there are also more expression of genes that relate to signal transduction and DNA transcription which will result in more gene expression.

We have found zero overlap of genes between *X. laevis* and *S. salar* after filtering for statistical significance and significant gene expression. This may be the case due to the lack of significant GO terms in the brain region of *S. salar* returned from GO analysis, which majorly reduced the number of possible overlap between the species. Additionally, we only examined the brain and midbrain from *S. salar* and *X. laevis*, respectively, due to the midbrain being the only region with the earliest (stage 44) and latest stage (stage 66) available for the species. Had the other brain regions from *X. laevis* had both those stages then we would be able to analyze them as well, but it still would not solve the number of *S. salar* GO terms issue.

Another issue we faced before GO analysis was gene IDs. The gene IDs for *S. salar* could not be converted into the same type as *X. laevis*, so we had to resort to using the GO terms instead of gene IDs in order to compare between the species. This led to another issue with GO terms. A gene can have multiple GO terms, but we used one GO term that was chosen through GO analysis for each gene ID. In doing so, we could only compare to find the same GO term between species, so even if the genes could be present in both species but had differing GO terms selected through GO analysis, then those genes would not be considered as an overlap.

We also initially filtered for overlap between the species without consideration for significant expression by using 0 as the basis for determining up- or down-regulation instead of 1 and -1, respectively. We did not find any genes that relate to neural plasticity in our results. The genes that are shared between *S. salar* and *X. laevis* relate to metabolic pathways and fine motor functions.

This indicates that *S. salar* and *X. laevis* do not share any genes that relate to plasticity during their major life-history transition. Both of these species do not have any overlap of significantly expressed genes between them. Even when we did and did not account for significance in gene expression, both of these species do not share genes that relate to neural plasticity. This leads us to conclude with our null hypothesis.

We recognize there are likely better methods to analyze the data to answer our question and hope to revisit this kind of comparison between species for neural plasticity with more species and a better understanding of analytical methodologies.

**Data Availability**

All scripts, figures, data are saved in the EDUpod directory: “/stor/work/BIO321G\_BY\_Spring2022/Fish\_n\_Frogs\_Scripts”.

Original data for *X. laevis* done by Ta et al. can be found in the GEO repository under GSE183193, as well in the EDUpod directory: “/stor/work/BIO321G\_BY\_Spring2022/Fish\_n\_Frogs\_Scripts/Data/Xenopus/XenopusBrain\_metamorphosis”.

The metadata, raw counts and normalized counts are provided in an easy to view format and are labeled with “fixed” in the previously mentioned directory.

The *X. laevis* scripts we created are in the EDUpod directory: “/stor/work/BIO321G\_BY\_Spring2022/Fish\_n\_Frogs\_Script/Personal\_Directories” under “Alfred\_Zhu”, “Sara\_Ansari’, and “Catherine\_Lewis”.

DESeq script for this species can be found in: “/stor/work/BIO321G\_BY\_Spring2022/Fish\_n\_Frogs\_Script/Personal\_Directories/Alfred\_Zhu”.

GO analysis script for this species can be found in: “/stor/work/BIO321G\_BY\_Spring2022/Fish\_n\_Frogs\_Script/Personal\_Directories/Sara\_Ansari”.

The *S. Salar* original data by Mohamed et al. is found in the GEO repository under GSE157001 and in the EDUpod: “/stor/work/BIO321G\_BY\_Spring2022/Fish\_n\_Frogs\_Scripts/Data/Salmo”.

The data includes raw counts and the series matrix.

*S. salar* script and modified metadata are found in the EDUpod: “/stor/work/BIO321G\_BY\_Spring2022/Fish\_n\_Frogs\_Scripts/Personal\_Directories” under “Aileen\_Li’, “Loralei\_Miller, and “Nirvana\_Maleki”.

DESeq and modified metadata can be found in: “/stor/work/BIO321G\_BY\_Spring2022/Fish\_n\_Frogs\_Scripts/Data/Salmo/Aileen\_Li”.

GO analysis script for this species and cross-comparison can be found in: “/stor/work/BIO321G\_BY\_Spring2022/Fish\_n\_Frogs\_Scripts/Data/Salmo/Loralei\_Miller”.

Figures for both species are found in: “/stor/work/BIO321G\_BY\_Spring2022/Fish\_n\_Frogs\_Scripts/Figures”.

