University of Puerto Rico

Río Piedras Campus

Department of Biology

Molecular mechanisms of intestinal organogenesis: A transcriptomic approach

by:

Noah A. Auger ©

A thesis submitted to the Biology Graduate Program in fulfillment of the requirements for the degree of

Master of Science in Biology

Date of Defense:

May 5, 2023

Date of Thesis Submission:

June 2, 2023

This thesis has been accepted by the faculty of the

Department of Biology

Faculty of Natural Sciences

University of Puerto Rico

Río Piedras Campus

In fulfillment of the requirements for the degree of

Master of Science in Biology

José E. García-Arrarás, Ph. D.

Thesis advisor

Humberto Ortiz-Zuazaga, Ph. D.

Thesis Committee

Alfredo Ghezzi, Ph. D.

Thesis Committee

Imilce Rodríguez-Fernández, Ph. D.

Thesis Committee

Table of Contents

List of Tables

List of Figures

Abbreviations

DET: differentially expressed transcripts
DGE: differential gene expression
DPE: days post-evisceration
GO: gene ontology
GSEA: gene set enrichment analysis
HPE: hours post-evisceration
KCl: potassium chloride
ORA: over representation analysis
WGCNA: weighted gene co-expression network analysis
L2FC: log2FoldChange
EMT

Acknowledgements

This project was funded by NIH (1R15GM124595) and the University of Puerto Rico. I also acknowledge support from the Sequencing and Genomics Facility and the High-Performance Computing Facility of the University of Puerto Rico that was supported by an Institutional Development Award (IDeA) from the National Institute of General Medical Sciences of the NIH under grant number P20GM103475.

Abstract

Organogenesis is a process normally reserved for developing organisms, however, some adult organisms are capable of such a feat. Adult sea cucumbers, in the highly regenerative clade Echinodermata, can undergo intestinal organogenesis after a process of evisceration where the intestine is severed then expelled through the cloaca. Studying this phenomenon has implications in regenerative biology to help discover the cellular and molecular processes that allow some species to regenerate better than others. Also, the field of regenerative medicine can benefit by enhancing existing biomedical applications, particularly that of intestinal tissue engineering. Thus, the sea cucumber *Holothuria glaberrima* is being used in this study to investigate intestinal organogenesis. The cellular properties of its regeneration have largely been elucidated, but the molecular mechanisms are now just starting to be unveiled. In this study we use a bioinformatic approach to investigate molecular mechanisms by collecting RNA-seq data from various early- and late-stage intestinal regenerate timepoints. These new timepoints not only contain newly sequenced tissue with luminal epithelium, but also samples that are spatially separated, i.e., anterior and posterior regenerates. Analyses from differential gene expression, gene set enrichment analyses, and weighted gene co-expression revealed differences in gene expression and molecular pathways operating at distinct temporal and spatial timepoints. For example, various signaling components of the Wnt/β-catenin pathway appear to be upregulated during early-stage rudiment regeneration, while signaling components of the Wnt/Planar Cell Polarity pathway are upregulated during late-stage regeneration in tissue that contains luminal epithelium. Even distinct gene ontological terms are functionally enriched in anterior regenerates that are not in posterior regenerates such as chromosome organization and cilium assembly. Altogether, this study demonstrates the diversity of molecular mechanism that occur during a dynamic process like intestinal organogenesis and serves as a baseline to guide future hypothesis-driven molecular studies.

Keywords: intestinal organogenesis, echinoderm, molecular mechanisms, transcriptome, differential gene expression, weighted gene co-expression network analysis, gene ontology

Chapter 1: Introduction to Model Organism

Introduction

Regenerative biology seeks to discover the cellular and molecular mechanisms that underly regeneration. Many questions can be investigated such as why some species regenerate better than others, and why tissues in the same organism differ in regenerative capacity. Regenerative medicine then leverages this knowledge to create biomedical applications like tissue engineering. One area of particular interest is intestinal tissue engineering where biosynthetic constructs are inserted and integrated at a functional capacity into the intestine. This can be applied to diseases such as Short Bowel Syndrome where large portions of the intestine are removed due to cancer or Crohn's disease, thus causing malnutrition from lack of proper intestinal length. Many other diseases can also benefit, like Hirschsprung disease where parts of the intestine lack the enteric nervous system. However, many hurdles exist to achieve these biomedical applications: (1) selecting a proper cell source for regenerating different layers of the intestine, (2) regulation of cellular processes that occur during regeneration such as proliferation and differentiation, (3) maintaining the contractile phenotype of muscle cells, (4) and determining differences in anterior and posterior regenerates (Bitar & Raghavan, 2012). These hurdles can be overcome by studying the process of intestinal regeneration in various species in disparate taxa.

Several model organisms exist to study intestinal regeneration. Many of them such as rodents and fruit flies are limited in regenerative capacity to self-renewal of intestinal stem cells. However, species in the Echinodermata clade possess a greater regenerative capacity than most other model organisms and are more closely related to humans (Figure 1). For example, the echinoderm *Holothuria glaberrima*, the model organism used in this study, can undergo intestinal organogenesis as an adult after induced evisceration and is conveniently found in abundance off the coast of Puerto Rico. Many studies have examined the histological and cellular events during *H. glaberrima* intestinal regeneration. At the histological level, the intestine is composed of three main layers: an outer mesothelium, an inner connective tissue, and the luminal epithelium (Figure 2A) (Mashanov & García-Arrarás, 2011). The intestine is attached to the body wall by the mesentery, which is continuous with the outer two layers of the intestine but lacks luminal epithelium (García-Arrarás et al., 2019). Intestinal regeneration begins when evisceration is induced by injecting the sea cucumber with KCl. The intestine then detaches from its connections to the esophagus, cloaca, and mesentery and is then dispelled through the cloaca (Figure 2B). This tears the edges of the mesentery at 0 hours post-evisceration (hpe). By 7 days post-evisceration (dpe), cells from the mesothelium and inner connective tissue of the torn mesentery have migrated to provide the cell source of the intestinal rudiment. At 14-dpe, the luminal epithelium from the remaining anterior esophagus and posterior cloaca ingress into the rudiment. By 21-dpe, the luminal epithelium connects to form a continuous lumen. These histological observations have been further analyzed at the cellular level.

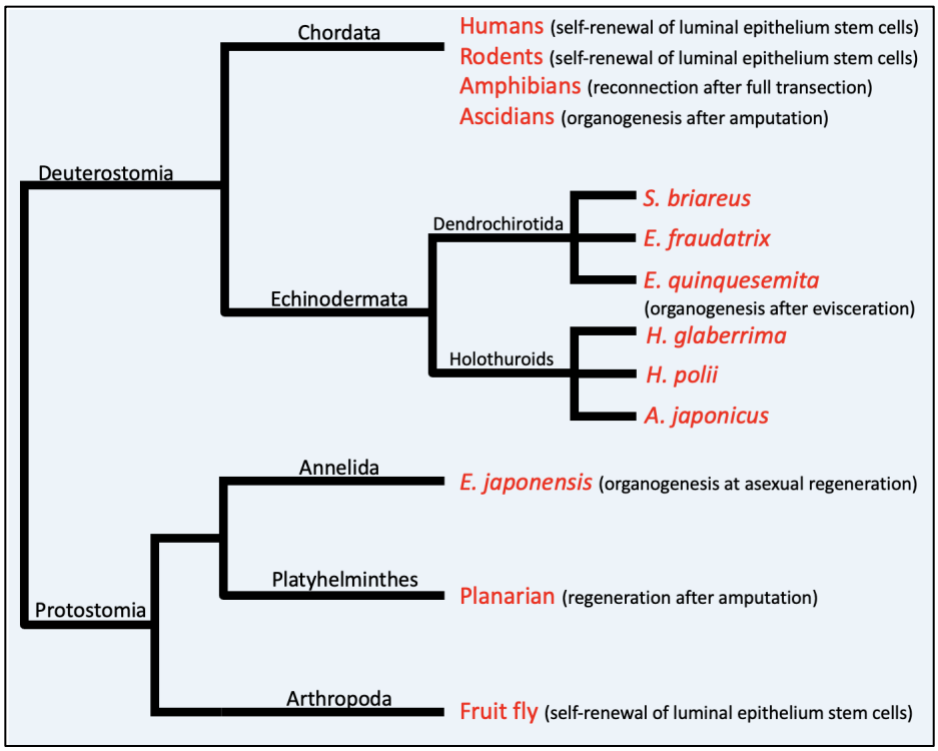


Figure 1. Evolutionary Relationship of Model Organisms Used for Intestinal Regeneration Studies. This cladogram depicts the evolutionary relationship of various species that have been studied for intestinal regeneration, in red, within various phyla. Next to each species in parenthesis is the extent to which the species can regenerate its intestine. For example, the greatest extent of intestinal fruit fly regeneration is the self-renewal of intestinal stem cells. The Echinodermata phylum is split into two classes, Dendrochirotida and Holothuroids, which are all capable of regeneration following evisceration. The model organism of focus in this study is the holothuroid *Holothuria glaberrima*.

The cellular processes of intestinal regeneration are well-studied in all three layers of the intestine, but emphasis has been placed on the muscle cells of the mesothelium, intrinsic and extrinsic nerves, and enterocytes of the luminal epithelium. Figures 3 and 4 provide an overview of the many cellular processes that occur over the course of regeneration. However, here only a major few will be discussed. Muscle cells in the mesothelium undergo dynamic changes in cellular states including de-differentiation and myogenesis (Murray et al., 2001; Murray & García-Arrarás, 2004; Candelaria et al., 2006) as well as epithelial-to-mesenchymal transitions (García-Arrarás et al., 2011). Various parts of the enteric nervous system have been studied such as a study dedicated to neuroendocrine cells (Huang & García-Arrarás, 2013) as well as a broader study that looked at extrinsic and intrinsic enteric system regeneration, which demonstrated that neurodegeneration precedes reinnervation of the regenerating intestine (Tossas et al., 2014). More recently, the nervous system component of the mesentery was investigated, showing that most of it remains in place during regeneration while a contraction occurs at the distal end (Nieves-Ríos et al., 2020). Enterocytes have also been viewed where they de-differentiate into a pluripotent state and migrate into the rudiment to then proliferate and re-differentiate to form the new luminal epithelium (García-Arrarás et al. 1998).

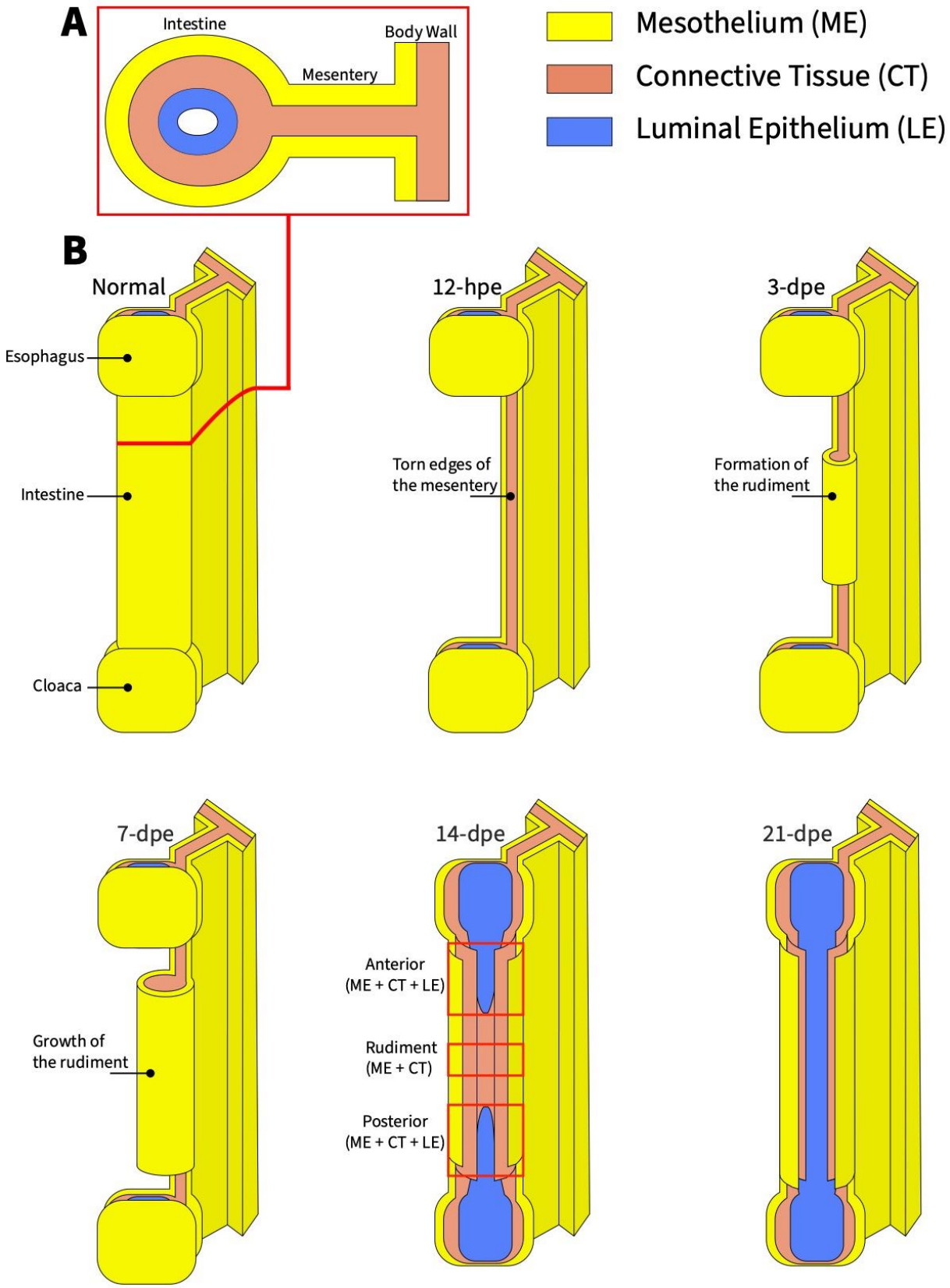

A
Intestine
Body Wall
Mesentery
Mesothelium (ME)
Connective Tissue (CT)
Luminal Epithelium (LE)
B
Normal
Esophagus
Intestine
Cloaca
12-hpe
Torn edges of
the mesentery
3-dpe
Formation of
the rudiment
7-dpe
Growth of
the rudiment
14-dpe
Anterior
(ME + CT + LE)
Rudiment
(ME + CT)
Posterior
(ME + CT + LE)
21-dpe

Figure 2. Intestinal Histology and Process of Regeneration in the Sea Cucumber. **(A)** Transverse cut of the intestine that is bound to the body wall by the mesentery. The intestine is composed of three layers the Mesothelium (ME), Connective Tissue (CT), and the Luminal Epithelium (LE). The mesentery is continuous to the intestine but only by the ME and CT layers. **(B)** Histological overview of intestinal regeneration after evisceration. The normal intestine is attached to the anterior esophagus and posterior cloaca, but once evisceration occurs those connections are severed. Thus, immediately after evisceration at 0-hpe the torn edges of the mesentery are exposed in the coelomic cavity. At 3-dpe the mesentery serves as the cell origin for the rudiment by forming thickenings along the torn edges to create a blastema-like structure composed of mesothelium and inner connective tissue. At 7-dpe, these mesenterial thickenings continue to grow, forming a larger blastema-like structure. At 14-dpe the intestine has been cut longitudinally to demonstrate what is occurring in all three layers of the intestine. Cells from the luminal epithelium of the esophagus and cloaca proliferate and migrate into the intestinal rudiment. Here the regenerating digestive tract can be spatially divided where anterior and posterior regenerates contain luminal epithelium while in between, only rudiment is found. The three sections demarcated in red indicate where samples were extracted from and what tissue types they contain for 14-dpe RNA-seq samples. By 21-dpe, the luminal epithelium extensions have connected to create a continuous lumen covered in luminal epithelium.

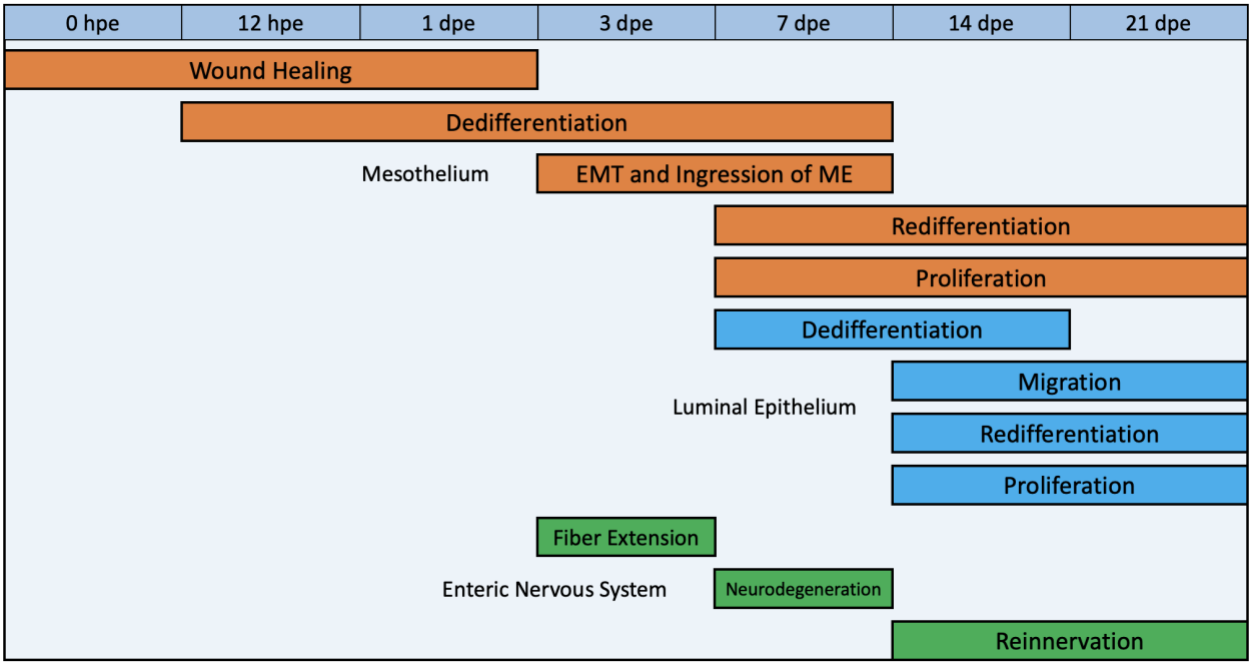


Figure 3. Cellular Processes during Intestinal Regeneration. This schematic shows the various cellular processes that occur during different timepoints of regeneration within different tissues. On the top row are timepoints of regeneration where cellular studies have been conducted. Underneath are color coded boxes that display cellular processes occurring within a cell or tissue type. Boxes in orange are associated to the mesothelium layer. Boxes in blue are related to the luminal epithelium tissue, while green boxes concern the neuronal cell types of the enteric nervous system. EMT: epithelial-to-mesenchymal transition. ME: mesothelium.

Underlying molecular mechanism of intestinal regeneration are also being explored. There are in situ hybridization studies of *Mortalin* and *Survivin* (Donato-Santana et al., 2008) as

well as of *Wnt9*, *TCTP* and *Bmp1* (Mashanov et al., 2012). Also, there are studies on the ubiquitin proteasome system (Pasten et al., 2010), and a RNAi study on the Wnt/β-catenin pathway (Alicea-Delgado & García-Arrarás, 2021). However, these studies are attempts to dissect the molecular mechanisms of sea cucumber intestinal regeneration by using what is already known in other species. What needs to be explored are the molecular mechanisms unique to *H. glaberrima* regeneration. In this way, novel regulators of regeneration might be found that are activated in the sea cucumber but perhaps lie dormant in other less regenerative species. This aim was achieved by combining previous RNA-seq data with new RNA-seq of intestinal regenerates into a single transcriptome, to then perform a series of bioinformatic analyses. The preexisting RNA-sew data consisted of samples from 12-hpe, 1-, 3-, and 7-dpe (Quispe-Parra et al., 2021). This data was completed in the present study by performing RNA-seq at 14-dpe of the rudiment. The RNA-seq data was further extended by adding timepoints from tissue of 14-dpe anterior and posterior regenerates as well as 21-dpe and normal intestine, which acted as the control. The goal was to add these missing timepoints then analyze the data to find potential molecular regulators of intestinal regeneration.

Aims

General Aim

Create an extensive transcriptomic profile of intestinal regeneration by adding timepoints to the rudiment regeneration timeline and by adding RNA-seq data from regenerating intestine that contains luminal epithelial tissue, as well as samples of anterior and posterior intestinal regenerates. Then, perform downstream bioinformatic analyses to provide a preview of the potential molecular mechanisms unique to sea cucumber intestinal regeneration. This will serve as a jumping off point for investigators to study specific genes or signaling pathways during intestinal regeneration in *H. glaberrima*.

Specific Aims

1. Compile all RNA-seq data into a single transcriptome then perform an unbiased assessment of gene expression data by performing various bioinformatic analyses such as differential gene expression, gene set enrichment, and weighted gene co-expression network analysis, and over-representation analysis. This will be the baseline for subsequent studies into genes and signaling pathways required for intestinal regeneration.

2. Perform a biased assessment of potentially important genes during intestinal regeneration by focusing on the Wnt gene family. This will be a case study for future investigators to see what type of analyses can be done with the gene expression data collected in the first aim. The Wnt gene family will have its coordinates manually annotated in the genome and its expression pattern will be mapped across all timepoints of intestinal regeneration.

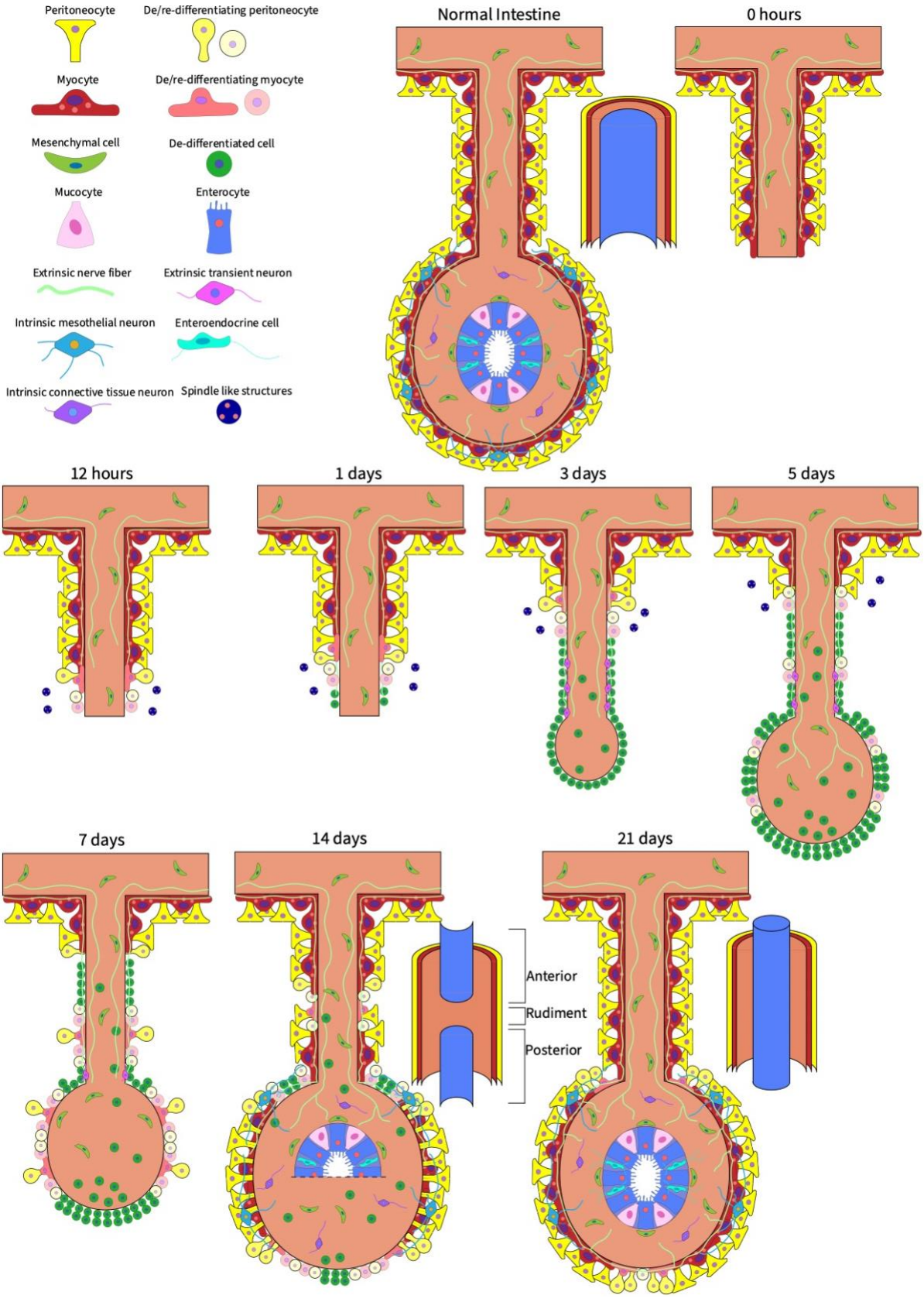

Peritoneocyte
De/re-differentiating peritoneocyte
Myocyte
De/re-differentiating myocyte
Mesenchymal cell
De-differentiated cell
Mucocyte
Enterocyte
Extrinsic nerve fiber
Extrinsic transient neuron
Intrinsic mesothelial neuron
Enteroendocrine cell
Intrinsic connective tissue neuron
Spindle like structures
Normal Intestine
0 hours
12 hours
1 days
3 days
5 days
7 days
14 days
21 days
Anterior
Rudiment
Posterior

Figure 4. Process of Intestinal Organogenesis after Evisceration**.** Listed in the top left corner are all the cell types that make up a normal (uneviscerated) intestine, except for the extrinsic transient neurons. In sequential order starting from the top row is a representation of the normal intestine at the cellular level (left) and at a tissue level (right). At the tissue level, yellow and red represent the mesothelium (peritoneocytes and myoepithelium, respectively), tannish brown is the connective tissue while blue is the luminal epithelium. At 0-hpe the mesenterial torn edge is represented at the distal tip of the mesentery. At 12-hpe, the spindle like structures appear and are represented as the blue spheres projecting from either side of the mesentery. At 1-dpe the cells of the mesothelium have de-differentiated into a pluripotent cell state (circular green cells). At 3-dpe, the rudiment is depicted as the bulging growth from the distal tip of the mesentery and continues to grow at 7-dpe. At 14 days, the anterior and posterior luminal epithelium from the esophagus and cloaca invade the rudiment.