They are saved as PDFs. *S. salar* will begin with “S” and *X. laevis* will begin with “x” and is titled according to region.

DESeq data frames are saved as CSVs in the directory: “/stor/work/BIO321G\_BY\_Spring2022/Fish\_n\_Frogs\_Scripts/Data/DEQseqResults”.

**Citations**

Cuestas Torres, Diana Marcela, and Fernando P. Cardenas. “Synaptic Plasticity in Alzheimer’s Disease and Healthy Aging.” *Reviews in the Neurosciences,* vol. 31, no. 3, 2020, pp. 245–68, <https://doi.org/10.1515/revneuro-2019-0058>.

Gargaglioni, Luciane H., et al. “Role of Midbrain in the Control of Breathing in Anuran Amphibians.” *American Journal of Physiology-Regulatory, Integrative and Comparative Physiology,* vol. 293, no. 1, 2007, pp. R447–57, <https://doi.org/10.1152/ajpregu.00793.2006>.

Harada, Yoichiro, et al. “Oligosaccharyltransferase: A Gatekeeper of Health and Tumor Progression.” *International Journal of Molecular Sciences*, vol. 20, no. 23, 2019, <https://doi.org/10.3390/ijms20236074>.

Hol, Elly M., and Yassemi Capetanaki. “Type III Intermediate Filaments Desmin, Glial Fibrillary Acidic Protein (GFAP), Vimentin, and Peripherin.” *Cold Spring Harbor Perspectives in Biology,* vol. 9, no. 12, 2017, <https://doi.org/10.1101/cshperspect.a021642>.

Khairnar, Sachin & Holeyappa, Shanthanagouda. (2018). Breeding and spawning of fishes: Role of endocrine gland. 6. 472-478.

Mohamed, A., et al. “Multitissue Transcriptomic Profiling during Onset of Salmon Maturation.” *NCBI,* 2020. [www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE157001](http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE157001).

Schiffer, Lina, et al. “Human Steroid Biosynthesis, Metabolism and Excretion Are Differentially Reflected by Serum and Urine Steroid Metabolomes: A Comprehensive Review.” *The Journal of Steroid Biochemistry and Molecular Biology*, vol. 194, 2019, <https://doi.org/10.1016/j.jsbmb.2019.105439>.

Ta, Aaron C., et al. “Temporal and Spatial Transcriptomic Dynamics across Brain Development in Xenopus Laevis Tadpoles.” G3 *Genes|Genomes|Genetics,* edited by B Andrews, vol. 12, no. 1, 2021. <https://doi.org/10.1093/g3journal/jkab387>.

Wolf, Jeffrey C., and Marilyn J. Wolfe. “A Brief Overview of Nonneoplastic Hepatic Toxicity in Fish.” Toxicologic Pathology, vol. 33, no. 1, 2005, pp. 75–85, <https://doi.org/10.1080/01926230590890187>.

Packages:

Readr

Hadley Wickham and Jim Hester (2020). readr: Read Rectangular Text Data. R package version 1.4.0. [https://CRAN.R-project.org/package=readr](https://cran.r-project.org/package=readr)

Apeglm

Zhu, A., Ibrahim, J.G., Love, M.I. Heavy-tailed prior distributions for sequence count data: removing the noise and preserving large differences Bioinformatics (2018)

Tidyverse

Wickham et al., (2019). Welcome to the tidyverse. Journal of Open Source Software, 4(43), 1686, <https://doi.org/10.21105/joss.01686>

DESeq2

Love, M.I., Huber, W., Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2 Genome Biology 15(12):550 (2014)

EnhancedVolcano

Kevin Blighe, Sharmila Rana and Myles Lewis (2019). EnhancedVolcano: Publication-ready volcano plots with enhanced colouring and labeling. R package version 1.4.0. <https://github.com/kevinblighe/EnhancedVolcano>

EnrichR

Wajid Jawaid (2021). enrichR: Provides an R Interface to 'Enrichr'. R package version 3.0. [https://CRAN.R-project.org/package=enrichR](https://cran.r-project.org/package=enrichR)

Gprofiler2

Kolberg L, Raudvere U, Kuzmin I, Vilo J, Peterson H (2020). “gprofiler2- an R package for gene list functional enrichment analysis and namespace conversion toolset g:Profiler.” \_F1000Research\_, \*9 (ELIXIR)\*(709). R package version 0.2.1.