Chapter 2: Gene Expression and Gene Networks of Intestinal Regeneration

Introduction

Organogenesis is a complex process defined by cell fate determination, cell proliferation, cell growth, cell migration, amongst other cellular processes (Jimenez-Rojo et al., 2012; Kupperman et al., 2000; Xiao et al., 2007). Underneath this cellular complexity lies further intricacies at the molecular level where genes are differentially expressed and signaling pathways are inhibited or activated (Lü et al, 2005; Xue et al., 2011). These major genes and signaling pathways must be understood to obtain a clearer picture of organogenesis. One approach to achieve this aim is to use bulk mRNA-seq, which aims to quantify the levels of mRNA in a tissue or organ at a certain area or timepoint. These data can then be probed to create a transcriptome that is used for various bioinformatic analyses. Several labs, including ours, studying intestinal regeneration in echinoderms have deployed transcriptomic approaches to study intestinal organogenesis in *Holothuria leucospilota* (Wu et al., 2020), *Apostichopus japonicus* (Yuan et al., 2019), and *Eupentacta fraudatrix* (Boyko et al., 2020).

One issue with these RNA-seq studies, including those from our lab, is that they consist of few timepoints and lack spatial separation, particularly that of tissue separation (rudiment versus advanced stages that contain luminal epithelium) as well as anterior versus posterior regenerates. The current RNA-seq data in our lab consists of two separate sequencing experiments that compose early-stage rudiment regeneration of 12-hpe, 1-, 3- and 7-dpe, but does not consider later stages of rudiment regeneration nor stages when the luminal epithelium appears. Therefore, this present study proposes to address these issues by first completing the timeseries of rudiment regeneration. Then a new timeseries was started at the beginning of lumen formation, which separates the anterior and posterior regenerates that contain luminal epithelium. All these experiments were then combined to form a singular transcriptome that was used for downstream bioinformatic analyses including Differential Gene Expression (DGE), Gene Set Enrichment Analysis (GSEA), Weighted Gene Co-expression Network Analysis (WGCNA), and Over-Representation Analysis (ORA). The result of these analyses provides an unbiased insight into the genes and molecular pathways that are involved during intestinal regeneration and serve future investigators with data to formulate hypothesis-driven experiments.

Differential Gene Expression (DGE)

Many bioinformatic methods were used in this study. The rest of the introduction is dedicated to explaining the concepts behind each method and what type of information can be extracted from them. The first major method is a DGE analysis. The first step in a DGE is to collect RNA-Seq samples and create a transcriptome, which is akin to a genome but only composes the exons of genes that have been transcribed in a cell. This allows the quantification of gene expression of a given biological state, thus allowing comparisons between a control and an experimental condition. In the case of this study, the controls are a normal un-eviscerated

mesentery and intestine. These will be compared to the rudiment and tissue that contains luminal epithelium, respectively. This method is effective for discovering changes in gene expression and can identify important genes involved during the regeneration process. Additional information on transcriptome building can be found in some excellent reviews (Dong et al., 2016; Garber et al., 2011) as well as some recent ones for DGE analyses (Costa-Silva et al., 2017; McDermaid et al., 2019).

Gene Set Enrichment Analysis (GSEA)

A lesser-known method used here is GSEA, which can determine whether genes in a gene-set, related to a gene ontological (GO) term, are enriched. This is done by ranking genes from a DGE analysis by their log2FoldChange value, highest positive value to lowest negative value (Figure 5A). These genes are not pre-filtered and do not consider p-adjusted values, as this could skew the statistical analysis. The next step is to determine if genes within a gene-set are enriched at the positive or negative end of the ranked list. The GO term ribosome biogenesis will be used as an example to describe the enrichment process. The genes highlighted in orange in Figure 5A are associated to the ribosome biogenesis gene-set (Figure 5B). Most of them are at the positive end of the ranked list, so ribosome biogenesis is an activated GO terms and will appear as such in the GSEA graph. The enriched GO terms from the analysis are ordered by GeneRatio. This value is calculated by dividing the total number of genes in the ribosome biogenesis gene-set by the number of genes associated to the ribosome biogenesis gene-set found within the ranked list. Ribosome biogenesis is shown to have a GeneRatio of about 0.55 (Figure 5C). The GSEA helps to understand at a broad spectrum the molecular events during intestinal regeneration.

A

Rank	Gene	L2FC	GO term
1	*AATF*	23	Ribosome biogenesis
2	*KIF14*	16	Cell division
3	*LSG1*	11	Ribosome biogenesis
4	*FER*	7	Diapedesis
5	*MTG2*	3	Ribosome biogenesis
...	...	...	...
700	*APC*	-5	Cell division
701	*VCP*	-11	DNA repair
702	*GNl1*	-14	Ribosome biogenesis
703	*GJA5*	-18	Vasomotion
704	*NBN*	-20	DNA repair

B

Gene Ontology	Gene Set
Ribosome biogenesis	*AATF, GNl1, LSG1, MTG2, . . .*
DNA repair	*BRCA1, NBN, RAD52, VCP, . . .*
. . .	

C

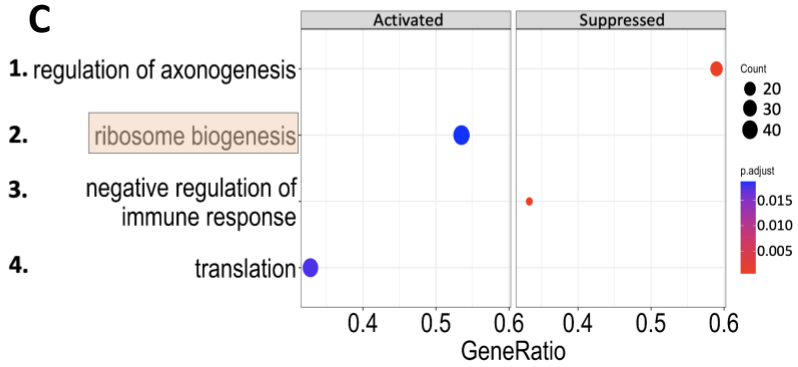


Figure 5. Explanation of a Gene Set Enrichment Analysis. **(A)** A ranked list of genes in descending order from highest positive to lowest negative log2FoldChange (L2FC) value. Each gene has a rank, a name, a L2FC value,

and an associated Gene Ontological (GO) term. Highlighted in orange are genes within the gene-set related to ribosome biogenesis. **(B)** Each GO term has a gene-set. Ribosome biogenesis is highlighted in orange to match the genes in A that are part of its gene-set. **(C)** The outcome of a GSEA is a graph that plots enriched GO terms that were found within the ranked list. Ribosome biogenesis was enriched at the positive end of the ranked list as seen in A, thus on the GSEA graph it appears as activated. Genes within a gene-set that are enriched at the negative end of the ranked list will have their associated GO term suppressed. These GO terms are ordered based on GeneRatio, which is calculated by dividing the number total number of genes within a gene-set by the number of genes in the ranked list associated to that gene-set. The size of the dot on the plot represents the number of genes found in the ranked list while the color is the p-adjusted value.

Weighted Gene Co-expression Network Analysis (WGNCA)

The other lesser-known method is a WGCNA, which is based on the notion that genes that interact with one another are likely co-expressed. Genes that are co-expressed form modules (groups of genes that display a similar expression pattern) (Figure 6A). Genes within each module have connections to one another, but a select few have many connections and are called hub genes, which may be considered at the top of a regulatory network. Discovering hub genes is beneficial to finding regulators of important molecular pathways. To perform a WGCNA, a few steps must be achieved. First, there must be raw gene expression counts (Figure 6A). Second, the gene counts must be transformed into a scale-free network where many genes have few connections, and few genes have many connections (Figure 6B). Lastly, the scale-free network can then be used to extract modules. These modules contain genes that are highly correlated (connected) to one another. The genes that have the greatest number of connections to other genes within a module are considered hub genes and can be visualized in a gene regulatory network (Figure 6C). Results of a WGCNA are candidate genes (hub genes) that may be regulators of important molecular pathways, which can be further assayed in knock down experiments.

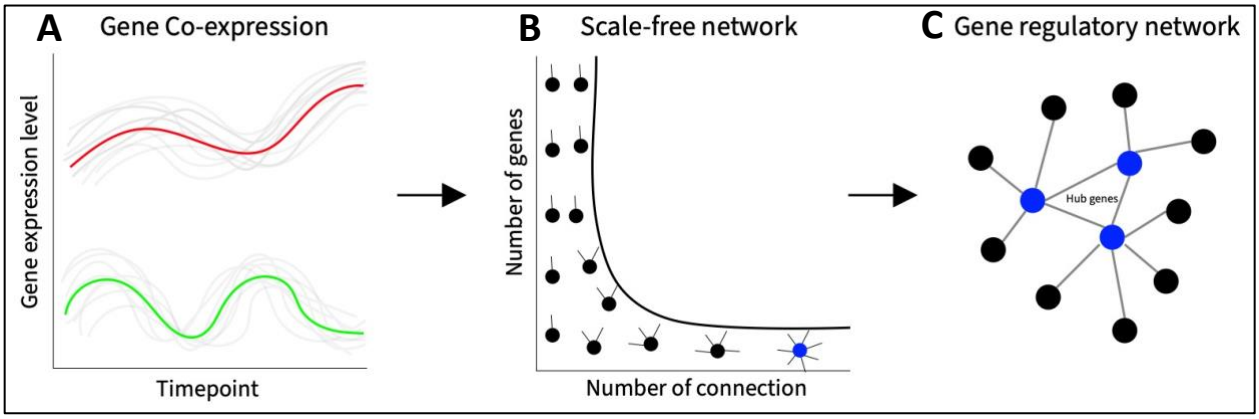


Figure 6. Explanation of a Weighted Gene Co-expression Network Analysis. **(A)** A plotted graph of gene expression over various timepoints. Gray lines represent the expression profile of an individual gene. These grays lines are clustered into two modules. The eigengenes of each module is represented as the red and green line. **(B)** A visual representation of a scale-free network. The dots represent genes (just as the gray lines in A), and the sticks projecting out of each dot represent the number of connections a gene has with other

genes within its module. A scale-free network is achieved when many genes within a module have few connections while few genes have many connections. The latter are considered hub genes and are colored in blue. **(C)** Gene regulatory network derived from the scale-free network transformation of the gene expression data. The blue dots are the hub genes within the network that have more connections to other genes than the black dots on the periphery.

Results

Validation of RNA-seq Samples

Two RNA-seq experiments using intestinal tissues have been previously performed in our lab that cover 12-hpe, 1-, 3-, and 7-dpe of the regenerating rudiment (Table 1). The normal mesentery was used as a control instead of a normal intestine since the former is the cell source of the regenerating rudiment. Novel data has been collected for this study that will serve to complete the timeseries of the rudiment by adding 14-dpe. Additionally, since RNA-seq data from different sequencing experiments will be combined into one transcriptome, the novel data will also include samples from 7-dpe to compare against the previously obtained 7-dpe sample. This will be done to identify possible batch effects. Moreover, the rest of the novel data will come from advanced stages of regeneration with samples that contain luminal epithelium at 14- and 21-dpe. The 14-dpe samples are divided into three spatial segments: anterior and posterior regenerates that contain luminal epithelium and a rudiment section that lacks luminal epithelium (these samples are demarcated in red in Figure 2B at 14-dpe). The reasoning behind sectioning tissue samples at 14-dpe is to see if the differences in tissue composition and spatial location reflect differences in gene expression.

Table 1. All RNA-seq Data Collected during Intestinal Organogenesis

	Normal Mesentery	12-hpe	1-dpe	3-dpe	7-dpe	14-dpe Rudiment	14-dpe Anterior	14-dpe Posterior	Normal Intestine
Quispe-Parra (2020)	X		X	X					
Quispe-Parra (2021)		X			**X**				
Auger (2022)					**X**	X	X	X	X

Table 1. Each row is a different RNA-seq experiment conducted in our lab. Each column is a timepoint where tissue was taken for sequencing. The Xs represent what timepoint was sequenced in an experiment. The Xs highlighted in yellow were used to determine if batch effects were present when all the sequencing data was combined into one transcriptome.

A single transcriptome was composed from all three RNA-seq experiments. The transcriptome quality was assessed by Transrate (Smith et al., 2016). It contained 620,734 contigs with a mean length per contig of 810 nucleotides (Table 2). The transcriptome was also

assessed for completeness by BUSCO (Simão et al., 2015). It showed 99% completeness for core genes while only detecting 0.4% fragmented genes and 0.6% missing core genes (Table 3).

Table 2. Transrate Transcriptome Assembly Quality

Metric	Result
Number of sequences	620,734
Total length (nt)	503,632,791
Longest sequence (nt)	49,901
Shortest sequence (nt)	176
Mean sequence length (nt)	810
N50 sequence length (nt)	1,768

Table 3. BUSCO Transcriptome Gene Content

Metric	Result
Core genes queried	954
Complete core genes detected	99 %
Complete single copy core genes	22.2 %
Complete duplicated core genes	76.8 %
Fragments core genes detected	0.4 %
Missing core genes	0.6 %

After assessing the transcriptome, gene expression was quantified using Salmon (Patro et al., 2017) and Corset (Davidson & Oshlack, 2014). The raw expression counts were then analyzed by a Principal Component Analysis (PCA) (Figure 7). This revealed that the 7-dpe samples from the two RNA-seq experiment clustered together. Other samples were also spatially distributed from one another, exemplifying gene expression uniqueness at each timepoint. However, when putting all timepoints together in a PCA plot, the early stages of rudiment regeneration appeared to be clustered together with no spatial distribution, thus a second PCA plot was constructed using just data from the rudiment. This demonstrated that the early timepoints indeed display spatial distribution by conforming to a chronological (Figure S1).

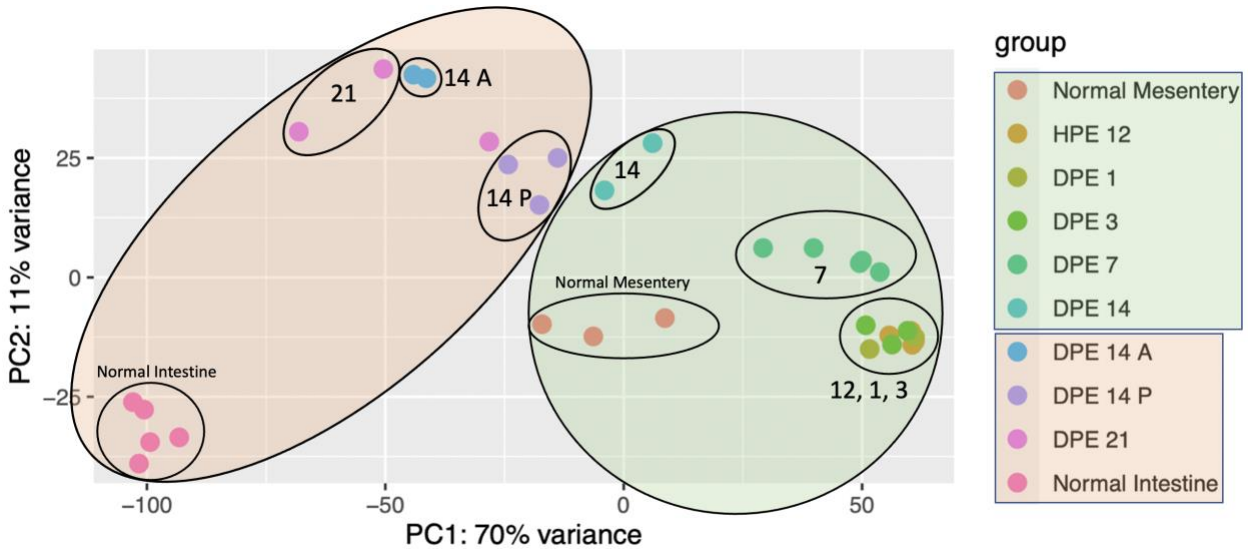


Figure 7. PCA of all RNA-seq Samples. Each colored dot represents a tissue sample from a different temporal and/or spatial timepoints of regeneration. Multiple dots of the same color represent replicates within that sample. Dots highlighted green are samples that contain only the rudiment whereas highlighted in orange are samples that contain luminal epithelium. The smaller circles show that replicates of a sample are clustered together. The numbers next to the smaller circlers indicate what samples are within. The two larger circles show that samples of rudiment and luminal epithelium are distributed spatially, thus evidencing their differences in tissue composition. Within the DPE 7, the two samples on the left are from Quispe-Parra (2021) while the three clustered together on the right are from Auger (2022).

Identifying cell specific markers in tissue samples

Although the PCA plots present evidence that each sample contains gene expression reflective of their spatial and temporal identity, it was also incumbent to provide evidence that gene markers could be correlated with specific cell types present in certain tissues. For example, do the samples with alleged luminal epithelium contain luminal epithelial specific gene markers? To parse out cellular identity in these samples, cell specific markers were searched for, to determine if they were enriched in any tissue samples. Samples that contained luminal epithelium were assessed. In 14-dpe anterior and posterior as well as 21-dpe, gene markers for enterocytes, goblet cells and enteroendocrine cells such as FABP2, FCGBP, and ADGRG4, were respectively enriched (Table 4). Another area of concern for validating RNA-seq data is confirming expression of various genes by PCR. In Figure 2D of a paper publish by Quispe-Parra (2021), various genes at 1- and 3-dpe had their RNA-seq and PCR values compared and showed that the values were complementary. These are the same data that were incorporated into the transcriptome of the present study. Currently underway is the validation of the remaining timepoints.

Table 4. Enriched Gene Markers of Samples with Luminal Epithelium Cells

Luminal Epithelium Human Cells	A	A_P	A_21	A_P_21	P	P_21	21
Enterocyte			ACE2 FABP2	MGAM			
Goblet cell	FCGBP		CLCA1	FCGBP, CLCA1	FCGBP		FCGBP, CLCA1
Glandular cell			PCNA	MUC4			PCNA
Enteroendocrine			SCGN	INSM1	ADGRG4		ADGRG4
Enterocyte & Goblet Cell		ADGRG4					
Enteroendocrine & Enterocyte			TMPRSS2				
Enteroendocrine, Enterocyte, Goblet Cell				ANXA13			

Table 4. This table shows enriched cell specific markers of luminal epithelium cells that were found within tissue samples that contained luminal epithelium during intestinal regeneration. Each row contains cell specific gene markers that are known to be found in human intestinal epithelium cells. Each column is either a single timepoint or combination of various timepoints where a cell specific marker was found. Abbreviations: A, 14 dpe Anterior. P, 14 dpe Posterior. 21, 21 dpe.

Differential Gene Expression Analysis

After validating the RNA-seq data, the dynamic state of differential gene expression during intestinal organogenesis could be explored by comparing differentially expressed transcripts (DETs) at each stage of regeneration. The term “transcript” is being used instead of “gene” since isoforms are present and since a de novo transcriptome is being used instead of a genome to quantify gene read counts. Moreover, the goal here is to create a Venn diagram representing which transcripts were uniquely differentially expressed at a timepoint or were uniquely expressed in a combination of two or more timepoints. This will provide a general insight into what’s occurring at the gene expression level. These analyses produced results that were both expected and surprising (Figure 8). First, it was expected that the tissue with luminal epithelium would be more similar among themselves than to the regenerating rudiment, which was the case. Samples with luminal epithelium also contained more DETs per timepoint than rudiment samples, perhaps reflecting that the former has an additional tissue layer. Surprisingly, 12-hpe had more than twice the number of DETs than 1-dpe. This is illustrative of the dynamic molecular processes that are occurring immediately after evisceration. It was also discovered that each timepoint contained a high percentage of unique transcripts, reflective of the unique combination of cellular events occurring at that timepoint.

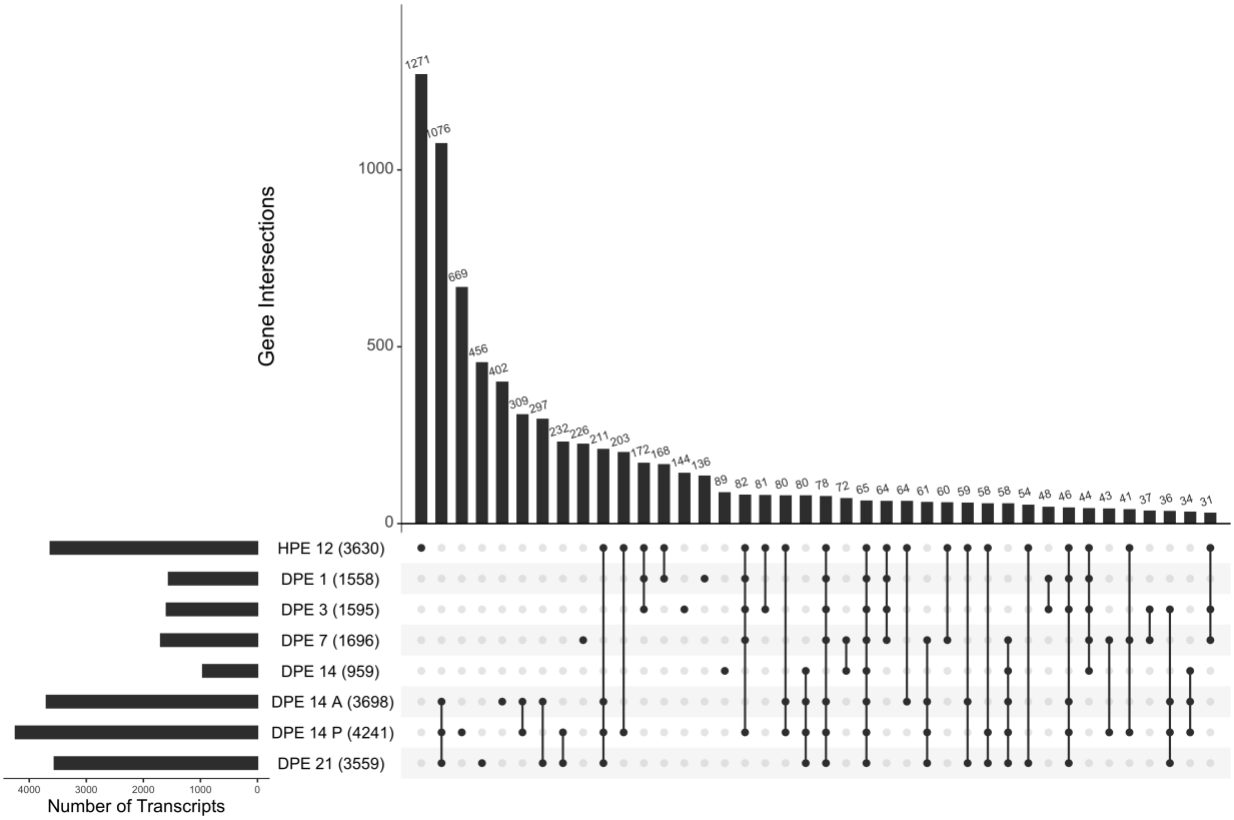


Figure 8. UpSetR Graph of Differentially Expressed Transcripts in Regenerating Samples. This graph is an alternative to a Venn Diagram but can be read in a similar manner. Each row is a different stage. Next to the stage name in parenthesis is the total number of DETs within the pooled samples representing the respective stage. The bar graph labeled Number of Transcripts is a visual representation of the total number of DETs in the samples of that stage. The bar graph labeled Gene Intersections represent the total number of DETs unique to a given stage or shared between one or more stages. For example, in the first column, HPE 12 has 3630 total DETs, but only 1271 are unique (not shared with any other timepoint). Similarly, in the second column DPE 14 A, DPE 14 P and DPE 21 are connected by lines, indicating that they share DETs, and the bar graph on top states that they share a total of 1076 DETs. DETs were defined as having an absolute log2FoldChange > 2 and a p-adjusted value < 0.05.

Also surprising was the finding that 14-dpe anterior and posterior display different DETs, suggesting that some of the molecular processes occurring at these locations may be unique. The differences between anterior and posterior regenerates will be further explored in the GSEA section. However, a list of the top 10 DETs in both anterior and posterior regenerates provide a snapshot of their differences. (Tables 5 & 6). Of the top 10, only the gene RPS27A was found in common. It is also interesting to note that the magnitude of upregulation appears to be stronger in the posterior regenerate than in the anterior.

Table 5. Top 10 Up-regulated Genes in the Anterior Intestine

Gene Name	Cluster ID	log2FoldChange	Adjust P-value
FGL2	137019.166948	27.1	1.473850e-03
RPS27A	56601.0	23.6	9.141100e-03
RpL18	63877.0	21.3	1.554267e-02
TUBB4B	62545.0	20.4	1.748147e-02
MLC1	63118.0	19.8	2.423061e-02
TRYP4	60332.2	19.3	2.764273e-02
DAP	66060.0	18.8	3.463434e-02
ATPsynF	60640.0	17.8	3.973292e-02
ERN2	6948.1	17.7	7.045030e-23
PCOLCE	64867.1	15.8	5.684820e-02

Table 6. The Cluster ID is a gene identifier from the Corset package. The Adjust P-value represents the false discovery rate for the log2FoldChange value.

Table 6. Top 10 Up-regulated Genes in the Posterior Intestine

Gene Name	Cluster ID	log2FoldChange	Adjust P-value
l(2)34Fc	58205.0	39.5	6.510121e-09
RpLP0	137019.121391	38.9	1.217922e-08
Pepck	64302.0	36.6	9.986924e-08
RPL26	66161.0	35.1	3.622796e-07
ANT	53187.0	34.7	5.068944e-07
RPS27A	56601.0	34.1	8.845984e-07
CYP9E2	55618.0	33.9	9.843082e-07
AK	60908.0	33.8	1.087509e-06
AchBP	62782.0	33.6	1.223207e-06
RpL-7	58765.0	33.4	1.430982e-06

Gene Set Enrichment Analysis

With the DETs extracted from the DGE analysis, the Gene Set Enrichment Analysis (GSEA) can be performed. This will delineate which gene ontological (GO) terms are significantly up- or down-regulated by using the log2FoldChange value of a transcript. For example, if a transcript is up-regulated and is annotated by certain GO terms, then they will appear as activated, if they have a p-value of < 0.05.

The results of the GSEA demonstrate that rudiment regeneration is dominated by activated terms associated to ribonucleoprotein complex biogenesis such as ncRNA metabolic process, RNA processing, ncRNA processing, rRNA metabolic process, ribosome biogenesis. There are two noticeable deviations from this pattern (Figure 9A-E). First, at 3-dpe, the

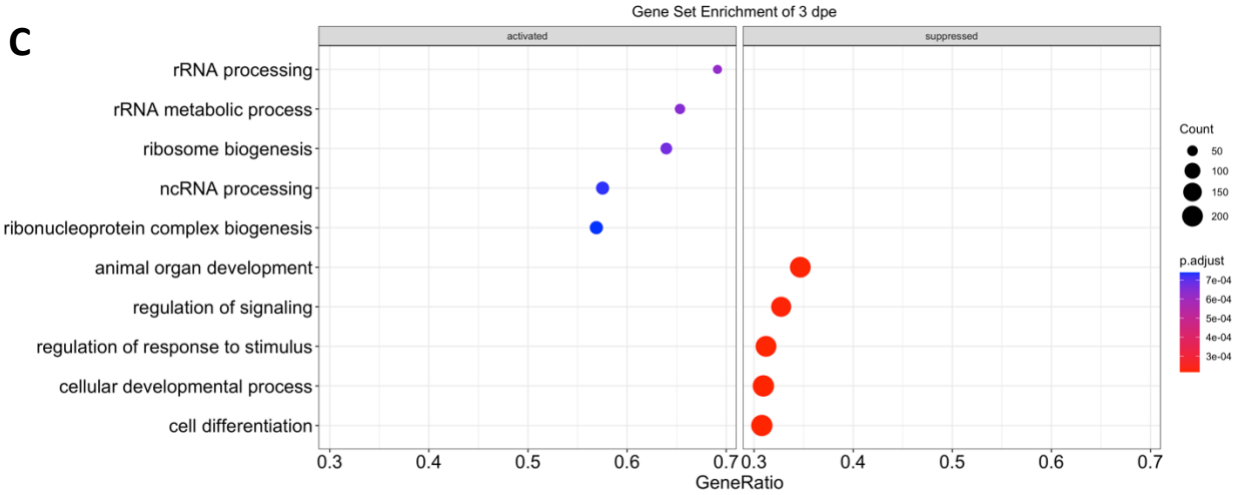
ribosome biogenesis GO terms are still activated, but suppressed GO terms overtake in gene count and are related to response to stimulus, cellular developmental process, cell differentiation, system development and response to chemical (Figure 9C). Second, at 7-dpe and 14-dpe ribosome biogenesis GO terms are still activated, but suppressed GO terms overtake in gene count and are related to cation transport, ion transmembrane transport, system processes and transmembrane transport (Figure 9D,E). Together these results suggest that ribosome biogenesis dominates early-stage rudiment regeneration, but other important molecular mechanism partake at 3-, 7- and 14-dpe.

Tissue that contains luminal epithelium displays different GO terms when compared to regenerating rudiment. For instance, no ribosome biogenesis related terms are present. As expected, 14-dpe anterior and posterior regenerates display different GO terms. The former contains termed related to cilium movement, microtubule-based process, and chromosome organization (Figure 9F). Whereas 14-dpe posterior contains terms such as PI3K regulation, ribonucleotide biosynthetic process, and development of muscle tissue (Figure 9G). 21-dpe shares similar terms with 14-dpe anterior, but contains unique terms such as apoptotic cell clearance, fatty acid biosynthetic process, and endothelium development. Together these results suggest that there are differences in regenerative molecular mechanisms not only between rudiment and tissue with luminal epithelium but also between anterior and posterior regenerates.

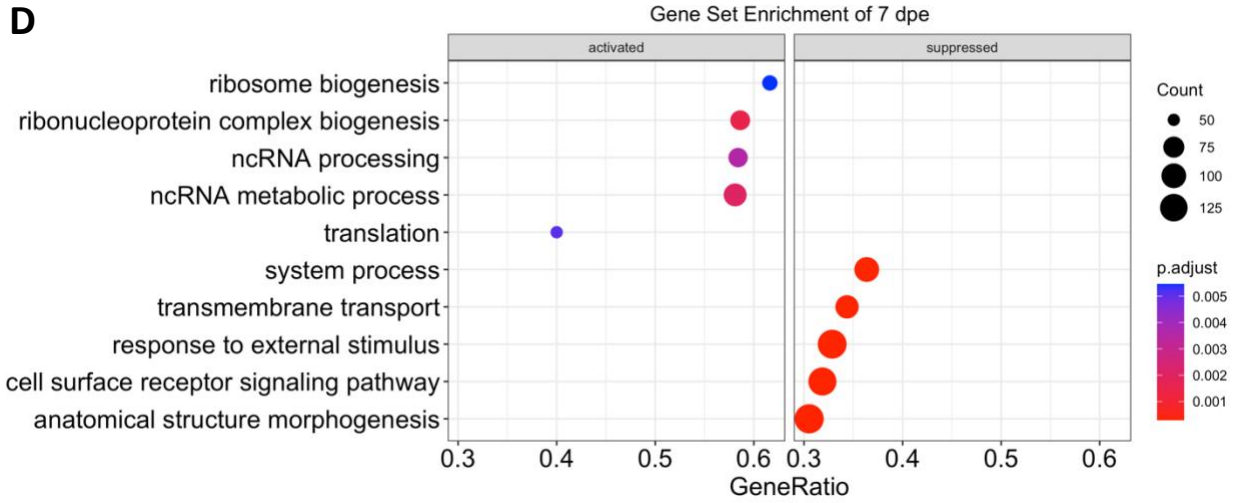
B


Gene Set Enrichment of 1 dpe
activated
suppressed
humoral immune response
ncRNA processing
ribonucleoprotein complex biogenesis
ncRNA metabolic process
RNA processing
regulation of innate immune response
defense response to bacterium
toll-like receptor signaling pathway
bone development
translation
0.4
0.5
0.6
GeneRatio
Count
30
60
90
120
p.adjust
0.00050
0.00045
0.00040
0.00035
0.00030

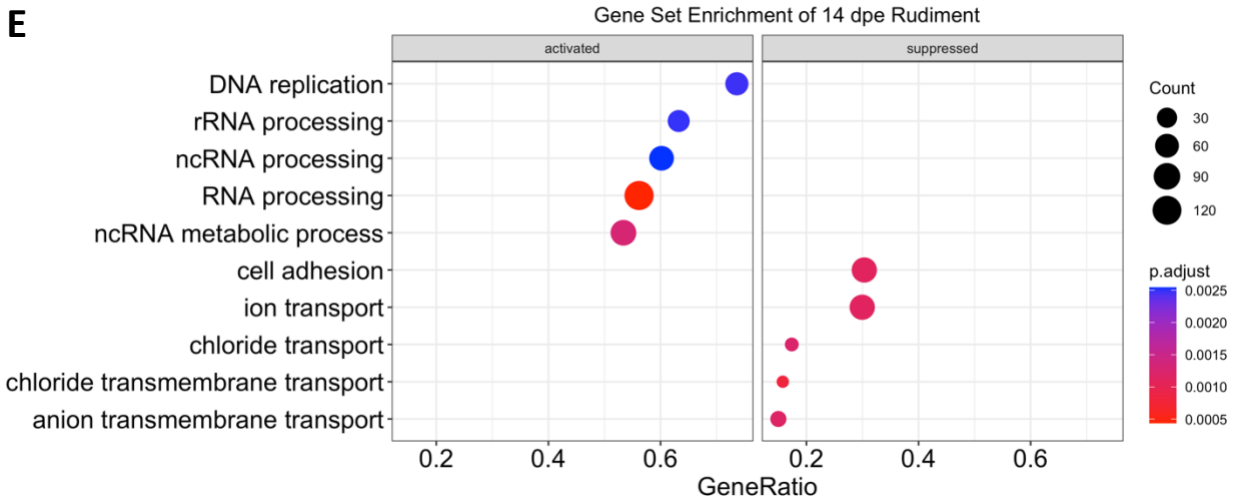
C


Gene Set Enrichment of 3 dpe
activated
suppressed
rRNA processing
rRNA metabolic process
ribosome biogenesis
ncRNA processing
ribonucleoprotein complex biogenesis
animal organ development
regulation of signaling
regulation of response to stimulus
cellular developmental process
cell differentiation
0.3
0.4
0.5
0.6
0.7
GeneRatio
Count
50
100
150
200
p.adjust
7e-04
6e-04
5e-04
4e-04
3e-04

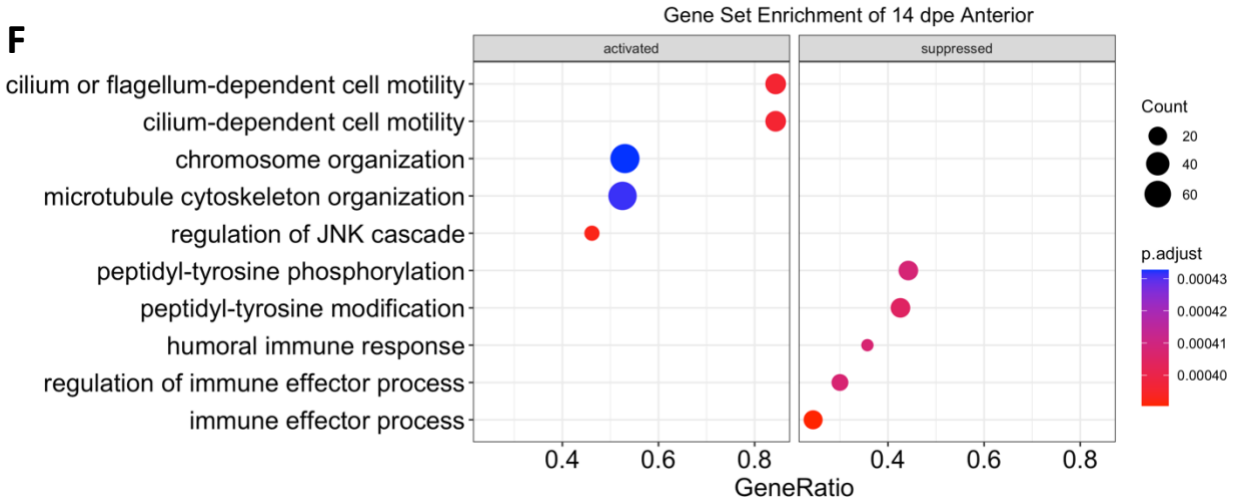
D


Gene Set Enrichment of 7 dpe
activated
suppressed
ribosome biogenesis
ribonucleoprotein complex biogenesis
ncRNA processing
ncRNA metabolic process
translation
system process
transmembrane transport
response to external stimulus
cell surface receptor signaling pathway
anatomical structure morphogenesis
0.3
0.4
0.5
0.6
GeneRatio
Count
50
75
100
125
p.adjust
0.005
0.004
0.003
0.002
0.001

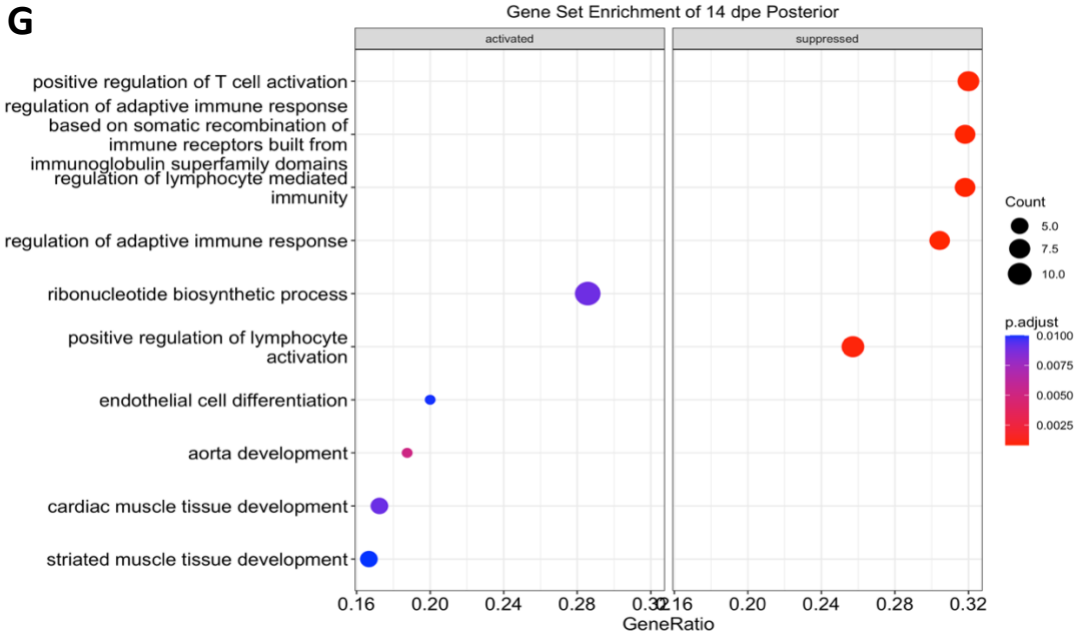
E


Gene Set Enrichment of 14 dpe Rudiment
activated
suppressed
DNA replication
rRNA processing
ncRNA processing
RNA processing
ncRNA metabolic process
cell adhesion
ion transport
chloride transport
chloride transmembrane transport
anion transmembrane transport
0.2
0.4
0.6
GeneRatio
Count
30
60
90
120
p.adjust
0.0025
0.0020
0.0015
0.0010
0.0005

F


Gene Set Enrichment of 14 dpe Anterior
activated
suppressed
cilium or flagellum-dependent cell motility
cilium-dependent cell motility
chromosome organization
microtubule cytoskeleton organization
regulation of JNK cascade
peptidyl-tyrosine phosphorylation
peptidyl-tyrosine modification
humoral immune response
regulation of immune effector process
immune effector process
0.4
0.6
0.8
GeneRatio
Count
20
40
60
p.adjust
0.00043
0.00042
0.00041
0.00040

G


Gene Set Enrichment of 14 dpe Posterior
activated
suppressed
positive regulation of T cell activation
regulation of adaptive immune response based on somatic recombination of immune receptors built from immunoglobulin superfamily domains
regulation of lymphocyte mediated immunity
regulation of adaptive immune response
ribonucleotide biosynthetic process
positive regulation of lymphocyte activation
endothelial cell differentiation
aorta development
cardiac muscle tissue development
striated muscle tissue development
0.16
0.20
0.24
0.28
0.32
GeneRatio
Count
5.0
7.5
10.0
p.adjust
0.0100
0.0075
0.0050
0.0025

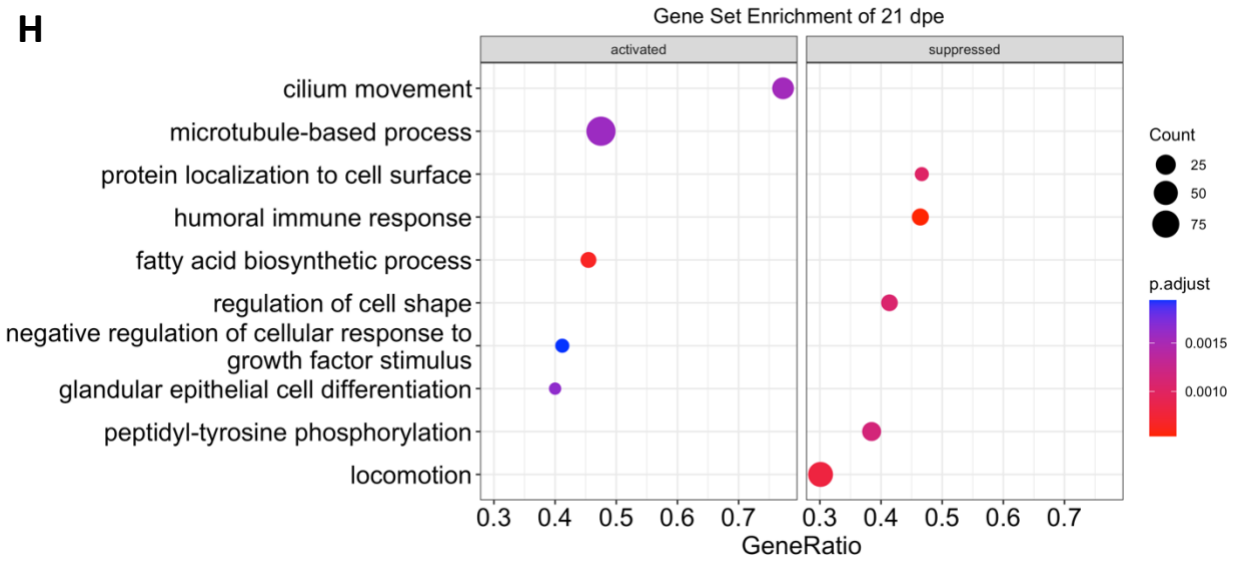
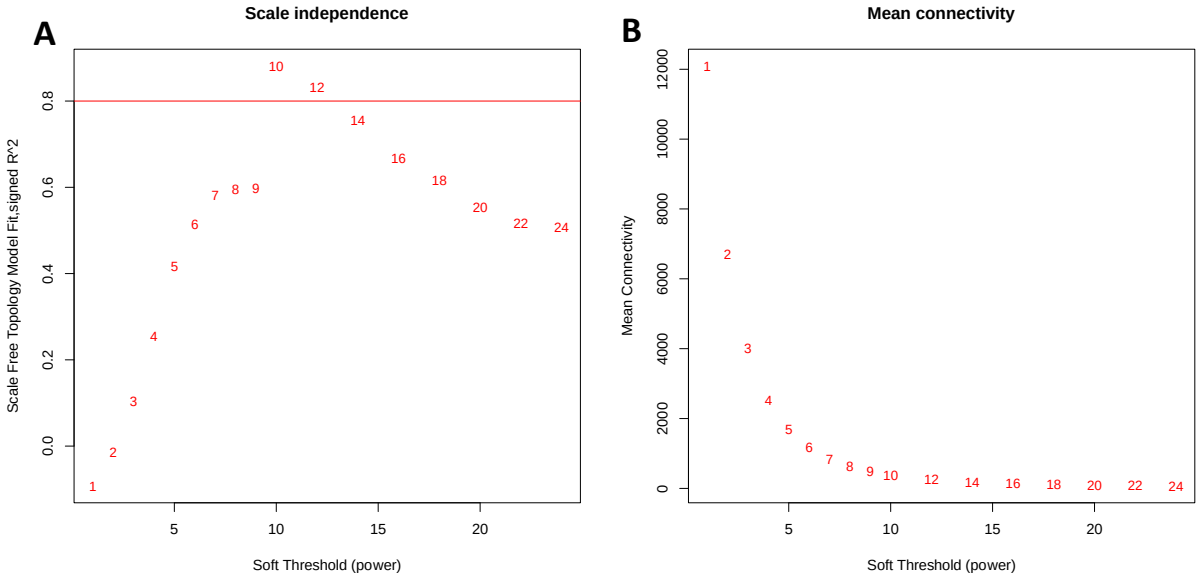


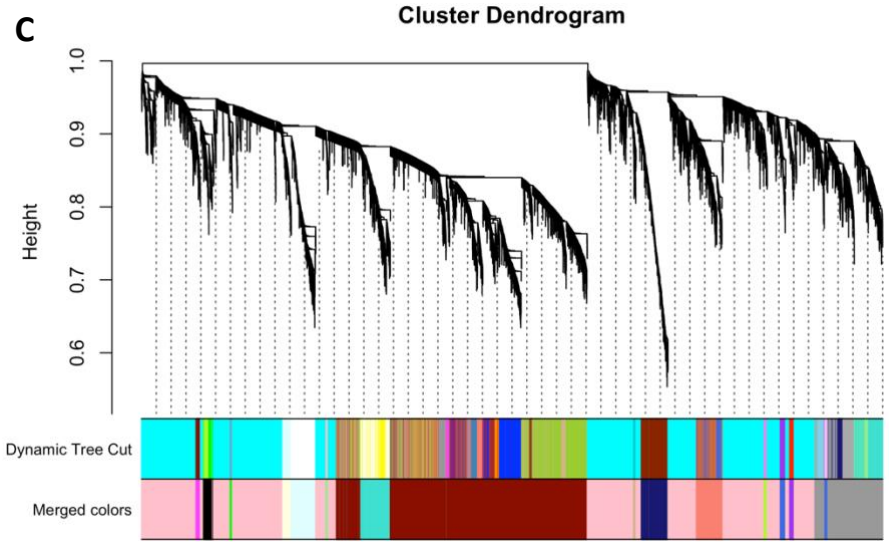
Figure 9. Gene Set Enrichment Graphs for all Regeneration Timepoints. Each graph contains biological process terms derived from GO. Each GO term has a corresponding dot present on either the activated or suppressed side of the graph. If a term is activated, then the genes that correspond to its gene-set were found to be enriched at the positive end of the ranked gene list, vice versa. The size of the dot represents the number of genes found within the ranked list that correspond to the gene-set of a GO term. The color of the dot represents the p-adjusted value for the significance of the enriched GO term. The GeneRatio is calculated by dividing the total number of genes in a gene-set of a GO term by the number of genes associated to that GO term found within the ranked list. A more comprehensive explanation can be found in Figure 5.

Gene Co-Expression Network Construction and Module Identification

The information derived from a GSEA is good for a broad overview of the molecular mechanisms of regeneration. However, more interesting are the potential drivers of these GO terms. Therefore, to explore potential gene regulators, a WGCNA was performed on all the RNA-seq data, which comprised 30 samples with a total of 23,791 transcripts. The assumption of this analysis is that genes that are co-expressed together are functioning in the same regulatory network. In this network each gene has connections to other genes. The genes with the most connections are considered at the top of the regulatory network. To perform a WGCNA analysis, a scale-free network must be constructed, whereby most transcripts have a low number of connections, and few transcripts have a high number of connections. The latter are considered hub genes. First, a similarity matrix was calculated by taking the raw expression counts of the transcripts and running the pickSoftThreshold function, which exponentiates the similarity matrix to a power that resembles a scale-free network. The scale-free network was reached when the correlation coefficient reached above a signed R^2 of 0.80 for the first time, which was at the soft-threshold β = 10 (Figure 10A). The scale free network is reflected by the mean connectivity (Figure 10B). Second, the power 10 was used to construct a signed network (where positive and negative correlations are accounted for) using the blockwisemodule function, and the modules were obtained using the dynamic cut method where the minimum number of transcripts in a module was 30, resulting in a total of 21 modules after merging the modules

(Figure 10C). The clustered dendrogram represents the clustering of modules based upon their module eigengene (Figure 10D).





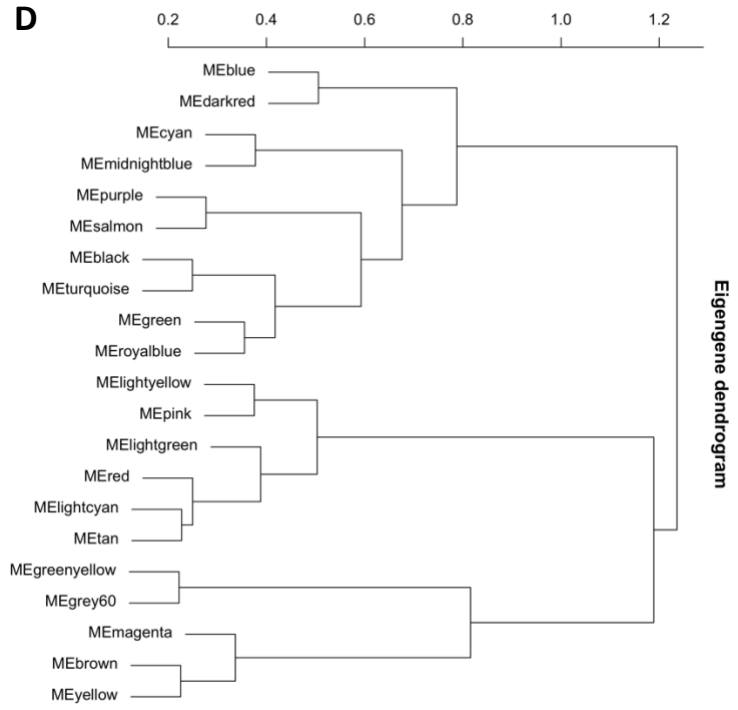


Figure 10. Construction of WGCNA Modules. **(A)** The scale independence graph demonstrates the point at which the scale-free network was reached. This occurs when the soft Threshold reaches above an R^2 value for the first time. **(B)** Mean connectivity is the data transformed into a scale-free network, which reflects the graph that was seen in Figure 6B. **(C)** The cluster dendrogram shows the clustering of genes into modules that are labeled by colors. The dynamic tree cut is the initial unfiltered clustering while the merged colors are the modules after choosing a filtering height. **(D)** The modules that were formed from the cluster dendrogram are represented in a hierarchical clustering where modules are grouped together by similarity of their module eigengene.

Module Eigengenes and Gene Ontology

The next step to parsing out useful information from the WGCNA is to validate some of the modules by using data previously collected from other experiments in the lab. A few modules can be validated. For example, in the Cyan module, 77 of its total 111 transcripts are annotated as Serum amyloid (SAA). The Cyan module eigengene shows high expression in the early stages of rudiment regeneration (Figure 1A). This data is corroborated by northern blots that show *SAA* expression increases during intestinal regeneration (Santiago et al., 2000). Another example is the Tan module where 42 of the total 139 transcripts are annotated as Tubulin. The Tan module eigengene shows an increase in expression at later stages of intestinal regeneration (Figure 1B). This is corroborated by northern blots that not only identified different tubulin isoforms being expressed during intestinal regeneration, but also found that one was upregulated during later stages of regeneration (Tossas et al., 2004). These results suggest that the WGCNA captured an accurate representation of intestinal organogenesis, and other modules may now be explored.

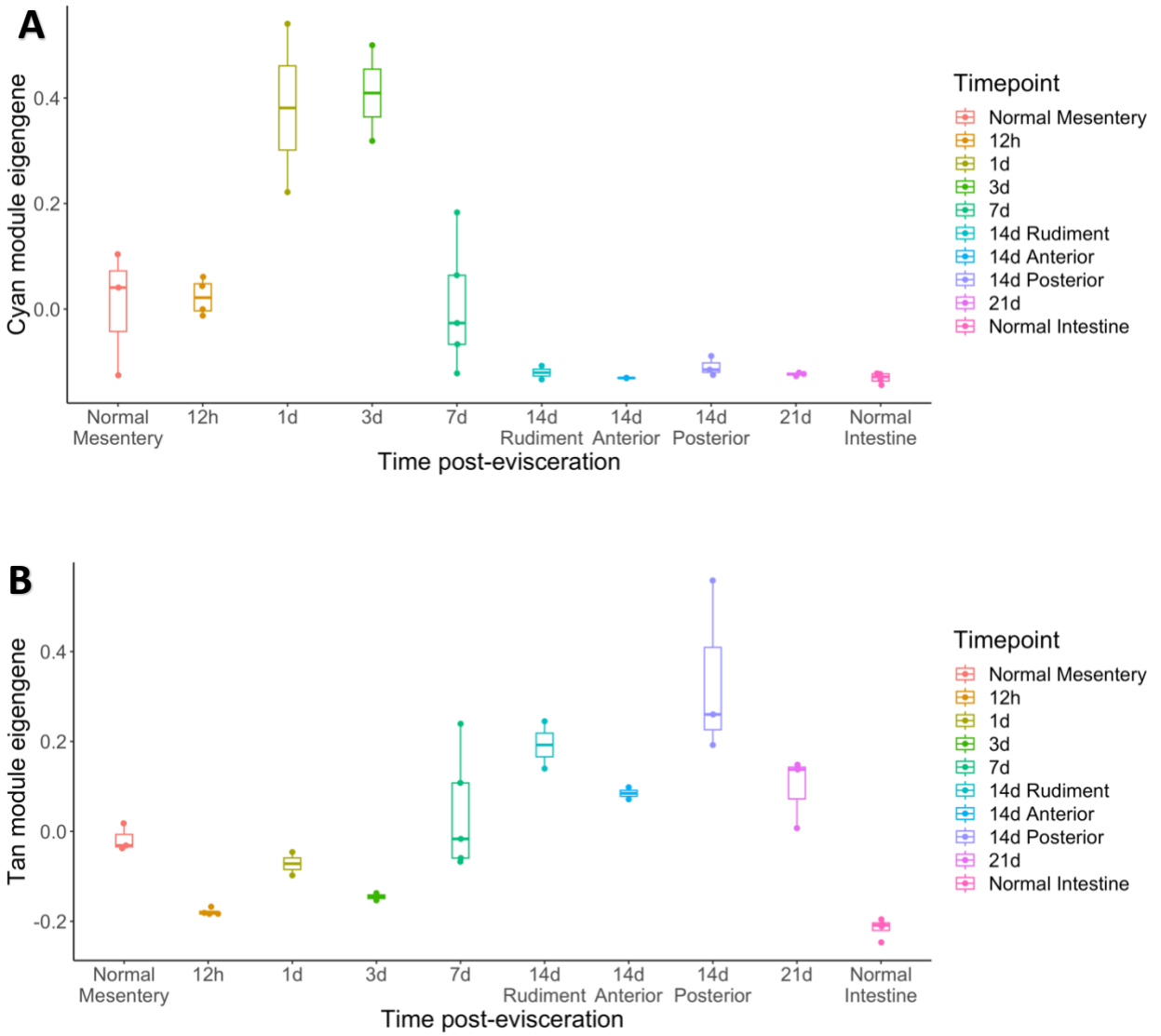
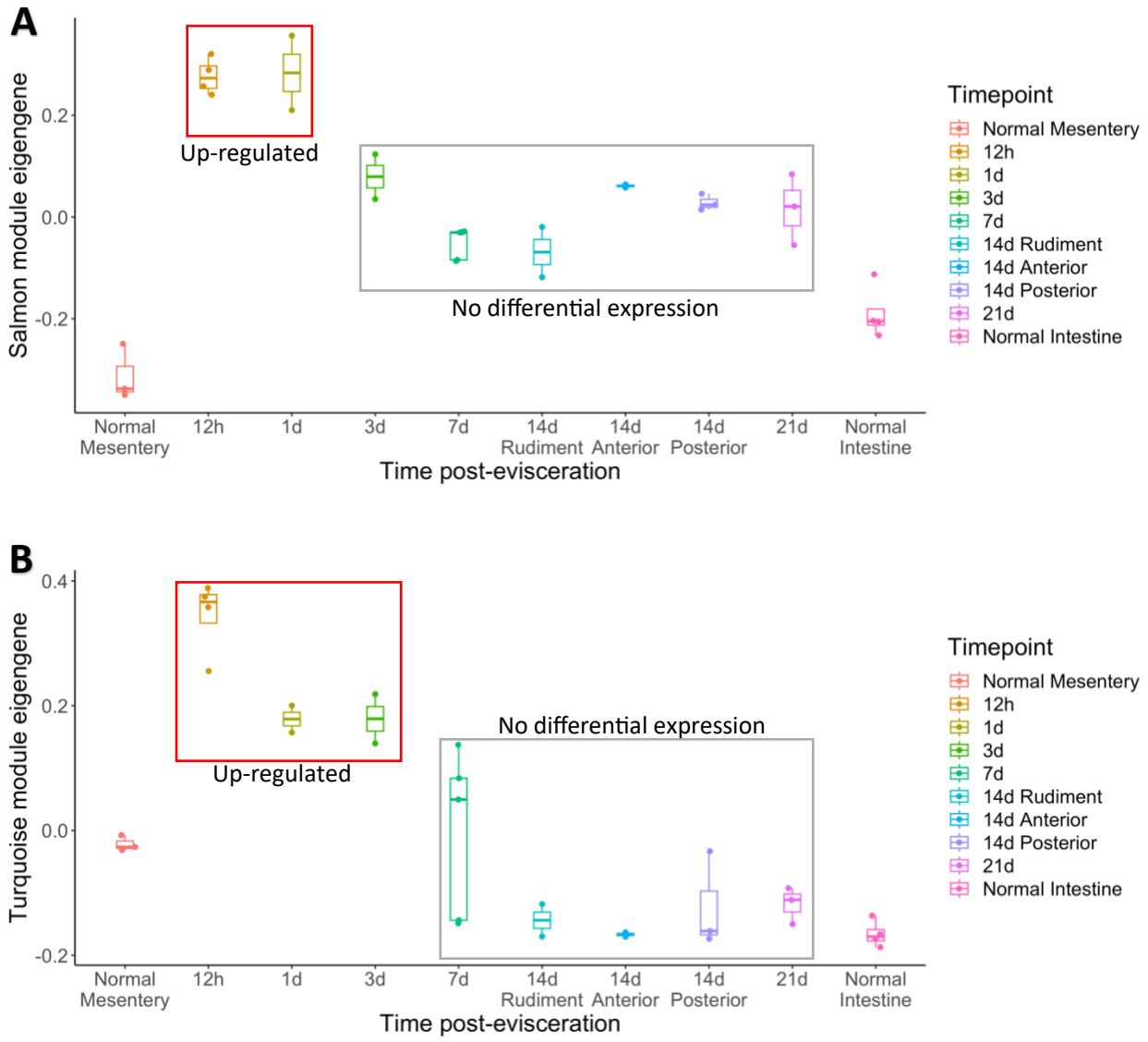


Figure 11. Validation of WGCNA Results. Module eigengenes of **(A)** Cyan and **(B**) Tan Modules. These graphs represent the eigengene of a module during intestinal regeneration. Each boxplot represents the overall expression of all transcripts within a module at a given timepoint. All the timepoints together represent the whole module eigengene across intestinal regeneration. The y-axis is the raw expression counts of the transcripts transformed into a log2FoldChange scale. Thus, if a value is negative, it means that the expression value is very low.

The Cyan and Tan modules serve to validate the results derived from the WGCNA. The next analysis with this data was to identify modules that were most likely biologically significant. Therefore, the modules were filtered in two ways. First, an Over Representation Analysis (ORA) was performed on each module with a p-value of < 0.05. This was done to isolate modules that perhaps perform a specific biological function. A GSEA was not run on these modules, since WGCNA performs calculations on unfiltered data without gene level summary statistics (i.e., log2Foldchange or p-values). Thus, modules that did not return any enriched GO terms were

discarded (Figure S2). Second, modules were filtered by looking at the transcripts at each timepoint to determine if they were significantly differentially expressed. In this way, modules that show strong enrichment of GO terms and contain transcripts that show differential expression would be isolated. This left 5 modules: Salmon, Turquoise, Green, Pink, and Red modules. Salmon and Turquoise modules were upregulated during early stages of intestinal regeneration (Figure 12A & B). Green module was upregulated in all samples that contain tissue with luminal epithelium (Figure 12C). Pink module was upregulated only in 14-dpe anterior and 21-dpe (Figure 12D). Red module was initially downregulated at 12-hpe but was then up-regulated in all samples with luminal epithelium (Figure 12E). Together these results suggest that the rudiment and samples that contain luminal epithelium have unique gene regulatory networks governing their respective cellular processes.



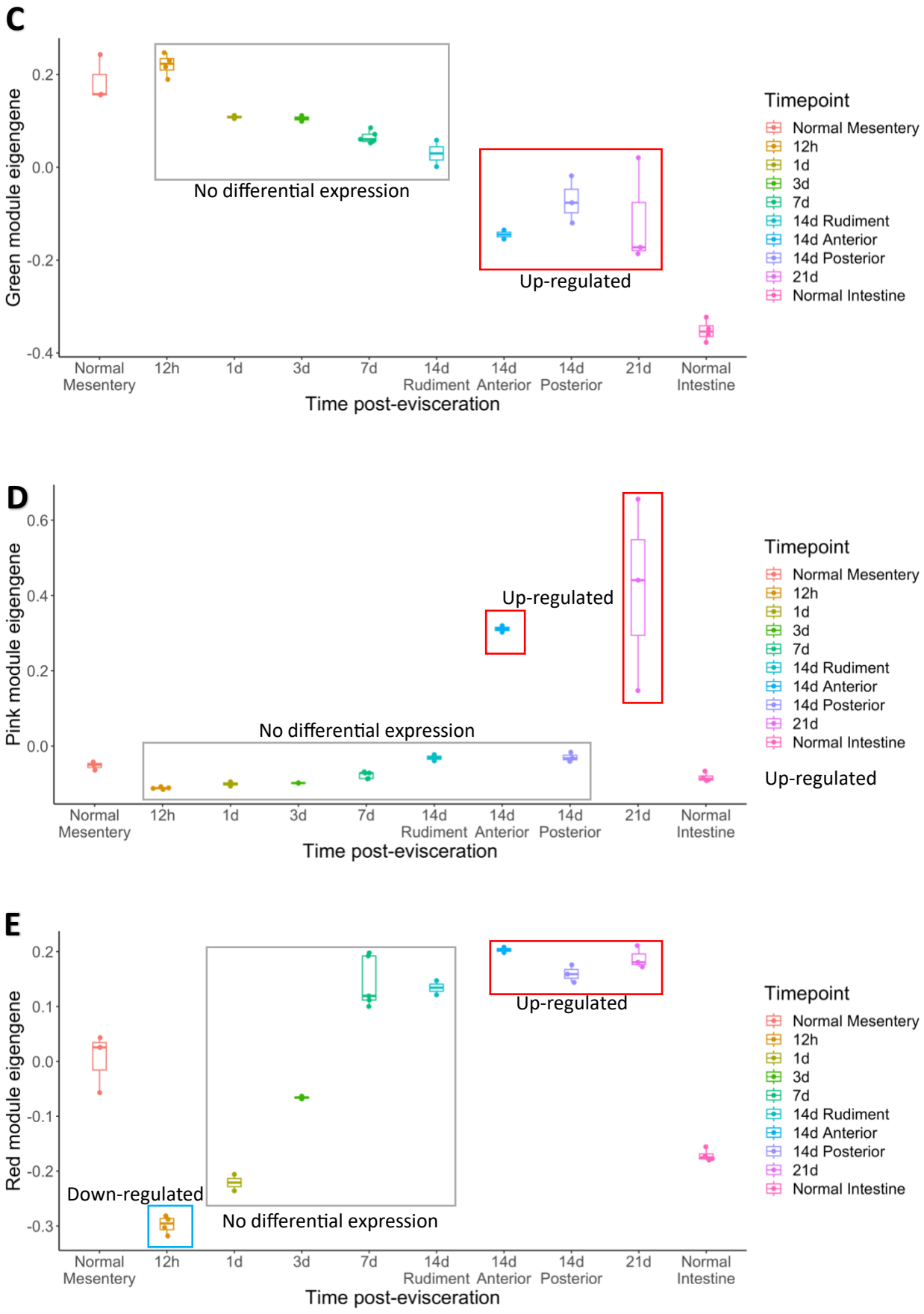

C
Green module eigengene
0.2
0.0
-0.2
-0.4
No differential expression
Up-regulated
Normal Mesentery
12h
1d
3d
7d
14d Rudiment
14d Anterior
14d Posterior
21d
Normal Intestine
Time post-evisceration
Timepoint
Normal Mesentery
12h
1d
3d
7d
14d Rudiment
14d Anterior
14d Posterior
21d
Normal Intestine
D
Pink module eigengene
0.6
0.4
0.2
0.0
Up-regulated
No differential expression
Normal Mesentery
12h
1d
3d
7d
14d Rudiment
14d Anterior
14d Posterior
21d
Normal Intestine
Time post-evisceration
Timepoint
Normal Mesentery
12h
1d
3d
7d
14d Rudiment
14d Anterior
14d Posterior
21d
Normal Intestine
Up-regulated
E
Red module eigengene
0.2
0.1
0.0
-0.1
-0.2
-0.3
Up-regulated
Down-regulated
No differential expression
Normal Mesentery
12h
1d
3d
7d
14d Rudiment
14d Anterior
14d Posterior
21d
Normal Intestine
Time post-evisceration
Timepoint
Normal Mesentery
12h
1d
3d
7d
14d Rudiment
14d Anterior
14d Posterior
21d
Normal Intestine

Figure 12. Module Eigengenes of Filtered Modules. Along the x-axis are the timepoints while along the y-axis is the raw gene count expression value of the transcripts within a module. A value on the y-axis above 0 indicates high expression while a negative value indicates low expression. If a boxplot is squared in blue, it means relative to its control that value is down-regulated. If in gray, then not differentially expressed. If in red, then up-regulated. The control for 12-hpe, 1-, 3-, 7-, and 14-dpe rudiment is normal mesentery. The control for 14-dpe anterior and posterior and 21-dpe is the normal intestine.

Module Hub Genes

The next step was to discover which are the hub genes of each module. Thus, the top 10 transcripts with the greatest connections in each module were extracted. The transcripts were then cross-referenced in the proteinatlas database to look at tissue, single cell, and cancer clustering patterns. These clusters are genes that are co-expressed within a specific tissue or single cell type as well as within a specific cancer line. By viewing how the gene is clustered (or co-expressed) with other genes, their specific functions might be postulated. Tables 7-11 provide summaries of the top 10 hub genes in each module along with information about their known functions and expression clustering in tissue types, single cell types, and cancer types.

Table 7. Hub Genes of Salmon Module

Transcript	Tissue Cluster	Single Cell Cluster	Cancer Cluster	Functions
SSB	Non-specific - Ribosome	Non-specific - Mixed function	Non-specific - Cell proliferation	The protein encoded by this gene is involved in diverse aspects of RNA metabolism, including binding and protecting poly(U) termini of nascent RNA polymerase III transcripts from exonuclease digestion, processing 5' and 3' ends of pre-tRNA precursors, acting as an RNA chaperone, and binding viral RNAs associated with hepatitis C virus. Autoantibodies reacting with this protein are found in the sera of patients with Sjogren syndrome and systemic lupus erythematosus. (ITAF)
NAT10	Non-specific - Mitochondria	Non-specific - RNA binding	Non-specific - Basic cellular processes	RNA cytidine acetyltransferase involved in histone acetylation, tRNA acetylation, the biosynthesis of 18S rRNA, and the enhancement of nuclear architecture and chromatin organization.
DHX37	Non-specific - Ribosome	Non-specific - RNA binding	Non-specific - Cell proliferation	Putative RNA helicase. Involved in alteration of RNA secondary structure such as translation initiation, nuclear and mitochondrial splicing, and ribosome and spliceosome assembly. Based on their distribution patterns, some members of this family are believed to be involved in embryogenesis, spermatogenesis, and cellular growth and division. Required for the release of the U3 snoRNP from pre-ribosomal particles
PHB1	Non-specific - Mitochondria	Non-specific - Mixed function	Breast cancer - Unknown function	Protein with pleiotropic attributes mediated in a cell-compartment- and tissue-specific manner, which include the plasma membrane-associated cell signaling functions, mitochondrial chaperone, and transcriptional co-regulator of transcription factors in the nucleus. Proposed to play a role in human cellular senescence and tumor suppression. Contributes to pulmonary vascular remodeling by accelerating proliferation of pulmonary arterial smooth muscle cells. Exact molecular functions is unclear.

EMG1	Non-specific - Mitochondria	Non-specific - Mixed function	Non-specific - Unknown function	Methylates 18s rRNA.
WDR12	Non-specific - Mitochondria	Non-specific - Mixed function	Non-specific - Cell proliferation	Involved in maturation of 12s and 5.8s rRNA. Component of the PeB oW complex, which is required for maturation of 28S and 5.8S ribosomal RNAs and formation of the 60S ribosome
ABCF1	Non-specific - Translation	Non-specific - RNA binding	Non-specific - Protein binding	This protein may be regulated by tumor necrosis factor-alpha and play a role in enhancement of protein synthesis and the inflammation process
RRP1	Non-specific - Mitochondria	Non-specific - RNA binding	Myeloid leukemia - Oxygen transport	Involved in the late stages of nucleologenesis at the end of mitosis and may be required for the generation of 28S rRNA. Isoform 2 is required for efficient Cap- and IRES-mediated mRNA translation initiation. Isoform 2 is not involved in the ribosome biogenesis
ZNRD2	Non-specific - Translation	Non-specific - RNA binding	Non-specific - Basic cellular processes	Might play a role in mitosis. Antigenic molecule. Could be a centromere-associated protein. May induce anti-centromere antibodies
ATP2C1	Choroid plexus - Transmembrane transport	Neurons - Neuronal signaling	Non-specific - Mitochondria	Supplies Ca and Mn ions to Golgi apparatus, which are necessary cofactors for processing/trafficking newly synthesized proteins.

Table 8. Hub Genes of Turquoise Module

Transcript	Tissue cluster	Single Cell Cluster	Cancer Cluster	Function
Actin				
RPL35A	Non-specific - Ribosome	Non-specific - Translation	Non-specific - Translation	Can bind initiator and elongator tRNA.
EEF1A2	Striated muscle - Unknown function	Skeletal myocytes - Muscle contraction	Breast cancer - Unknown function	Encodes alpha subunit of the elongation factor-1 complex.
CFL1	Non-specific - Cell cycle regulation	Pancreatic cells - Mixed function	Non-specific - Basic cellular processes	Can polymerize and depolymerize F-actin and G-actin.
RPS3A	Non-specific - Ribosome	Non-specific - Translation	Non-specific - Translation	Differentiation.
ARF1	Non-specific - Mitochondria & Proteasome	Non-specific - Mixed function	Non-specific - RNA binding	Small GTPase involved in protein trafficking between different compartments (PubMed:8253837). Modulates vesicle budding and uncoating within the Golgi complex (PubMed:8253837). In its GTP-bound form, triggers the recruitment of coatomer proteins to the Golgi membrane (PubMed:8253837). The hydrolysis of ARF1-bound GTP, which is mediated by ARFGAPs proteins, is required for dissociation of coat proteins from Golgi membranes and vesicles (PubMed:8253837). The GTP-bound form interacts with PICK1 to limit PICK1-mediated inhibition of Arp2/3 complex activity; the function is linked to AMPA receptor (AMPAR) trafficking, regulation of synaptic plasticity of excitatory synapses and spine shrinkage during long-term depression (LTD)

RPL7	Non-specific - Ribosome	Non-specific - Translation	Non-specific - Basic cellular processes	Binds 28S rRNA. May have regulatory role in the translation apparatus.
RPS27	Non-specific - Ribosome	Non-specific - Translation	Non-specific - Translation	Selectively regulates the expression and alternative splicing of inflammation and immune response genes. Required for proper maturation of 18S.
RPS27A	Non-specific - Ribosome	Non-specific - Translation	Non-specific - Translation	Fusion protein. Contains ubiquitin and ribosomal protein. Appears to be a novel stress sensor in the cell which amplifies p53 response to arrest cell cycle. Knockdown inhibits proliferation and induces cell cycle arrest.
A2M	Liver & Placenta - Transport via ER	Endothelial cells - Angiogenesis	Liver cancer - Metabolism	Protease inhibitor and cytokine transporter.

Table 9. Hub Genes of Green Module

Transcript	Tissue cluster	Single Cell Cluster	Cancer Cluster	Function
TENM2	Heart - Cardiac muscle contraction	Neurons - Neuronal signaling	HUVEC & TIME - Signal transduction	Involved in neural development, regulating the establishment of proper connectivity within the nervous system. Promotes the formation of filopodia and enlarged growth cone in neuronal cells. Induces homophilic cell-cell adhesion. May function as a cellular signal transducer.
KY	Striated muscle - Muscle contraction	Neurons & Oligodendrocytes - Nervous system development	Connective tissue cells - ECM organization	Probable cytoskeleton-associated protease required for normal muscle growth. Involved in function, maturation and stabilization of the neuromuscular junction. May act by cleaving muscle-specific proteins such as FLNC.
LRRC71	Ciliated cells - Cilium organization	Late spermatids - Spermatogenesis	Connective tissue cells - ECM organization	Unknown function.
CFAP251	Ciliated cells - Cilium & Cell projection	Ciliated cells - Cilium assembly	Keratinocytes - Epithelial cell function	Involved in spermatozoa motility (PubMed:30122540, 30122541). May also regulate cilium motility through its role in the assembly of the axonemal radial spokes.
WNT9	Striated muscle - Muscle contraction	Ciliated cells - Cilium assembly	Keratinocytes - Epithelial cell function	Required for normal timing of IHH expression during embryonic bone development, normal chondrocyte maturation and for normal bone mineralization during embryonic bone development.
IQCH	Ciliated cells - Cilium & Cell projection	Ciliated cells - Cilium assembly	Non-specific - Cilium assembly	May play a regulatory role in spermatogenesis.
VASH2	Testis - Cell cycle regulation	Plasmacytoid dendritic cells - Unknown function	L-1236 & L-428 - Unknown function	Tyrosine carboxypeptidase that removes the C-terminal tyrosine residue of alpha-tubulin, thereby regulating microtubule dynamics and function (PubMed:29146869). Critical for spindle function and accurate chromosome segregation during mitosis since microtuble detyronisation regulates mitotic spindle length and positioning (PubMed:31171830). Acts as an activator of angiogenesis: expressed in infiltrating mononuclear cells in the sprouting front to

				promote angiogenesis (PubMed:19204325). Plays a role in axon formation. Positive regulation of endothelial cell proliferation
PROKR2	Brain - Neuronal signaling	Early spermatids - Flagellum & Golgi organization	Breast cancer - Unknown function (mainly)	Can promote angiogenesis and induce strong gastrointestinal smooth muscle contraction.
MEOX2	Adipose tissue - ECM organization	Endothelial cells - Angiogenesis	Rabdoid cancer - Embryonic development (mainly)	Mesodermal transcription factor that plays a key role in somitogenesis and somitogenesis and limb muscle differentiation (By similarity). Required during limb development for normal appendicular muscle formation and for the normal regulation of myogenic genes (By similarity). May have a regulatory role when quiescent vascular smooth muscle cells reenter the cell cycle (By similarity). Also acts as a negative regulator of angiogenesis.
FAM183BP	Humans don't have this gene.			Predicted to be active in ciliary base.

Table 10. Hub Genes of Pink Module

Transcript	Tissue cluster	Single Cell Cluster	Cancer Cluster	Function
ACE	Intestine - Transmembrane transport	Proximal enterocytes - Transmembrane transport	Neuroblastoma - Neuronal signaling	Involved in blood pressure regulation and electrolyte balance. It catalyzes the conversion of angiotensin I into a physiologically active peptide angiotensin II.
MRC1	Lung - Lung function	Macrophages - Innate immune response	Rhabdoid cancers - Neuronal signaling	Mediates the endocytosis of glycoproteins by macrophages. The protein has been shown to bind high-mannose structures on the surface of potentially pathogenic viruses, bacteria, and fungi so that they can be neutralized by phagocytic engulfment.
DMBT1	Intestine - Transmembrane transport	Serous glandular cells - Salivary secretion	Non-specific - Enzymes	May be considered as a candidate tumor suppressor gene for brain, lung, esophageal, gastric, and colorectal cancers. May play roles in mucosal defense system, cellular immune defense and epithelial differentiation. May play a role as an opsonin receptor for SFTPD and SPAR in macrophage tissues throughout the body, including epithelial cells lining the gastrointestinal tract. Required for terminal differentiation of columnar epithelial cells during early embryogenesis

MALRD1	Intestine - Transmembrane transport	Proximal enterocytes - Transmembrane transport	Myeloma - Humoral immune response	Expression of this gene is enriched in the small intestine and is upregulated during differentiation of a human cell line that exhibits properties of intestinal epithelial cells. The encoded protein has been shown to modulate production of FGF19 in a human intestinal cell line and may regulate bile acid metabolism in the liver.
ENDOU	Esophagus - Epithelial cell function	Suprabasal keratinocytes - Cornification	Ovarian & Endometrial cancers - Unknown function	Endoribonuclease for ssRNA
MRC1	Lung - Lung function	Macrophages - Innate immune response	Rhabdoid cancers - Neuronal signaling	
CPB1	Pancreas - Digestion	Airway & Pancreas - Proteolysis	Rhabdoid cancers - Neuronal signaling	
MRC1	Lung - Lung function	Macrophages - Innate immune response	Rhabdoid cancers - Neuronal signaling	
MRC1	Lung - Lung function	Macrophages - Innate immune response	Rhabdoid cancers - Neuronal signaling	
ENDOU	Esophagus - Epithelial cell function (mainly)	Suprabasal keratinocytes - Cornification (mainly)	Ovarian & Endometrial cancers - Unknown function (mainly)	Endoribonuclease for ssRNA

Table 11. Hub Genes of Red Module

Transcript	Tissue cluster	Single Cell Cluster	Cancer Cluster	Function
PIN1	Non-specific - Unknown function	Non-specific - RNA binding	Non-specific - mRNA processing	Has profound impact on key proteins involved in the regulation of cell growth, genotoxic and other stress response, the immune system , induction and maintenance of pluripotency, germ cell development neuronal differentiation and survival.
TBCA	Non-specific - Vesicular transport	Non-specific - Translation	Non-specific - Translation	Tubulin-folding protein; involved in the early step of the tubulin folding pathway.
PARP1	Immune cells - Immune response	Cell type enriched (Testis - Spermatogonia)	B-cell cancers - Adaptive immune response	This gene encodes a chromatin-associated enzyme, poly(ADP-ribosyl)transferase, which modifies various nuclear proteins by poly(ADP-ribosyl)ation. The modification is dependent on DNA and is involved in the regulation of various important cellular processes such as differentiation, proliferation, and tumor transformation and also in the regulation of the molecular events involved in the recovery of cell from DNA damage. In addition to DNA repair, also involved in other processes, such as transcription

				regulation, programmed cell death, membrane repair, adipogenesis and innate immunity
FMO5	Liver - Metabolism	Hepatocytes - Oxidoreductase activity	Liver cancer - Metabolism	Flavin-containing monooxygenases are NADPH-dependent flavoenzymes that catalyzes the oxidation of soft nucleophilic heteroatom centers in drugs, pesticides, and xenobiotics
NCAPG	Non-specific - Cell cycle regulation	Non-specific - Cell proliferation	Non-specific - Cell proliferation	This gene encodes a subunit of the condensin complex, which is responsible for the condensation and stabilization of chromosomes during mitosis and meiosis. Phosphorylation of the encoded protein activates the condensin complex.
Rv2258c				
RIDA	Liver - Metabolism (mainly)	Proximal tubular cells - Absorption	Non-specific - Basic cellular processes	Enables 2-iminobutanoate deaminase activity and mRNA binding activity. Involved in mRNA catabolic process; mRNA destabilization; and organonitrogen compound catabolic process.
DHX9	Non-specific - Transcription	Non-specific - RNA binding	Non-specific - RNA binding	Multifunctional ATP-dependent nucleic acid helicase that unwinds DNA and RNA in a 3' to 5' direction and that plays important roles in many processes, such as DNA replication, transcriptional activation, post-transcriptional RNA regulation, mRNA translation and RNA-mediated gene silencing
DAP	Non-specific - Vesicular transport	Breast - Lactation	Non-specific - Enzymes	DAP may reduce adhesion and migration in breast cancer cell lines. Plays a role in apoptosis.
ALDH9A1	Kidney & Liver - Metabolism	Suprabasal keratinocytes - Cornification	Non-specific - RNA binding	Belongs to the aldehyde dehydrogenase family of proteins. It has a high activity for oxidation of gamma-aminobutyraldehyde and other amino aldehydes. The enzyme catalyzes the dehydrogenation of gamma-aminobutyraldehyde to gamma-aminobutyric acid (GABA)

Discussion

This study has aimed to compile various RNA-seq experiments done on the regenerating intestine of *H. glaberrima*, then to perform various bioinformatic analyses. This pursuit has delivered several major advancements in our lab: (a) a global transcriptome was made that is composed of normal mesentery, normal intestine, and timepoints from regenerating rudiment and tissue containing luminal epithelium, (2) the RNA-seq data was validated by cell specific markers and by comparing results to other experiments, (3) a GSEA displayed the dominant biological processes occurring at each stage of regeneration, (4) gene regulatory networks were found using WGCNA as well as potential hub genes (5) differences in anterior and posterior regenerates were shown as well as differences between rudiment and tissue with luminal epithelium. The rest of the discussion will be putting into perspective the potential biological significance of each module.

Salmon Module

The Salmon module appears to act during early stages of rudiment regeneration by activating ribosome biogenesis. The Salmon module eigengene is up-regulated at 12-hpe and 1-dpe whereas at all other timepoints there is no differential expression. The top 5 GO terms from the Over Representation Analysis (ORA) are rRNA processing, ribosome biogenesis, rRNA metabolic process, ncRNA processing, and ncRNA metabolic processing (Figure S2 A). All these GO terms fall under the umbrella of ribosome biogenesis, indicating that it is a paramount event during regeneration. Seven of the ten Salmon hub genes are related to ribosome biogenesis: SSB, NAT10, DHX37, WDR12, and EMG1 have functions in processing, assembly, and modifications of the pre-rRNA and ribosomal proteins (Gottlieb & Steitz, 1989; Ito et al., 2014; Choudhury et al., 2019; Hölzel et al., 2005; Singh et al., 2021). RRP1 may be necessary for the formation of 28s rRNA (Savino et al., 1999). ABCF1 is involved in protein synthesis once the mature 80s ribosome is formed (Paytubi et al., 2009). ATP2C1 aids the Golgi apparatus, which acts to package and transport newly synthesized proteins (Dode et al., 2005) (Figure 13). The remaining hub genes have various functions. PHB1 has roles in cellular senescence (Coates et al., 2001) and acts as a mitochondrial chaperone (Strub et al., 2011). ZNRD2 might have roles in mitosis, possibly associated to the centromere (Muro et al., 1998). Overall, the salmon hub genes seem to code for proteins that facilitate ribosome biogenesis, with possible ancillary functions in replication.

Turquoise Module

The Turquoise module appears to also act during early stages of rudiment regeneration through ribosome biogenesis but extends one day further than the Salmon module. The Turquoise module eigengene is up-regulated at 12-hpe, 1-, and 3-dpe but is not differentially expressed at any other timepoint. The top 5 GO terms from the ORA are protein metabolic process, organonitrogen compound metabolic process, cellular macromolecule metabolic process, cytoplasmic translation, and ribosome biogenesis (Figure S2 B). The first four GO terms denote the production of macromolecules, which could be proteins since translation is one GO term. Specifically, it could be the production of ribosomal proteins (RPs), since five of the ten hub genes are RPL35A, RPS3A, RPL7, RPS27, and RPS27S. The remaining 5 are involved in various functions. EEF1A2 delivers tRNAs to the ribosome to aid in translation (Crepin et al., 2014). ARF1 is involved in protein trafficking within the Golgi apparatus (Tanigawa et al., 1993) (Figure 13). A2M is a protease inhibitor and cytokine transporter (Sun et al., 2023). CFL1 can polymerize and depolymerize F- and G-actin (Yeoh et al., 2003). The final hub gene is Actin. Overall, the Turquoise module seems to be providing the ribosomal proteins necessary for ribosome production, and possibly performing ancillary functions related to ribosome biogenesis.

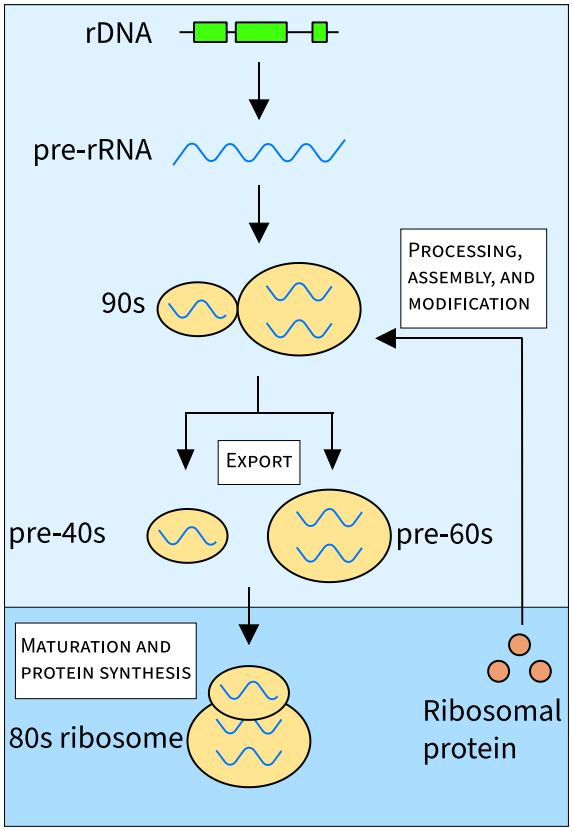


Figure 13. Role of Salmon and Turquoise Hub Genes in Ribosome Biogenesis. The graph shows a simplified version of the process of Ribosome Biogenesis. First, rDNA is transcribed into pre-rRNA, which is then processed, assembled, and modified with ribosomal proteins that are imported into the nucleus from the cytoplasm. This creates the 90s pre-ribosome, which is the precursor to the mature ribosome. The 90s pre-ribosome is then exported out of the nucleus into the cytoplasm as two individual subunits, the pre-40s and pre-60s subunits. These subunits then mature in the cytoplasm where they perform protein synthesis. This figure was adapted from Figure 1 from Pelletier et al., 2018.

Green Module

The Green module seems to be functioning during later stages of intestinal regeneration. The Green module eigengene is not differentially expressed during rudiment regeneration but is up-regulated in all timepoints that contain luminal epithelial tissue. It cannot be certain in which tissue layer(s) this module is operating. However, given a few observations, it can be postulated that the Green module is operating strictly in the mesothelium layer. First, the top 5 GO terms from the ORA are cilium organization, cilium movement, microtubule-based movement, cilium assembly, and plasma membrane bounded cell projection assembly (Figure S2 C). These GO terms are related to the cilia, but no known cell type has cilia within the intestinal luminal epithelium. However, the peritoneocytes located in the mesothelium are mono-ciliated and neuronal cells are known to have primary cilia, which are non-motile, but still contain some of the core components of motile cilia. Second, one of the hub genes is Wnt9, which has been shown to only be expressed in the mesothelium of anterior and posterior regenerates at 14-

and 21-dpe from in situ hybridization experiments (Mashanov et al., 2012). Since genes within a module are co-expressed, it could be postulated that the other hub genes show a similar expression pattern. Together this data would suggest that the Green module is confined to the mesothelium, although only up-regulated in tissue that contains luminal epithelium.

Moreover, the hub genes of the Green module tell an interesting story. Four of the hub genes are related to functional components of cilia: LRRC71, CFAP251, IQHC, and FAM183BP. Two others are TENM2 and PROKR2, both of which are involved in growth cone extension (Silva et al., 2011; Engle, 2010). VASH2 is involved in axon formation (Wang et al., 2019). MEOX2 is known to have various functions including roles in limb myogenesis and is a negative regulator of angiogenesis (Lin et al., 2005; Chen et al., 2010). KY is involved in the neuromuscular junction formation and function and may have roles in normal muscle growth (Blanco et al., 2001) (Figure 14). These results seem to suggest that the Green module is acting in the mesothelium during later stages of regeneration, possibly performing a function that develops and connects the nervous system and muscle cells. This could be done by signaling mechanisms from the primary cilia of neurons or could be a result of the mono-ciliated peritoneocytes. The former is more likely since it has been shown that primary cilia play vital roles in the development of the nervous system (Lee & Gleeson, 2011).

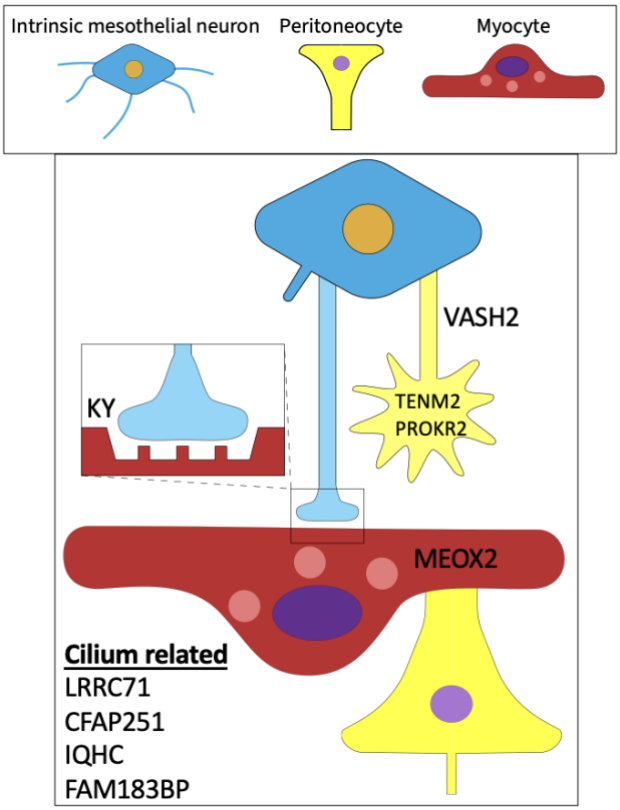


Figure 14. Possible Biological Function of the Green Module. This graph is a representation of the possible function of the Green module as inferred from the top 10 hub genes. The image is visualized as the cell types

in the mesothelium as in Figure 4. The cilia related genes in the figure are associated to the primary cilia of the neuron and/or peritoneocytes, which are depicted as the small protrusion. The yellow structure extending from the neuron is a growth cone and is associated to TENM2 and PROKR2. VASH2 is associated to the axon, depicted as a blue extension from the neuron. KY is associated to the synapse between the neuron and myoepithelial cell, enlarged in the square.

Pink Module

The pink module appears to be isolated at later stages of regeneration. The pink module eigengene is up-regulated at 14-dpe anterior and 21-dpe but is not differentially expressed at any other timepoint. The top 5 GO terms from the ORA are lipid metabolic process, cellular lipid metabolic process, small molecule metabolic process, cellular lipid catabolic process and oxoacid metabolic process (Figure S2 D). There are only 6 hub genes as some are isoforms. MRC1 mediates the endocytosis of glycoproteins by macrophages (Miller et al., 2008). DMBT1 plays roles in mucosal defense and has been shown to participate in epithelial differentiation (Kang & Reid, 2003; Mollenhauer et al., 2000). ENDOU is an endoribonuclease that binds single-stranded RNA (Laneve et al., 2008). ACE converts angiotensin I into angiotensin II (Takeuchi et al., 1986). MALRD1 may regulate bile acid metabolism (Vergnes et al., 2013). CPB1 is a procarboxypeptidase that typically acts upon amino acids (Brodrick et al., 1976). These results suggest two possible conclusions. The first is that at 14-dpe the anterior intestine is metabolically active, but the posterior portion is not. The second is that at 14-dpe the anterior intestine is performing a different metabolic function than the posterior portion. This makes sense since it is well-known in other species that the anterior and posterior intestines perform different functions. And since MRC1, DMBT1, and ENDOU are all related to immune defense, it is possibly a result of the forming lumen being exposed to outside pathogens.

Red Module

The Red module seems to be deactivated during early-stage regeneration but then reactivated during later stages. At 12-hpe the red module eigengene is down-regulated then not differentially expressed until being up-regulated in all tissues that contain luminal epithelium. The top 5 GO terms from the ORA are chromosome organization, DNA recombination, cell cycle process, mitotic cell cycle process, and cell cycle (Figure S2 E). Four of the 10 hub genes have functions in cell proliferation: PIN1 (Atchison et al., 2003), PARP1 (Sobczak et al., 2019), NCAPG (Gong et al., 2019) and DHX9 (Hou et al., 2021). Other hub genes include the monooxygenase FMO5 (Fiorentini et al., 2017), the tubulin folding factor TBCA (Lewis et al., 1996), an inducer of cell death DAP (Koren et al., 2010), DAP which is a catalyst for the dehydrogenation of gamma-aminobutyraldehyde to gamma-aminobutyric acid (Vaz et al., 2000) as well as RIDA which is involved in mRNA catabolic processes (Park et al., 2019). The last hub gene is Rv2258c, which humans do not have, and may contribute to the unique regenerative properties possessed by sea cucumbers. These results taken with the fact that studies have shown that the proliferating tip of the luminal epithelium that invades the rudiment at 14-dpe anterior and posterior has a proliferative rate of about 60%, whereas the highest that the rudiment reaches is about 10% (Unpublished data from García-Arrarás). Therefore, it can be speculated that the Red module is isolated to the luminal epithelium and facilitates proliferation.

The previous sections analyzed each independent module. However, the following sections will attempt to synthesize the data in a broader view.

Rudiment Regeneration

Early regeneration of the rudiment is dominated by ribosomal biogenesis, as revealed by GSEA. This dominance was then recapitulated by WGCNA in the Salmon and Turquoise modules. Ribosome biogenesis is the production, processing and assembly of distinct ribosomal proteins and ribosomal RNA (rRNA) that form mature ribosomes, the molecular machines responsible for protein synthesis. Ribosome biogenesis plays important roles in apoptosis, differentiation, and development with essential roles during cell growth and proliferation (Campbell & White, 2014). Hyperactivation of ribosome biogenesis can cause cancer initiation and progression (Orsolic et al, 2016; Truitt & Ruggero, 2016) and can also occur in cardiovascular, blood, neurodegenerative diseases (Prakash et al, 2019; Turi et al, 2019; Goncalves et al, 2011). These pathologies can result from dysregulation in the signaling pathways that regulate ribosome biogenesis such as MYC (Popay et al, 2021), mTOR (Iadevaia et al, 2014) and ncRNA particularly miRNA (McCool et al,2020), lncRNA (Xing et al, 2017) and cirRNA (Holdt et al, 2016). Processes associated with ribosome biogenesis have been associated to regeneration in several organisms. rRNA biosynthesis has regulatory roles during shoot regeneration in Arabidopsis (Shinohara et al., 2014) while in mouse liver *DEF* (nuclear protein that participates in ribosome biogenesis) was essential for liver homeostasis and regeneration (Jiao et al., 2023). Ribosome biogenesis is also implicated specifically in intestinal organogenesis. In zebrafish, the endonuclease Rcl1 that separates 18S rRNA from 5.8S and 25S rRNAs is essential for intestinal organogenesis (Zhu et al., 2021). Also, two nuclear proteins called Mpp10 and Sas1o, which are associated to the ribosomal small subunit processome, are essential for the development of digestive organs (Zhao et al., 2019).

This study suggests for the first time in echinoderms that ribosome biogenesis may be an essential initiator of intestinal organogenesis. The majority of the top hub genes in the Salmon and Turquoise modules are functionally related to ribosome biogenesis. Dealing with rRNA processing, assembly and modification are NAT10, RRP1, DHX37, WDR12, EMRG1 and SSB. While other top hub genes code for the ribosomal subunits Rpl35a, Rps3A, Rpl7, Rps27 and Rps27S. Moreover, these results provide support for the notation that ribosomes are specialized machines that selectively translate specific mRNA, as it was once believed that ribosomes translated any available mRNA (Genuth & Barna, 2018). For example, during mouse plantaris muscle hypertrophy, Rpl3 is upregulated while RPL3L, which has specific expression in skeletal and heart muscle, is downregulated (Chailou, 2019). In the same study it was also shown that during plantaris muscle regrowth following hindlimd suspension Rpl3 and Rpl7 were the only upregulated ribosomal proteins, suggesting differential expression of ribosomal proteins and possible ribosome specialization. This coincides with the data as Rpl7 is upregulated during early stages of intestinal regeneration in the Turquoise module, which was suggested to be isolated to the mesothelium where the muscle layer is present. These results suggest that ribosome biogenesis is an initiator of the regeneration process, possibly by supplying the cells with the molecular machinery to transcribe and translate necessary transcripts for subsequent shifts in cellular processes.

Anterior vs. Posterior Regeneration

Differential expression of genes in the anterior and posterior developing gut is well known, particularly of the cdx and homeobox gene families (Beck, 2002; Gamer & Wright, 1993). Therefore, it is not surprising that the regenerating sea cucumber intestine displays differences in anterior and posterior regenerates. GSEA revealed that the anterior intestine contains GO terms related to microtubule-based processes, cilium movement, JKN cascade regulation and chromosome organization. Whereas the posterior intestine contains GO terms related to PI3K signaling and ribonucleotide biosynthetic process. These differences are reinforced by the WGCNA results where the Pink module shows upregulation of genes in the anterior intestine that are not differentially expressed in the posterior intestine. It is also interesting to note that this module may be contained to the luminal epithelium. Since *Bmp1* is a part of the module and shows in situ hybridization only in the luminal epithelium at 14 dpe anterior and 21 dpe. Thus, one could assume that the other transcripts in this module follow a similar pattern of expression. Also, the hub genes in the Pink module include CPB1 and ACE (both enriched in proximal enterocytes) as well as DMBT1 (enriched in luminal glandular cells). These results suggest that the main difference between anterior and posterior regeneration occurs in the luminal epithelium and probably has to do with the absorptive properties of the luminal cells that define functions more closely related to those of the small intestine than to the large intestine.

Limitations

There are a few limitations to this study. For example, the earliest timepoint collected of RNA-seq is at 12-hpe, when over 3,600 transcripts are differentially expressed. This suggests that changes at the gene expression level occur rapidly and at a high volume. Thus, it may be necessary to perform RNA-seq at even earlier stages of intestinal regeneration, especially since it is known that some genes can be activated within minutes of stimulation and don't even require de novo protein synthesis. These are called immediate-early genes and have roles in stress response and differentiation (Bahrami & Drabløs, 2016), which are some of the same processes occurring early on during intestinal regeneration. Thus, these genes are not accounted for and could be essential for various cellular processes of regeneration. Moreover, the GO terms that are associated to all the genes within the analyses are derived from mammals, but the sea cucumber is an invertebrate. This creates problems as genes may have their function mischaracterized or perhaps no characterization exists at all for a gene in the context of sea cucumber intestinal organogenesis. Therefore, any function attributed to a gene in this study would have to be functionally assessed before any definitive statements can be made.

Materials and Methods

The data and code for this project can be accessed on GitHub: https://github.com/augernoah/WGCNA.git.

Animals and treatment

The animal handling and dissection methods applied in this study have been described previously (Quispe-Parra et al., 2021). In brief, adult sea cucumbers were collected from northern Puerto Rican shores at the coordinates 18°28'11.2"N 66°07'07.9"W and transported to the laboratory. They were placed in aerated sea water aquaria at room temperature until the time of evisceration, at which point they were given intracoelomic injections of 0.35 M KCl. The eviscerated sea cucumbers were then placed back into the aquaria and left to regenerate. They were later anesthetized by ice water immersion for 45 minutes and dissected under RNase-free conditions. The tissues collected after evisceration were 12-hours post-evisceration (hpe), 1-, 3-, 7-, 14- rudiment, 14- anterior, 14- posterior, and 21-days post-evisceration (dpe). However, the mesentery and intestine were collected from normal non-eviscerated sea cucumbers as controls. Timepoints 1- and 3- dpe as well as the normal mesentery used in this study were taken from a previous study (Quispe-Parra et al., 2021). The unpublished timepoints that were collected for this study are 12-hpe, 7-, 14- rudiment, 14- anterior, 14- posterior, and 21-dpe as well as a sample from the normal large intestine. The spatial division of 14-dpe can be visualized in Figure 2. The anterior and posterior portions contain the mesothelium, connective tissue, and luminal epithelium while the rudiment portion only contains mesothelium and connective tissue. Intestines containing luminal epithelium were lifted from the body wall using forceps and cut from their mesenterial connections with surgical scissors. The same dissection method was used for non-eviscerated intestine to isolate it from the mesentery. Dissected tissues were placed in an Eppendorf tube with RNAlater solution and stored at 4°C until the RNA extraction procedure was carried out. Dissected tissues were pooled together so that each Eppendorf tube contained two tissue samples, each from a different sea cucumber.

RNA extraction was done as published earlier (Quispe-Parra et al., 2021). A combination of the method established by Chomczynski (1993) using Tri-reagent (Sigma-Aldrich, St. Louis, USA) and the RNAeasy mini kit (Qiagen, Hilden, Germany) was used for RNA extraction. At least three replicates were processed for each timepoint. Each replicate contained at least two tissue samples, one sample from two different sea cucumbers. The concentration and quality of the extracted RNA were assessed using a 2100 Bioanalyzer (Agilent Technologies, USA). Only samples that showed a concentration greater than 200 ng/µL and an RNA Integrity Number (RIN) value of ≥ 8 were used for sequencing. The obtained RNA was sequenced at the Sequencing and Genotyping Facility of the University of Puerto Rico. Libraries were constructed based on the TruSeq Stranded mRNA Library Prep Kit (Illumina, USA), and paired-end sequencing was performed using an Illumina NextSeq 500 sequencer.

De novo transcriptome assembly

Raw reads from all timepoints were uploaded to the High-Performance Computing Facility of the University of Puerto Rico. Samples consisted of 12-hpe, 1-, 3-, 7-, 14- rudiment,

14- anterior, 14- posterior, and 21-dpe. FastQC v.0.11.5 (Andrews, 2010) was used to determine quality of the reads. Trimmomatic v.0.39 (Bolger et al. 2014) was used to trim the reads using the following parameters (ILLUMINACLIP:{}:2:40:15 LEADING:20 TRAILING:20 SLIDINGWINDOW:4:15 MINLEN:35). Digital normalization and k-mer trimming were carried out using the khmer package (Cruseo et al., 2015). Parameters for normalization were -p -k 20 -C 20 -M 4e9 and for khmer trimming were -V -Z 18. The de novo transcriptome was assembled with Trinity v.2.15.0 (Grabherr et ak., 2011) with the following parameters (–left left.fq –right right.fq –CPU 15 –max_memory 100G). The quality of transcriptome assembly was quantified by Transrate v.1.0.1 (Smith et al., 2016) using the parameters (--left cuke_R1.fastq --right cuke_R2.fastq --threads 32). Transcriptome completeness was measured with BUSCO v.5.2.2 (Simão et al., 2015) with parameters (-l metazoa -o buscoresults -m transcriptome).

Quantifying expression counts

The quality-controlled samples from the de novo transcriptome assembly were used to quantify gene expression. Salmon v.0.8.2 (Patro et al., 2017) with default parameters and the dumpEq flag were used to quantify gene expression and used to produce equivalence classes. Equivalence classes generated with Salmon were then passed through Corset v.1.09 (Davidson & Oshlack, 2014) with default parameters to hierarchically cluster contigs by sequence similarity and expression class.

Functional Annotation of Transcriptome

The transcriptome was annotated using two methods. The first was with Dammit v.1.1 using the databases Pfam-A (Finn et al., 2014) and uniref90 (Suzek et alk., 2007) as well as the protein sequences from the sea urchin *Strongylocentrotus purpuratus* (Accession: GCA_000002235.4). The second method was with BLAST+ v.2.9.0 (Madden, 2003) against the Swiss-Prot (Boeckmann et al., 2003) database with the parameters (-evalue 1e-5 -num_threads 16 -max_target_seqs 1). The files were then merged in R so that each transcript had two annotations.

Differential Gene Expression

Differential Gene Expression (DGE) analysis was performed following a protocol established by Quispe-Parra et al., 2021. DESeq2 v.1.38.3 (Love et al., 2014) was used to perform the DGE. Samples with less than 30 read counts were filtered from the biological replicates of each sample. Then the counts were normalized using the estimateSizeFactor command then transformed into log2FoldChange data using the variance stabilizing transformation command. The principal component analysis as seen in Figures 7 & 8 were created using this data with the plotPCA command to determine batch effects. The DGE analysis was then carried out using a pipeline documented at length here: https://github.com/devneurolab/HgWnt2023. Identification of cell specific markers of intestinal luminal epithelial in samples that contained luminal epithelium was done by gathering markers from various single-cell transcriptomic studies and from The Human Protein Atlas (Pontén et al. 2008). The expression levels of cell specific markers were searched in all timepoints to determine if they were enriched.

UpSetR graph of Differentially Expressed Transcripts

UpSetR v.1.4.0 (Conway et al., 2017) was used to create Figure 9. From each timepoint differentially expressed transcripts (DET) were extracted from the DGE analysis. DETs were defined as having an absolute log2FoldChange > 2 and a p-adjusted value of < 0.05. The transcript IDs from each timepoint were exported and inputted into the website https://bioinformatics.psb.ugent.be/webtools/Venn/ (no publication yet exists). The files created were then exported into R Studio where the UpSetR package was used to create the graph.

Gene Set Enrichment Analysis (GSEA)

The GSEA was performed using clusterProfiler v.2.1.4 (Yu et al., 2012). For each timepoint, genes were ranked in order of magnitude of lof2FoldChange, meaning that the highest positive value was ranked as number 1 while the lowest rank was the most negative value. Duplicated gene names were removed from the ranked list. The ranked list was then passed through the gseGO function with parameters (ont = BP, keyType = UNIPROT, nPerm = 5000, minGSSize = 15, maxGSSize = 800, pvalueCutoff = 0.05, verbose = T, OrgDB = human). The GSEA graphs were created with the output of the previous command by using the command dotplot with parameters (showCategory = 2, split = .sign) + facet_grid(.~sign). This step required DOSE v.3.24.2 (Yu et al., 2015).

Construction of Modules by Weighted Gene Co-Expression Analysis (WGCNA)

The raw expression counts obtained from the Cluster and Salmon procedure described above was used for the WGCNA v.1.71-1 (Langfelder & Horvath 2008). All timepoints were used which contained 30 samples and 23,791 transcripts after filtering each row for > 30 counts. This data was then put through the pickSoftThreshold function with parameters of (powerVector = powers, verbose = 5, networkType = signed). A scale-free network was achieved with the power of 10, where the R^2 first reached above 0.8. The Scale Independence and Mean Connectivity graphs in Figure 11 A & B were created by a custom code. The Cluster Dendrogram (Figure 11C) that formed the modules was made by first invoking the blockwiseModules command with parameters (maxBlockSize = 12000, TOMType = signed, power = 10, randomSeed = 1234, minModuleSize = 30, deepSplit = 2, networkType = signed, detectCutHeight 0.9). The modules were then merged using the dynamic tree cut method with the command plotDendroAndColors with parameters (hang = 0.3, guideHang = 0.05). All 21 module eigengenes were extracted from the previously formed data and graphed as a Eigengene dendrogram in Figure 11D.

Filtering for Biologically Significant Modules

The 21 modules were then filtered by two methods, attempting to find the modules that may be the most biologically significant. First, each modules' transcripts were subject to an Over Representation Analysis (ORA) by the R package UniprotR v.2.2.2 (Soudy et al., 2020). This was achieved by converting the Uniprot IDs of a transcript within a module into a GO term to determine if the GO term was functionally enriched. The command used was Enrichment.BP, which looks at GO terms related to biological processes, with the parameters (OS=hsapiens, p_value = 0.05). For a module to proceed to the next round of filtering, it had to have had at least one enriched GO term.

Modules were then further filtered by only selecting ones that had each one of the top 10 hub genes differentially expressed at most or all timepoints. Tob hub genes were identified by finding the transcripts that had the highest intramodular connectivity by looking at module memberships values and the associated p-value. The code for this can be seen in the GitHub page. Once identified, the log2FoldChange and p-adjusted values of each hub each at each timepoint were manually verified using the DGE data previously created. The module eigengene graphs were plotted using a custom code and can be seen in the GitHub page.

Supplemental Figures

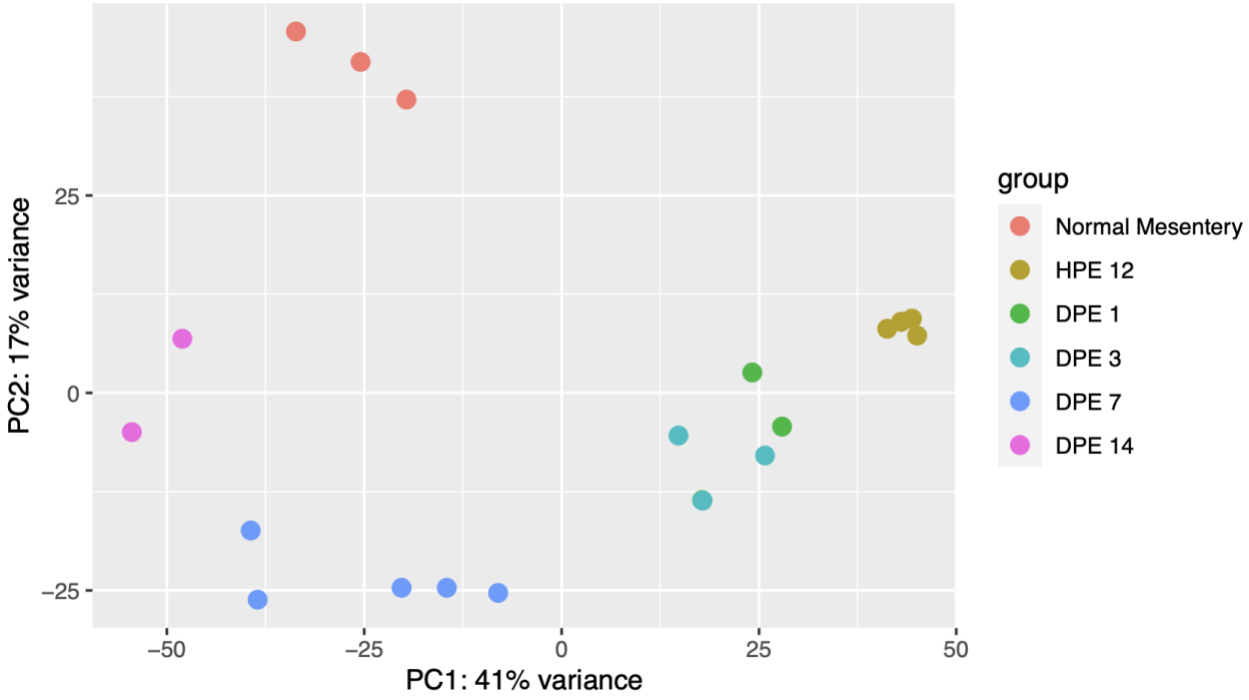
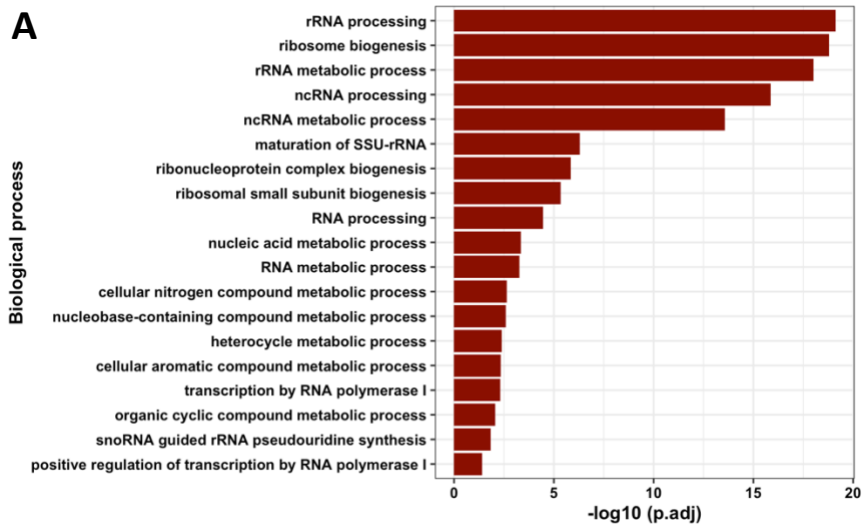
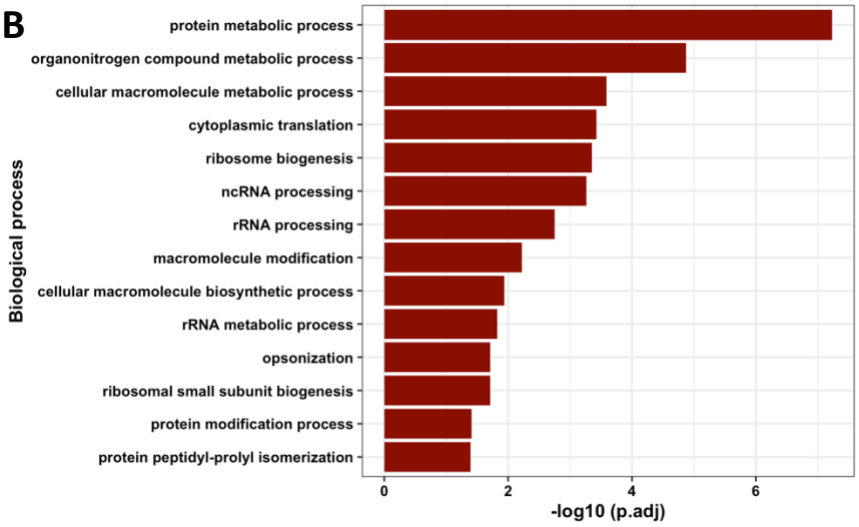
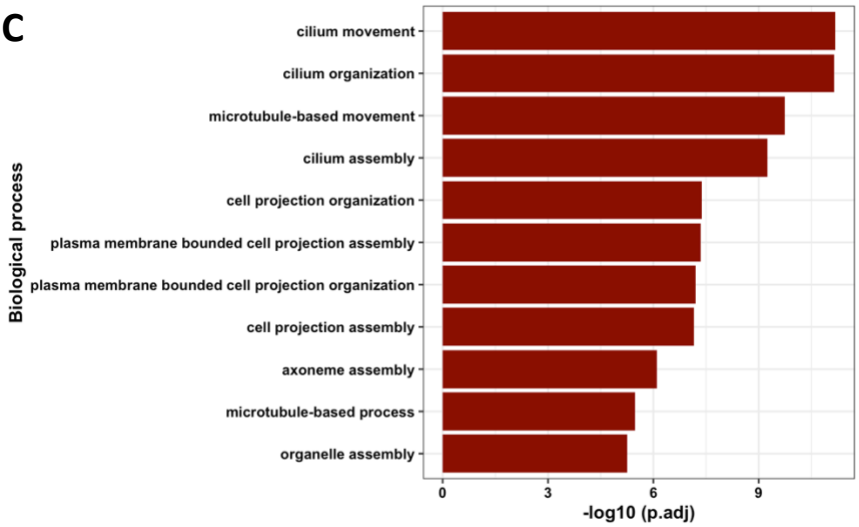


Figure S1. PCA of Samples from Rudiment. Each colored dot represents a timepoint, while multiple dots of the same color represent replicates.


A
rRNA processing
ribosome biogenesis
rRNA metabolic process
ncRNA processing
ncRNA metabolic process
maturation of SSU-rRNA
ribonucleoprotein complex biogenesis
ribosomal small subunit biogenesis
RNA processing
nucleic acid metabolic process
RNA metabolic process
cellular nitrogen compound metabolic process
nucleobase-containing compound metabolic process
heterocycle metabolic process
cellular aromatic compound metabolic process
transcription by RNA polymerase I
organic cyclic compound metabolic process
snoRNA guided rRNA pseudouridine synthesis
positive regulation of transcription by RNA polymerase I
Biological process
0
5
10
15
20
-log10 (p.adj)


B
protein metabolic process
organonitrogen compound metabolic process
cellular macromolecule metabolic process
cytoplasmic translation
ribosome biogenesis
ncRNA processing
rRNA processing
macromolecule modification
cellular macromolecule biosynthetic process
rRNA metabolic process
opsonization
ribosomal small subunit biogenesis
protein modification process
protein peptidyl-prolyl isomerization
Biological process
0
2
4
6
-log10 (p.adj)


C
cilium movement
cilium organization
microtubule-based movement
cilium assembly
cell projection organization
plasma membrane bounded cell projection assembly
plasma membrane bounded cell projection organization
cell projection assembly
axoneme assembly
microtubule-based process
organelle assembly
Biological process
0
3
6
9
-log10 (p.adj)

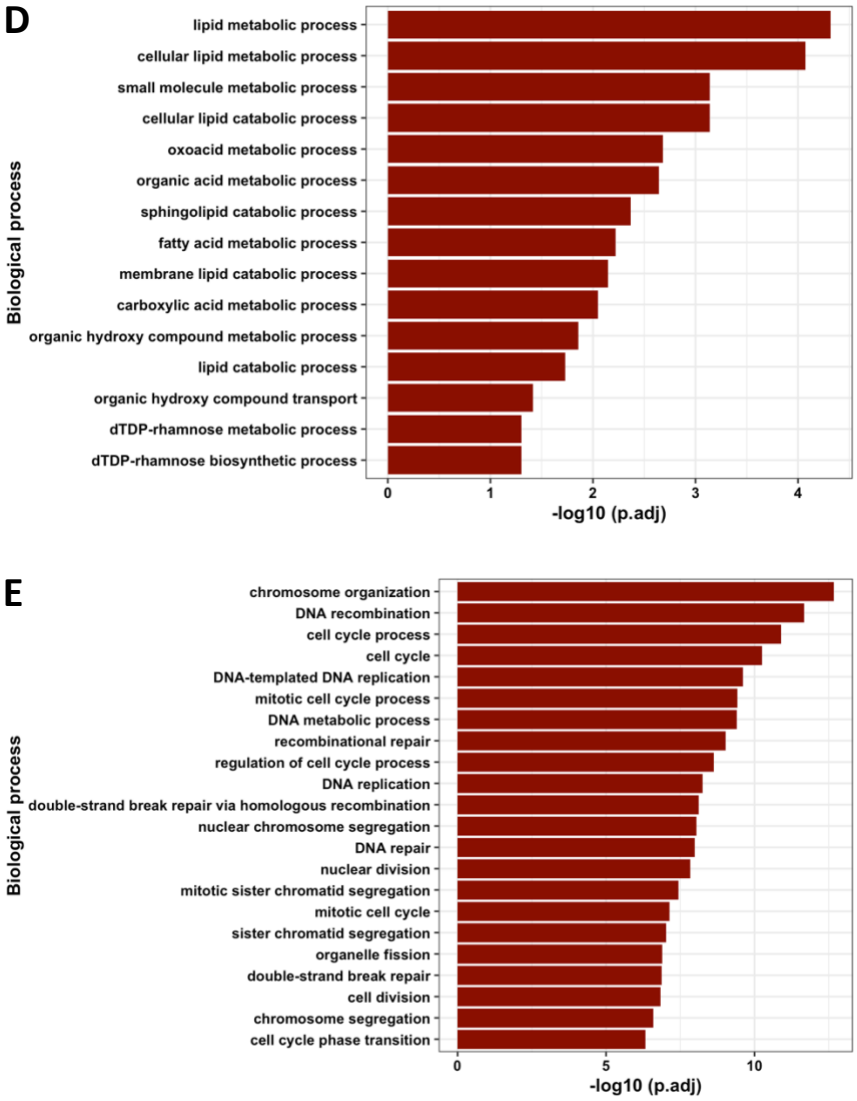


Figure S2. Over Representation Analysis results from filtered Modules. A GO term was considered enriched if found to have a p-adjusted value <0.05. All GO terms are related to biological processes.

Chapter 3: Wnt Gene Family as a Case Study

This chapter has been published in the journal *Genes*.

Auger, N. A., Medina-Feliciano, J. G., Quispe-Parra, D. J., Colón-Marrero, S., Ortiz-Zuazaga, H., & García-Arrarás, J. E. (2023). Characterization and Expression of Holothurian Wnt Signaling Genes during Adult Intestinal Organogenesis. *Genes*, *14*(2), 309. https://doi.org/10.3390/genes14020309

Introduction

The Wnt gene family encodes ligands that bind to the Frizzled (Fzd) family of cell surface receptors. This, in turn, activates canonical and noncanonical Wnt signaling pathways that are known to modulate various cellular processes such as morphogenesis, cell fate specification and proliferation during embryonic/adult development, and regeneration (Yaglova et al., 2019; Komiya et al., 2008; Otto et al., 2008). The canonical pathway, known as Wnt/β-catenin, can regulate proliferation, differentiation, and survival by affecting gene regulation through translocation of β-catenin into the nucleus that then interacts with the T cell factor/lymphoid enhancer factor (TCF/LEF) family (He et al., 1998; Tetsu et al., 1999; Kim et al., 2022; Wu et al., 2008).

The noncanonical pathway, known as the Planar Cell Polarity (PCP) pathway, can regulate cell migration and polarity by cytoskeleton rearrangements and activation of the transcription factors c-JUN and AFT2 (Amano et al., 2010; Golenia et al., 2017). These Wnt activated pathways share genes such as *Wnt* and *Fzd*, but some genes are unique to a pathway. For example, the Wnt/PCP pathway contains the so called “core” proteins, i.e., Celsr, Vangl, Prickle and Ankrd6 (Shi, 2022; Mentink et al., 2018). For a more in-depth view of the genes involved in and the signaling dynamics of the Wnt/β-catenin and Wnt/PCP pathways, see Figure 15.

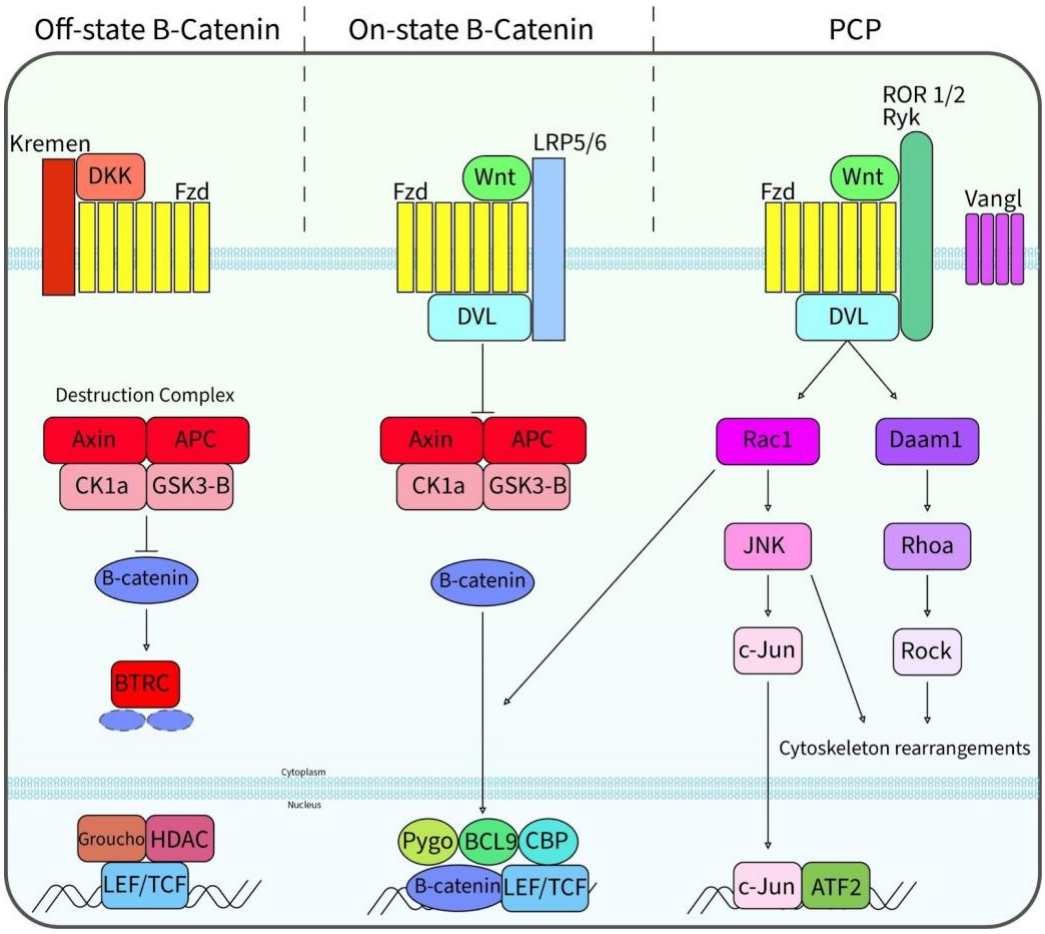


Figure 15. Overview of Signaling Genes in the Wnt/β-catenin and Wnt/PCP Pathways. Wnt/β-catenin has an Off- and On-state. When Wnt/β-catenin is in the Off-state, Kremen and DKK prevent Wnt from binding Fzd while β-catenin is labeled for proteasomal degradation by the destruction complex and in the nucleus the LEF/TCF transcription factor is inhibited by the corepressors Groucho and HDAC. In the On-state, Wnt binds Fzd and recruits the co-receptor LRP5/6, then DVl attaches and inhibits the destruction complex, allowing β-catenin to translocate in the nucleus where it binds to LEC/TCF and to other co-activators. One of the downstream effects of the Wnt/PCP is the recruitment of Vangl to the cell-membrane which can alter cell polarity. Although these pathways are distinct, there is crosstalk between them. For example, RAC1 has been shown to help translocate β-catenin into the nucleus, although being a part of Wnt/PCP pathway (Wu et al., 2008). Additionally, Vangl can inhibit canonical Wnt signaling (Mentink et al., 2018).

The cellular processes described above that are modulated by Wnt signaling can be species, organ, and cell specific. They also vary depending on developmental stage or whether the regenerative process is for cell maintenance (homeostatic) or injury response. One of the most studied functions of Wnt signaling is the self-renewal of the intestine luminal epithelium. Studies on distant groups, such as insects (Drosophila) and mammals, even hint at common mechanisms involving Wnt mediating the control of luminal epithelium regeneration (Tian et al., 2018; Perochon et al., 2018; Ouladan et al., 2021). In mammals, several Wnts have functions during gastrointestinal stem cell homeostasis, such as Wnt1 (Mao et al., 2014), Wnt2b (Suh et

al., 2017), Wnt3a (Farin et al., 2012), Wnt5a (Miyoshi et al., 2012), Wnt6 (Sigal et al., 2017), and Wnt9b (Gregorieff et al., 2005). In Drosophila, Wnt might play a role in stem cell maintenance (Lin et al., 2008), but more recent studies suggest a critical role for enterocyte homeostasis (Buchin et al., 2013; Wang et al., 2016; Tian et al., 2016).

Although the above studies characterize Wnt signaling during the maintenance (homeostatic regeneration) of certain cell types in the intestine, Wnt signaling is more complex and can be involved in the regeneration of a complete adult organ. Therefore, to obtain a holistic view of Wnt signaling during regenerative organogenesis, we employed a well-suited model organism, the sea cucumber *Holothuria glaberrima* (Quispe-Parra et al., 2021). This echinoderm, like most members of the Echinodermata phylum, has remarkable regenerative properties. The histological and cellular events during sea cucumber intestinal regeneration are well-characterized and have been described in previous publications (Quispe-Parra et al., 2021; García-Arrarás et al., 2019; Medina-Feliciano et al., 2021), as well as in the introduction section of Chapter 1.

The molecular processes that define some of the cellular events during sea cucumber intestinal organogenesis were recently investigated. *Wnt* genes are seen as key molecular regulators (Medina-Feliciano et al., 2021). For example, microarray analysis and qPCR showed that *Wnt9* was upregulated at 3-, 7-, and 14-dpe during intestinal regeneration (Ortiz-Piñeda et al., 2009). Moreover, in situ hybridization confirmed high expression of Wnt9 in the luminal epithelium of early regenerates (Mashanov et al., 2012). More recently, qPCR confirmed *Wnt6* upregulation during early intestinal regeneration (Quispe-Parra et al., 2021). It was even shown that Wnt/β-catenin signaling controls cell proliferation but not cell differentiation or apoptosis after RNAi of β-catenin during early-stage regeneration (Alicea-Delgado et al., 2021).

Wnt signaling is also upregulated during intestinal regeneration in two other holothurians: *Apostichopus japonicus* and *Eupentacta fraudatrix*. In *A. japonicus,* it was shown by qPCR that *WntA* and *Wnt6* were upregulated in early- and late-stage regeneration stages (Li et al., 2017; Sun et al., 2013). A different study, also using qPCR, showed that *Wnt7* and *Wnt8*, along with *Fzd7* and Dishevelled (*Dvl*), were upregulated at early-stage regeneration but either lowly expressed or downregulated in the later stages (Zhang et al., 2017; Yuan et al., 2019). Studies of intestinal regeneration in *E. fraudatrix* showed the *Wnt4*, *Wnt6*, *Wnt16*, *Fzd1/2/7*, *Fzd4*, and *Fzd5/8* genes to be differentially expressed by qPCR in early- and late-stage regeneration (Girich et al., 2017). All these studies provide some insight into the presence and expression of *Wnt*s and the genes involved in Wnt pathways during intestinal regeneration. However, they also present problems when trying to make comparisons. To name a few, some studies focused only on some *Wnt* or *Fzd* genes while neglecting downstream genes in the signaling cascade. In other studies, the exact developmental stage of the tissue remained unclear and, in some cases, even the organ or tissue composition of the sample could not be definitively determined from the authors' description. The other major problem is that these studies often used inadequate tissue comparisons as controls. All these problems create a patched landscape of studies of Wnt signaling during intestinal regeneration.

Therefore, the present work focuses on *Wnt* gene identification and the expression of Wnt signaling genes during intestinal organogenesis from a holistic point of view. To do this, we first identified the *Wnt* genes found in the sea cucumber genome and compared them to those of other echinoderm species. We next characterized their expression at various regeneration stages and in different tissue types. Lastly, we tried to determine which Wnt signaling pathways were active at different timepoints. This study adds to the growing field of research of intestinal regeneration in echinoderms by supplying new spatiotemporal timepoints with appropriate controls for early- and late-stage regenerates, as well as tying together the available data on Wnt signaling during intestine regeneration in echinoderms and other species. We hope that this extensive spatiotemporal RNA-seq data will aid in the identification of master genetic regulators of organ regeneration that can later be assayed for downstream functional analysis.

Results

Wnt Gene Identification and Manual Annotation

The initial identification of *H. glaberrima Wnt* genes was done using *Wnt* gene sequences from the sea urchin *S. purpuratus* and the sea cucumber *E. fraudatrix.* These sequences were used to probe the *H. glaberrima* transcriptome database for sequences that showed considerable similarities. The obtained sequences were further characterized by BLAST against the NCBI non-redundant database (Table 12). Comparisons were also made between *Wnt* genes from three additional echinoderm species, i.e., the green sea urchin *Lytechinus variegatus,* the sea cucumber *A. japonicus,* and the sea star *Anchaster planci*, all of which demonstrated high levels of similarity with their corresponding homologs. Thus, we were able to identify the presence of transcripts for 12 different *Wnt* genes: *Wnt1*, *Wnt2*, *Wnt3*, *Wnt5*, *Wnt6*, *Wnt7*, *Wnt9*, *Wnt10*, *Wnt16*, *WntA,* and a possible duplication of *Wnt4* (*Wnt4a* and *Wnt4b*).

Table 12*. Wnt Gene Characterization through NCBI Basic Local Alignment.*

Wnt	Match Species	Accesion	% Identity	E-Value
Wnt1	*L. variegatus*	MK029663.1	68.19	4e-67
Wnt2	*E. fraudatrix*	MK318552.1	74.92	3e-177
Wnt3	*E. fraudatrix*	MK318553.1	71.60	2e-135
Wnt4a	*E. fraudatrix*	KU061282.2	67.83	2e-83
Wnt4b	*E. fraudatrix*	KU061282.2	72.17	8e-146
Wnt5	*H. leucospilota*	AB969706.1	96.63	0.0
Wnt6	*A. japonicus*	JQ753331.1	76.18	0.0
Wnt7	*E. fraudatrix*	MK318555.1	73.12	1e-155
Wnt9	*E. fraudatrix*	MK318557.1	66.49	2e-59

Wnt10	*A. planci*	XM_022242684.1	75.16	2e-15
Wnt16	*E. fraudatrix*	KT362220.1	70.00	5e-10
WntA	*A. japonicus*	KU888892.1	77.46	0.0

The identification of these transcripts allowed us to manually annotate all of them in the recently published draft genome of *H. glaberrima* (Medina-Feliciano et al., 2021) (Figure 16). Confirming the complete sequence of the identified transcripts, exons of annotated Wnt transcripts were located at distinct, non-overlapping regions in the genome, where we obtained their sequence from start to stop codon. Although most of the *Wnt*s exons spanned across distinct scaffolds, the order of the exons with exact matches for nucleotide and amino acid sequences was maintained for each *Wnt***,** providing sufficient information to map each exon to its respective Wnt gene. The annotations showed that all Wnt family genes in *H. glaberrima* ranged from four to six exons, with similar lengths compared to their intronic regions, which differed from gene to gene. To allow such comparisons, in Figure 16 we depict the length based on unit conversions; therefore, exon and intron sizes are based on the distances between base pairs.

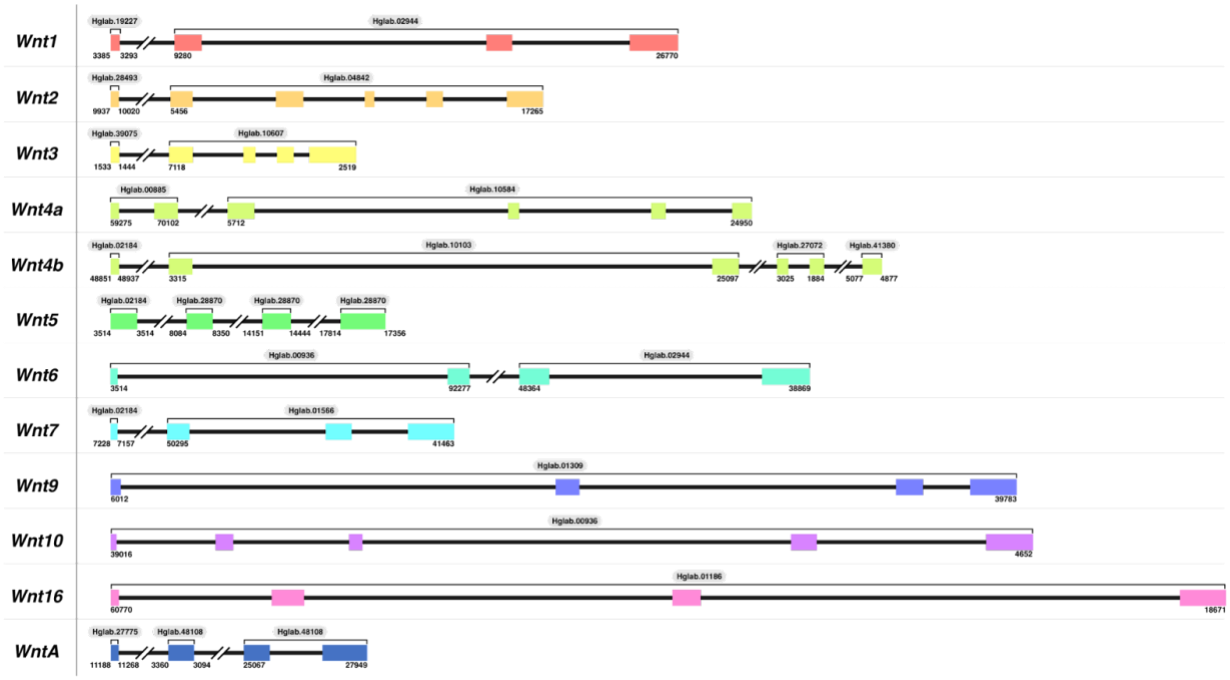
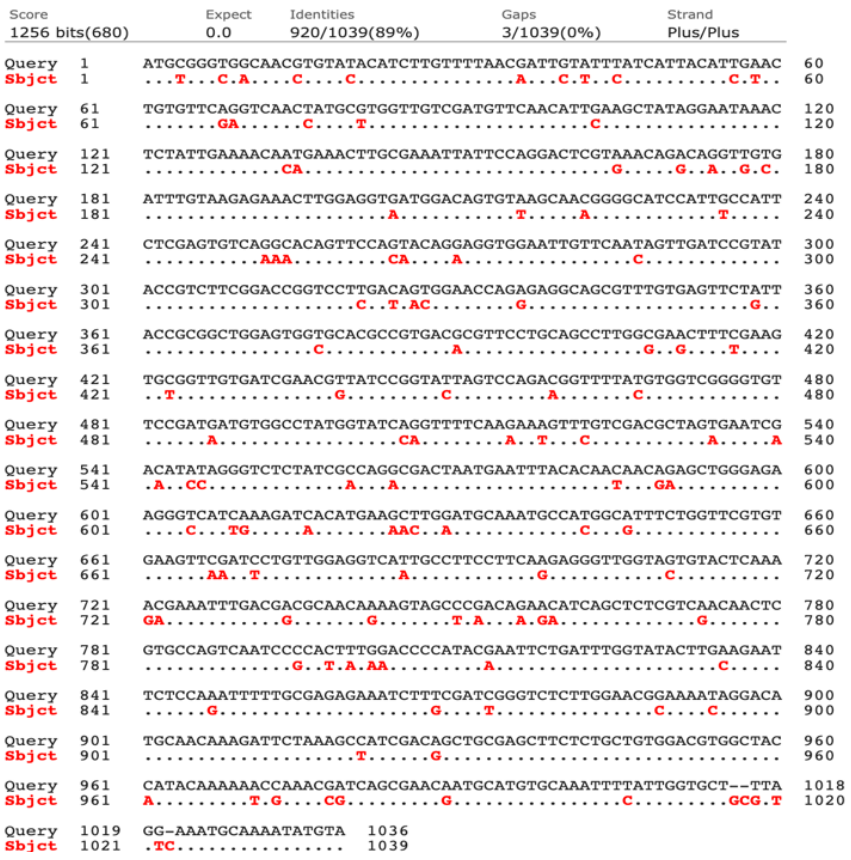


Figure 16. Structure of Annotated Wnt Genes in *H. glaberrima***.** The structure for all *Wnt* genes was characterized utilizing *Wnt* transcripts identified in the sea cucumber transcriptome data. All the scaffolds that contain each *Wnt* exon are identified by labels on top of each bracket (e.g., Hglab.02944). The start and end coordinates of the *Wnt*s span across each scaffold are shown at the bottom of each gene structure based on the genome nucleotide sequences (NCBI ID: XXX). Broken lines between exons represent a gene structure

break due to genome fragmentation. Exon and intron sizes are based on base pair lengths, in which each base pair is equal to 0.2 for exons and 0.05 for introns.

A genomic analysis also allowed us to further confirm the duplication of *Wnt4* by determining the distinct distribution of exons in the genome (Figure 16) and by the alignment of sequences with an 89% percent degree of similarity (Figure 17). Thus, we confirmed such cases to be biological duplications, rather than sequencing or assembly artifacts.



Score	Expect	Identities	Gaps	Strand
1256 bits(680)	0.0	920/1039(89%)	3/1039(0%)	Plus/Plus

```
Query  1     ATGCGGGTGGCAACGTGTATACATCTTGTTTTAACGATTGTATTTATCATTACATTGAAC  60
Sbjct  1     ...T...C.A....C....C..............A...C.T..C..........C.T..  60

Query  61    TGTGTTCAGGTCAACTATGCGTGGTTGTCGATGTTCAACATTGAAGCTATAGGAATAAAC  120
Sbjct  61    .......GA......C....T.....................C.................  120

Query  121   TCTATTGAAAACAATGAAACTTGCGAAATTATTCCAGGACTCGTAAACAGACAGGTTGTG  180
Sbjct  121   ..............CA.............................G.....G..A..G.C.  180

Query  181   ATTTGTAAGAGAAACTTGGAGGTGATGGACAGTGTAAGCAACGGGGCATCCATTGCCATT  240
Sbjct  181   ........................A...........T.....A...........T.....  240

Query  241   CTCGAGTGTCAGGCACAGTTCCAGTACAGGAGGTGGAATTGTTCAATAGTTGATCCGTAT  300
Sbjct  241   ...........AAA.........CA....A................C.............  300

Query  301   ACCGTCTTCGGACCGGTCCTTGACAGTGGAACCAGAGAGGCAGCGTTTGTGAGTTCTATT  360
Sbjct  301   ....................C..T.AC.........G.....................G..  360

Query  361   ACCGCGGCTGGAGTGGTGCACGCCGTGACGCGTTCCTGCAGCCTTGGCGAACTTTCGAAG  420
Sbjct  361   ................C............A..................G..G....T....  420

Query  421   TGCGGTTGTGATCGAACGTTATCCGGTATTAGTCCAGACGGTTTTATGTGGTCGGGGTGT  480
Sbjct  421   ..T..............G.........C.........A.......C.............  480

Query  481   TCCGATGATGTGGCCTATGGTATCAGGTTTTCAAGAAAGTTTGTCGACGCTAGTGAATCG  540
Sbjct  481   ......A................CA........A..T...C...........A.....A  540

Query  541   ACATATAGGGTCTCTATCGCCAGGCGACTAATGAATTTACACAACAACAGAGCTGGGAGA  600
Sbjct  541   .A..CC.............A...A.....................T...GA..........  600

Query  601   AGGGTCATCAAAGATCACATGAAGCTTGGATGCAAATGCCATGGCATTTCTGGTTCGTGT  660
Sbjct  601   ....C...TG.....A.......AAC..A............C...G..............  660

Query  661   GAAGTTCGATCCTGTTGGAGGTCATTGCCTTCCTTCAAGAGGGTTGGTAGTGTACTCAAA  720
Sbjct  661   ......AA..T.............A............G...........C..........  720

Query  721   ACGAAATTTGACGACGCAACAAAAGTAGCCCGACAGAACATCAGCTCTCGTCAACAACTC  780
Sbjct  721   GA...........G.......G.......T.A...A.GA.............G.......  780

Query  781   GTGCCAGTCAATCCCCACTTTGGACCCCATACGAATTCTGATTTGGTATACTTGAAGAAT  840
Sbjct  781   ..............G..T.A.AA.........A.....................C.....  840

Query  841   TCTCCAAATTTTTGCGAGAGAAATCTTTCGATCGGGTCTCTTGGAACGGAAAATAGGACA  900
Sbjct  841   ......G.....................G....T..............C....C......  900

Query  901   TGCAACAAAGATTCTAAAGCCATCGACAGCTGCGAGCTTCTCTGCTGTGGACGTGGCTAC  960
Sbjct  901   ....................T......G.................................  960

Query  961   CATACAAAAAACCAAACGATCAGCGAACAATGCATGTGCAAATTTTATTGGTGCT--TTA  1018
Sbjct  961   A.........T.G....CG.........G................C.........GCG.T  1020

Query  1019  GG-AAATGCAAAATATGTA  1036
Sbjct  1021  .TC.................  1039
```

Figure 17. NCBI Blast Alignment of Wnt4b against Wnt4a. Both genes were found in the transcriptome. To ensure that the transcripts were distinct and not a sequencing error, a sequence alignment was made. The query is *Wnt4b* and the subject is *Wnt4a*. Both genes were of similar length and displayed an 89% identity with only 3 gaps in the sequence alignment.

It is important to highlight that (i) all *Wnt* genes that were found in the transcriptome database were also found in the draft genome, and (ii) that no new *Wnt* genes were found in the draft genome that were not present in the transcriptome database. This was true for *Wnt8* and *Wnt11*, which were not detected in the transcriptome or the genome, even after attempting to find a homologous sequence by using multiple sequences from other echinoderms.

Comparison of Wnt Genes in Echinodermata

The protein coding sequences of the *Wnt* genes annotated in the *H. glaberrima* genome were compared to those found in NCBI or EchinoBase for species from other classes of the Echinodermata clade. Only sequences from three classes were reported: Asteroidea (sea stars), Echinoidea (sea urchins and sand dollars), and Holothuroidea (sea cucumbers) (Table 13). Therefore, *Wnt* sequences for the Crinoidea (sea lilies) and Ophiuroidea (brittle stars) classes were not included in the cross-species comparisons.

Table 13. NCBI Accession of Sequences of Distinct Echinoderms

Species	Wnt	NCBIAccession	Comments
L. variegatus	Wnt1	XP_041475650.1	
	Wnt3	XP_041478374.1	
	Wnt4	XP_041479642.1	
	Wnt5b	XP_041460654.1	
	Wnt6	XP_041475661.1	
	Wnt7b	XP_041459365.1	
	Wnt8a	XP_041485541.1	
	Wnt9a	XP_041475638.1	
	Wnt10b	XP_041477044.1	
	Wnt16	XP_041457833.1	
	WntA	XP_041464773.1	Appears as Wnt1-like
P. miniata	Wnt1	XP_038047182.1	
	Wnt2	XP_038054103.1	
	Wnt3	XP_038073218.1	
	Wnt4	XP_038061751.1	
	Wnt5b	XP_038054542.1	
	Wnt6	XP_038047180.1	
	Wnt7b	XP_038069369.1	Appears as an isoform of Wnt7
	Wnt8	XP_038059603.1	
	Wnt9	XP_038047188.1	
	Wnt10b	XP_038047890.1	
	Wnt11	XP_038051134.1	
	Wnt16	XP_038054431.1	
	WntA	XP_038076029.1	
S. purpuratus	Wnt1	NP_001116972.1	
	Wnt3	XP_030830597.1	
	Wnt4	XP_030842454.1	
	Wnt5b	XP_011670176.1	
	Wnt6	XP_790077.1	
	Wnt7b	XP_787051.3	
	Wnt8	NP_999832.1	
	Wnt9	XP_030830874.1	
	Wnt10	XP_011664244.1	
	Wnt16	XP_796616.2	
	WntA	XP_030832838.1	Appears as Wnt7a
A. japonicus	Wnt2	PIK43830.1	
	Wnt3	PIK62708.1	
	Wnt4	PIK52961.1	
	Wnt5	PIK40288.1	
	Wnt6	AGA62464.1	
	Wnt7A	PIK56278.1	
	Wnt7B	PIK48022.1	
	Wnt8A	PIK51024.1	
	Wnt8B	PIK51023.1	

	Wnt9	PIK51469.1
	WntA	PIK57158.1
E. fraudatrix	Wnt2	QEF51147.1
	Wnt3	QEF51148.1
	Wnt5	QEF51149.1
	Wnt7	QEF51150.1
	Wnt8	QEF51151.1
	Wnt9	QEF51152.1
	Wnt10	QDW65356.1
	Wnt16	ALT56982.1
	WntA	QDW65349.1

The phylogenetic analysis was performed using *Wnt* sequences from sea cucumbers *A. japonicus* and *E. fraudatrix (Holothuroidea),* sea urchins *L. variegatus* and *S. purpuratus* (Echinoidea), and the sea star *Patiria miniata* (Asteroidea). The results showed that all the sequences were in the same clade of their proposed *Wnt* homologs, confirming the discovery and annotation of the *Wnt* genes mentioned above (Figure 18). Furthermore, it demonstrated a high conservation of these within the major classes of the Echinodermata clade, obtaining bootstrap values of more than 95 for each specific *Wnt* gene type (Figure 18, nodes with pink circles). Similarly, for those *Wnt*s that are present in two or more holothurian species, these sequences were grouped in sister clades to those of echinoid species.

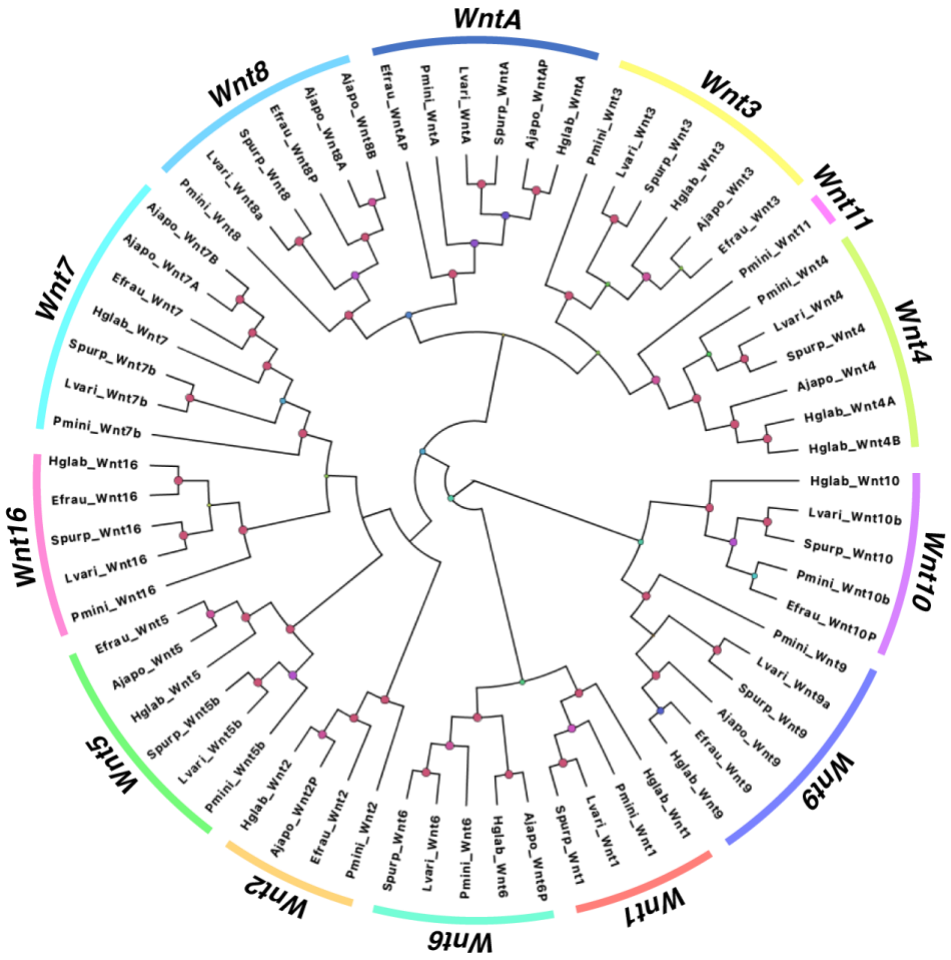


Figure 18. Phylogenetic Analysis of Wnt from Distinct Echinoderm Species. Gene tree of all *Wnt* genes identified in *H. glaberrima* (Hglab), along with those deposited in NCBI or Echinobase for *S. purpuratus* (Spur), *L. variegatus* (Lvari), *P. miniata* (Pmini), *A. japonicus* (Ajapo), and *E. fraudatrix* (Efrau). For all partial sequences, a letter P was added to its label; all other letters are based on the gene name shown in NCBI. The size and color of circles in nodes are dependent on their bootstrap value. Nodes with bootstrap values over 95 are shown in pink.

As it is known that *Wnt* genes are conserved in clusters, we wanted to know if clusters persist in *H. glaberrima* (Figure 19). Initially, we searched for the genomic coordinates of the *Wnt* genes within the genomes of *L. variegatus* and *S. purpuratus* for the sake of comparison and found *Wnt9, Wnt3, Wnt1, Wnt6, and Wnt10* clustered together. Based on the fragmentation of the currently available *H. glaberrima* draft genome, we could only confirm the cluster conservation of *Wnt1, Wnt6,* and *Wnt10.* It appears that the cluster is not only conserved but also oriented in the same manner as that of *L. variegatus.* However, the cluster in *S. purpuratus* shows a different orientation for *Wnt1* and *Wnt6* compared to both *L. variegatus* and *H. glaberrima.* Similarly, when the same cluster is compared with that in *Drosophila,* the structure and order are maintained, albeit with a change in *Wnt6* orientation.

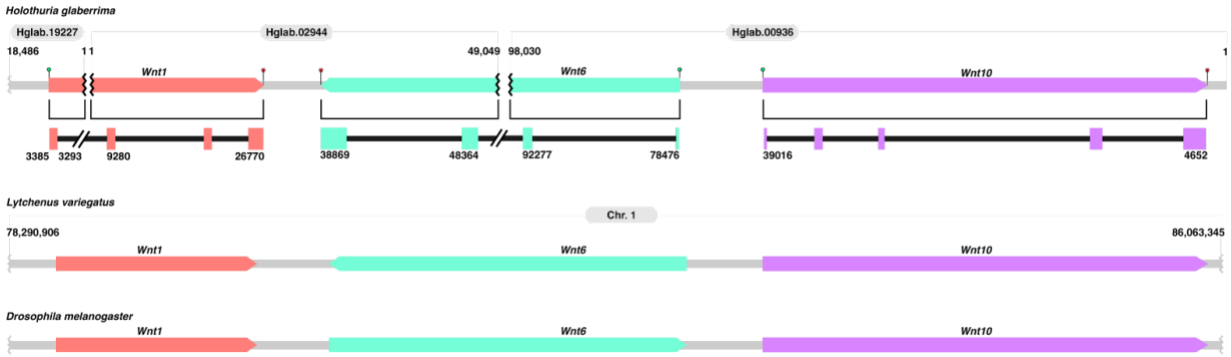


Figure 19. Wnt Cluster Conservation. Schematic illustrating the conservation of the *Wnt1-Wnt6-Wnt10* cluster in *H. glaberrima* compared to *L. variegatus* and *Drosophila melanogaster* (adapted from Vosburg et al., 2017). For *H. glaberrima*, we show the distribution of these genes across the draft assembly by complete squares with arrow heads pointing in the direction of the gene in the scaffold. The coordinates on top of the gene structures show scaffold size. The IDs on top of each gene structure represent the scaffold ID (NCBI ID: GCA_009936505.2). Separations of each *Wnt* due to fragmentation are shown as black line breaks. Start and stop codons are represented by green circles and red stars, respectively. The cluster of *L. variegatus* was characterized using its latest genome (Lvar_3.0; NCBI ID: 3495).

Wnt Expression

To determine the differential expression of *Wnt* genes during intestinal regeneration, we performed differential gene expression (DGE) analyses using various RNA-seq timepoints. These included previously published data from normal mesentery and from 1- and 3-dpe rudiments Quispe-Parra et al., 2021), as well as unpublished transcriptomes that include normal uneviscerated large intestine, 12 hours post-evisceration (hpe), 3-, 7-, 14- and 21-dpe. At 14-dpe, the regenerating rudiment is in the process of forming the lumen. Therefore, tissues from 14-dpe animals were used to make three different transcriptomes, corresponding to 14-dpe anterior (with lumen), 14-dpe rudiment (middle, no lumen), and 14-dpe posterior (with lumen). Since the transcriptomes correspond to different batches, some timepoints were repeated to be able to assess and eliminate possible batch effects. With all these timepoints, we performed a principal component analysis (PCA) to determine the relationship of gene expression at different regeneration stages (Figure 7 & S1). The PCA results showed two distinct groups that represent the normal mesentery and normal intestine. The PCA results for regenerating samples displayed a major difference between early-stage regenerates (12-hpe, 1- and 3-dpe) and late-stage regenerates (14- and 21-dpe) with the 7-dpe samples separating the two groups.

The differential expression of *Wnt* genes during intestinal regeneration was analyzed using all RNA-seq timepoints mentioned. Therefore, two different controls were used for these analyses. Regenerating intestinal stages that lacked a lumen (and, therefore, a luminal epithelium) were compared to the normal mesentery, while regenerating intestines that had a lumen (and, therefore, a luminal epithelium) at 14-dpe anterior and posterior and 21-dpe were compared to normal intestine.

Wnt3, *Wnt6,* and *Wnt9* were found to be upregulated both in early- and late-stage regeneration. However, only advanced regenerative intestines that had luminal epithelium showed differential expression of *Wnt4a*, *Wnt4b*, *Wnt5,* and *WntA*. Only one Wnt, *Wnt7*, showed a decrease in expression, specifically, at early-stage regenerating intestine (12-hpe) when compared to the normal mesentery and in regenerating intestines at advanced stages when compared to normal intestine. An interesting finding was the difference in *Wnt* gene expression between 14-dpe anterior and 14-dpe posterior intestine. In the anterior intestine with luminal epithelium, *Wnt6* was uniquely upregulated, while in posterior tissue, *Wnt5* and *Wnt10* were uniquely upregulated (Figure 20).

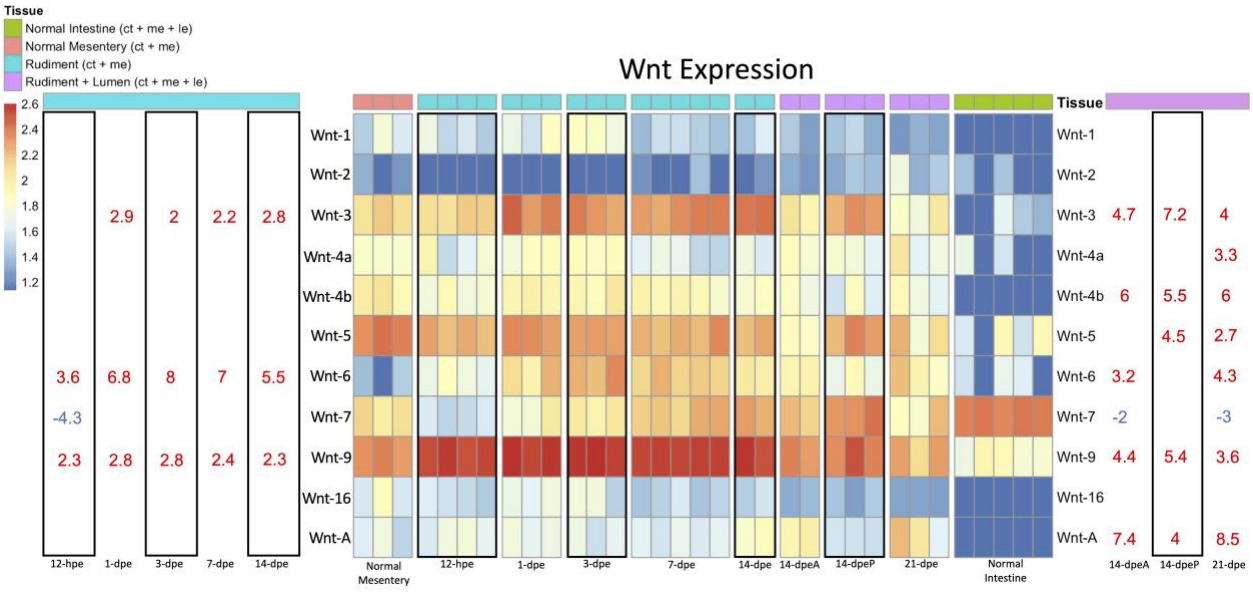


Figure 20. Heatmap of Wnt Expression during Early- and Late-stage Regeneration. This heatmap contains all the RNA-seq timepoints collected from the transcriptomic databases. The timepoints 12-hpe through 14-dpe should be referenced to Normal Mesentery, as this was the control. Similarly, the timepoints 14-dpeA, 14-dpeP, and 21-dpe should be referenced to Normal Intestine. On either side of the heatmap are the Log2foldchange (L2FC) values of a *Wnt* gene at a given timepoint. A significance threshold was set at L2FC <-2 or >2 with a pADJ value of 0.001. Above the columns are color coded labels that represent the tissue composition of the samples at a given timepoint. Abbreviations—CT: connective tissue, ME: mesothelium, and LE: luminal epithelium.

Expression of Wnt-Associated Genes

The availability of differential gene expression data from different regeneration stages and tissues offers the additional possibility of exploring the expression profile of other genes related to Wnt signaling pathways. Thus, we explored the expression of the members of the Wnt receptor family Frizzled (Fzd) and the Wnt signaling pathway protein Dishevelled (Dvl) (Figure 21). These results showed that while *Dvl-3* was expressed in the intestinal transcriptomes, there was no clear differential expression among the different stages or tissues.

In contrast, three out of the five *Fzds* that were found in the intestinal transcriptomes showed differential expression. *Fzd1* and *Fzd10* were upregulated in the late regeneration stages, i.e., posteriorly *Fzd10* at 14-dpe while anteriorly *Fzd1* at 21-dpe. *Fzd4* was down-regulated during early-stage regeneration in the rudiment but was then up-regulated in late regenerating intestines when compared to the normal intestine.

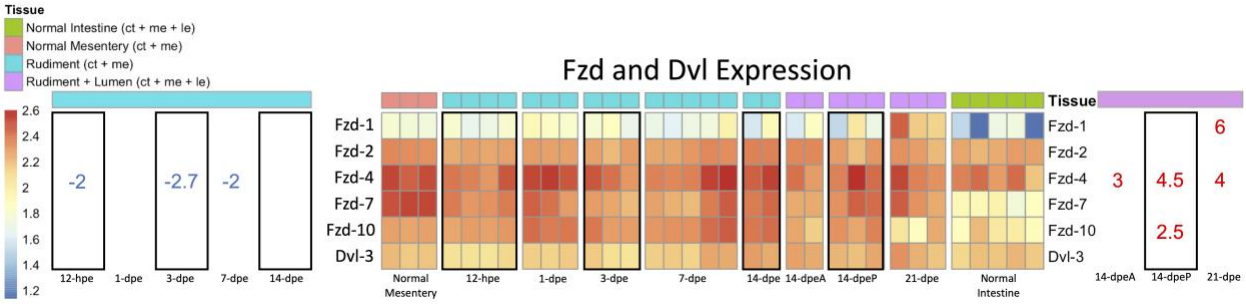


Figure 21. Heatmap of Fzd and Dvl Expression during Early- and Late-stage Regeneration. This figure can be read the same as the heatmap from Figure 20. The significance threshold is the same.

Wnt/β-Catenin and Wnt/PCP Pathway Gene Expression

Two major branches of Wnt signaling pathways exist, i.e., the canonical and the non-canonical Wnt signaling pathways, which can mediate different functions in regenerative processes (Komiya & Habas, 2008). A brief overview of both pathways can be found in Figure 15. The differential gene expression data were also used to explore the possibility that a particular pathway could be associated with an individual regeneration stage or tissue. For this, we searched for the transcripts of genes associated with either the canonical (Wnt/β-catenin) pathway or non-canonical (Wnt/PCP) pathway in the transcriptomic database and then analyzed these for differentially expressed genes. First, we assessed genes in the β-catenin pathway (Figure 22). In the early regenerative stages, two genes, *Groucho* (a repressor of TCF/LEF) and *Kremen1* (an inhibitor of the Wnt/Fzd/LRP6 complex), were down-regulated at 12-hpe to 3-dpe when compared to the normal mesentery. *DKK3* was upregulated only in the 12-hpe stage following evisceration, while *Myc* (a target gene) was upregulated during early-stage regeneration from 12-hpe to 3-dpe. Conversely, in the late regenerative stages, *Kremen1* and *Axin2* (a target gene and self-inhibitor) were upregulated when compared to the normal intestine. Additional genes associated with the canonical pathway that were found to be upregulated in some late-stage regenerates were *Twist* and *Slug* (specifically in the anterior 14-dpe stage) and *EDNRA*, *SP5*, and *BAMBI* in anterior and posterior 14-dpe, as well as in the 21-dpe.

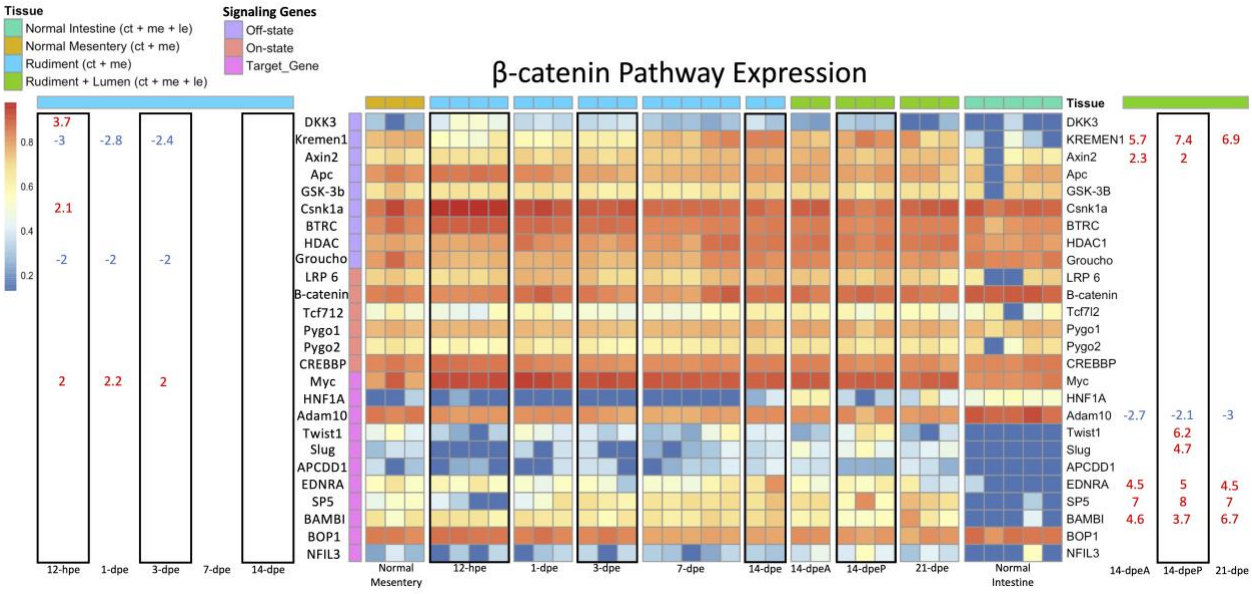


Figure 22. Heatmap of the Signaling Genes in the Wnt/β-catenin Pathway. This figure is read the same as in Figure 20. The significance threshold remains the same. The only difference is that the rows are color coded to indicate the role of the gene in the Wnt/B-catenin pathway. Genes in the Off-state inhibit β-catenin signaling, while genes in the On-state facilitate Wnt/β-catenin signaling. The Target Genes are the downstream targets of the TCF/LEF transcription factor family that is turned on by β-catenin.

As for the Wnt/PCP pathway, several associated genes were downregulated during the early regenerative stages, including *Lamc1* and *Ror1* at 12-hpe and *c-Jun* and *Ctnnal1* from 12 hpe to 3-dpe (Figure 23). Two genes were over-expressed at the same time periods, i.e., *Mfhas1* at 12-hpe and *Rac1* from 12-hpe to 3-dpe. In the later stages of regeneration, we found several Wnt/PCP associated genes to be over-expressed, including the core proteins *Ankrd6* and *Vangl2* and other genes such as *Pvr*, *Mfhas1*, *Lamc1*, *Daam2*, and *Ror1*. Only one Wnt/PCP associated gene, *Map1ic3b*, was found to be down-regulated in the late regenerative stages (14-dpe anterior and posterior) when compared to the normal intestine.

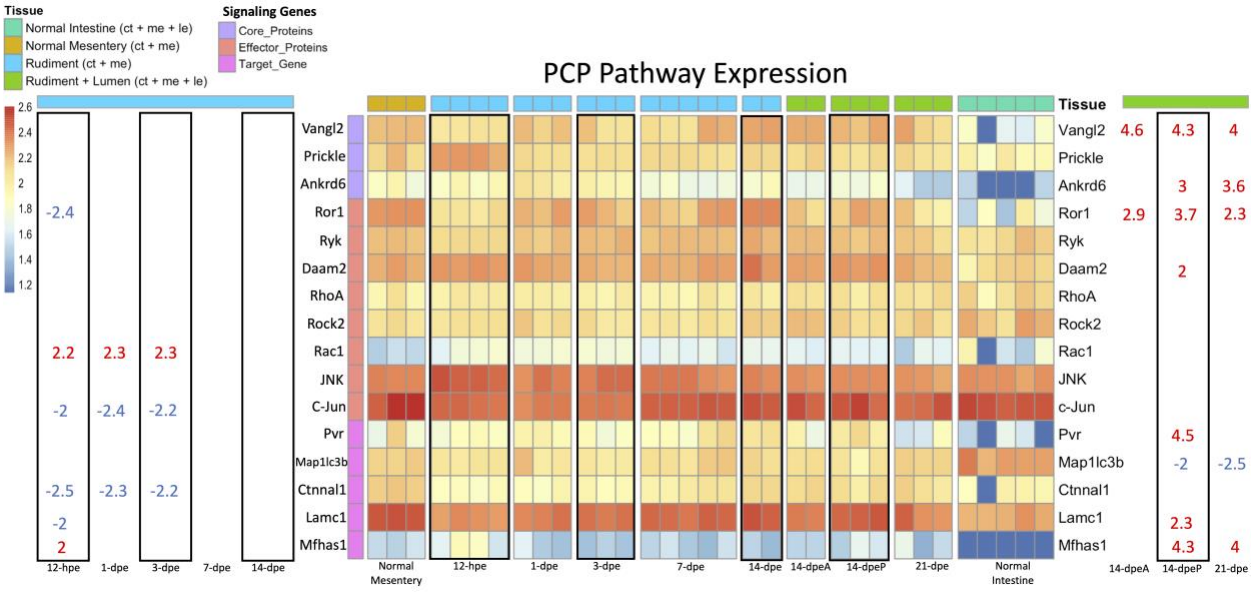


Figure 23. Heatmap of the Signaling Genes in the Wnt/PCP Pathway. This figure is read the same as in Figure 20. The significance threshold remains the same. The only difference is that the rows are color coded to indicate the role of the gene in the Wnt/PCP pathway. Genes labeled as Core Proteins are unique and essential to the Wnt/PCP pathway and, therefore, are strong indicators of its presence. The effector proteins are genes involved in carrying out the signaling cascade, while target genes are activated by the transcription factor c-Jun.

Discussion

Comparison of Wnt Genes in Echinodermata

In the present study, we found 12 *Wnt* genes in the genome of *H. glaberrima*: *Wnt1, Wnt2, Wnt3, Wnt4a, Wnt4b, Wnt5, Wnt6, Wnt7, Wnt9, Wnt10, Wnt16, and WntA.* Surprisingly, all these genes were shown to be expressed in the intestinal tissue, where they were initially identified, and no other *Wnt* gene was found in the genome. This is somewhat unexpected, as *Wnt genes* are involved in diverse developmental and biological processes; thus, we expected to find *Wnt* genes in the genome that were not expressed in the intestinal transcriptomes. Further, the manual annotation and characterization of these genes allowed us to determine their genomic structure within the sea cucumber genome (Figure 16). The structure of these in *H. glaberrima* is similar across the gene family, with all containing between 4 and 6 exons, with a minimum size of 59 bp (mostly first exons) and a maximum of 494 bp. All the annotated *Wnt* genes, with exception of *Wnt5*, had a first exon smaller than the rest, and the last exon was the largest of the studied *Wnts*. Different from the exon structures, intronic regions were variable across the family, showing a minimum of 636 bp and a maximum of 21,268 bp, with an average of 7363 bp. To the current knowledge, there is no information available about the structure of *Wnt* genes in other echinoderms, a few studies have described the structure of some of these in other species, such as the zebrafish *Brachydanio rerio* and humans (Molven et al., 1991; Miller, 2001). *Wnt1* in *B. rerio* and humans shows a similar structure to that of *H. glaberrima*, i.e.,

composed of four exons. Similarly, the structures of human *Wnt5* and *Wnt16,* both with four exons, are the same as those found in *H. glaberrima.* Yet, when comparing the genomic structure of *Wnt2, H. glaberrima* contains six exons while humans have five exons.

We extended the characterization of the *Wnt* genes by performing a phylogenetic analysis of *Wnt* of various echinoderms species (Figure 24). This was done for two main reasons: (i) to confirm that each characterized *Wnt* was arranged in the same clade as its respective homologs, and (ii) to assess the conservation of these genes within the Echinodermata phylum. The phylogenetic tree shows adequate grouping of *Wnt* with the expected clades; however, when looking deep into each of these, there were several interesting results. For instance, as one would expect, *Wnt* from holothuroids was shown to be in the same clade sister to that of echinoids. However, in the cases where partial sequences were utilized, the arrangement of *WntA* and *Wnt10* in *E. fraudatrix* was found to be different, resulting in an independent branch (*WntA*) or joining *P. miniata* as a sister clade to the echinoid clade (*Wnt10*). Other than this, which we attribute to the lack of a complete sequence, conservation was maintained, as all holothuroids and echinoid *Wnt* were maintained in distinct clades. Similarly, in almost all cases, *Wnt* genes of *P. miniata* appeared as a single branch throughout the tree.

Furthermore, there are variations between and within the assessed echinoderm classes regarding the loss or absence of individual *Wnt* genes (Figure 24). Seven subfamilies are conserved among the five classes: *Wnt3*, *Wnt5*, *Wnt7*, *Wnt9*, *Wnt10*, *Wnt16,* and *WntA*. In contrast, within the class Holothuroidea, eight subfamilies are conserved: *Wnt2*, *Wnt3*, *Wnt5*, *Wnt7*, *Wnt9 Wnt10*, *Wnt16,* and *WntA*. As can be seen in the phylogenetic tree, we did not include *Wnt10* and *Wnt16* of *A. japonicus,* as they were not deposited in NCBI. Nevertheless, we opted to include results from a previous publication (Yuan et al., 2019) that investigated *A. japonicus Wnt* genes, albeit with some limitations. This study suggested that there is a duplication of *Wnt3*. However, both sequences deposited in NCBI (PIK62708.1; PIK45647.1) were 100% identical. Furthermore, the authors of that study suggested a duplication of *Wnt9*, but only one sequence could be found deposited. Additionally, we found four sequences identified as *Wnt8* in NCBI, two of which seemed to be duplications (PIK51024.1; PIK51023.1) and another with homology to *Wnt2* (PIK43830.1) in several species when performing a BLAST against the NCBI non-redundant database. Therefore, we only included the sequences for which we were certain of their homology, only one *Wnt3* and *Wnt9* sequences, *Wnt8* duplicates, and the *Wnt8* sequence that showed homology to *Wnt2* of other species (labeled here as *Wnt2*). Similarly, we used the *Wnt1* sequence from A. japonicus reported by Girich and colleagues (2019) (Girich & Boyko, 2019), although this Wnt had not been previously reported in this species (Yuan et al., 2019). *Wnt8* was not found in *H. glaberrima.* Additionally, the *Wnt4* duplication appears to be unique to *H. glaberrima* compared to the species considered here. Lastly, the asteroid *P. miniata* is the only species that has all 13 *Wnt* subfamilies but no duplications.

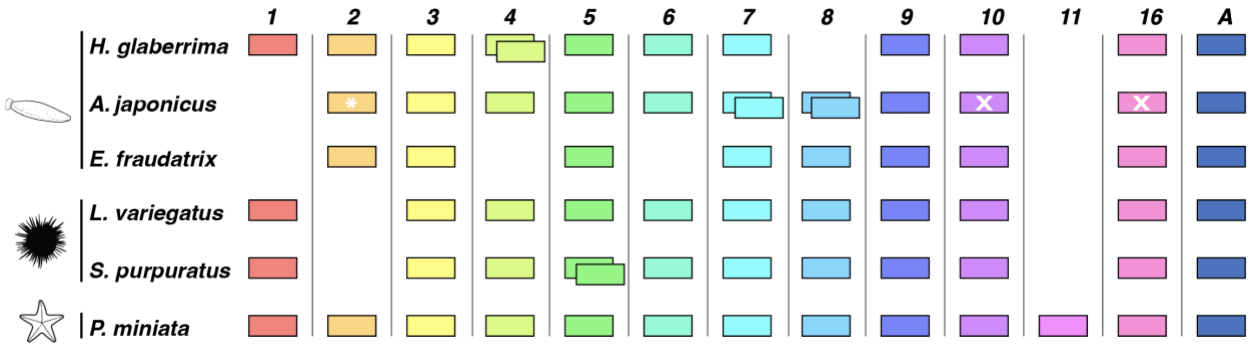


Figure 24. Summary of Wnt Gene Distribution in Select Echinoderms. This schematic is based on a phylogenetic analysis of Figure 18. Asterisks depict manually characterized genes. White X depicts genes that could not be found in the NCBI database.

Here, we also demonstrate the conservation of the *Wnt1, Wnt6, and Wnt10* cluster in *H. glaberrima.* This cluster is known to also contain *Wnt9* and *Wnt3* in several other species (Vosber et al., 2021). Nonetheless, the fragmented nature of the current draft genome limited the possibility of confirming the additional genes within the cluster. Still, we were able to manually confirm that in *H. glaberrima, Wnt1, Wnt6, and Wnt10* are clustered together (Figure 19). The annotation showed that four exons of *Wnt1* were present in scaffold Hglab.02944 in the positive direction, the same scaffold that contains two exons of *Wnt6* in the negative direction, separated by 12.09 Kb. Similarly, we found that the scaffold Hglab.00936 contained the full *Wnt10* gene in the positive direction, along with the two first exons of *Wnt6* in the negative direction, separated by 39.46 Kb. However, this scaffold appeared to be in the opposite direction to scaffold Hglab.02944, for which, in Figure 19, we rotated it to demonstrate the clustering of *Wnt1, Wnt6, and Wnt10.* Moreover, we believe this to be the reason why the assembler utilized for the draft genome was not able to merge these two scaffolds. We also identified the same gene organization and direction of this cluster in the latest genome of *L. variegatus* (Figure 19), which are also clustered with *Wnt3* and *Wnt9* (data not shown). All of this is consistent with clusters previously reported in other species, such as *Drosophila melanogaster* (Vosberg et al., 2021). However, the direction, and in some cases, the organization of the genes in such a cluster differs from species to species, while their grouping remains intact.

Wnt Signaling During Early-Stage Regeneration

Although correlative in nature, DGE analyses provide insight into which genes may participate in or modulate specific cellular processes. As for early-stage regeneration, which spans from 12-hpe to 7-dpe, *Wnt7* is downregulated at 12-hpe when injury and/or wound healing occurs. This contrasts with what was found in *A. japonicus*, where *Wnt7* was overexpressed during the injury/wound healing stage (Yuan et al., 2019). Upregulated genes during early-stage regeneration were *Wnt3*, *Wnt6* and *Wnt9*. These genes may be involved from 1- to 7-dpe when the mesothelium dedifferentiates and migrates to the mesentery, where the cells redifferentiate and proliferate to form the intestinal rudiment (Yuan et al., 2017).

To determine which Wnts may facilitate cellular processes during early-stage regeneration, we referenced previous work from this lab and others. First, it has been proposed that Wnt signaling is involved in proliferation (Bello et al., 2020), while a possible role of Wnts in apoptosis and/or dedifferentiation has been discarded (Alicea-Delgado et al., 2021; Bello et al., 2021). However, a specific Wnt has not been shown to be in control of cellular proliferation. Second, these data and other studies point to *Wnt6* as the most likely candidate to modulate the increase in proliferation since, in *H. glaberrima*, *E. fraudatrix*, and *A. japonicus, Wnt6* is up-regulated between 3- and 7-dpe. Lastly, this upregulation precedes the large spike in cell proliferation (~7-dpe) observed in the mesothelium of regenerating intestines (García-Arrarás et al., 2019; Zheng et al., 2018), while the overexpression of *Wnt*3 and *Wnt*9 is stable throughout early-stage regeneration. Further evidence comes from other species. For example, *Wnt6* expression is most abundant in the vertebrate intestine in the crypt epithelium, where proliferation mostly occurs (Gregorieff et al., 2005). *Wnt6* is also a known contributor to tumorigenesis and the development of colon cancer via its effects on cell proliferation, apoptosis, and migration (Benhaj et al., 2006).

We cannot rule out *Wnt3* and *Wnt9* as possible controllers of proliferation during early-stage regeneration or during the injury/wound healing phase, particularly since the Wnt9 protein is expressed in the regenerating mesothelium (Mashanov et al., 2012). Additionally, Wnts are known to have redundant functions (Matsura et al., 2016). Thus, *Wnt*3 and *Wnt*9 may supplement the effect of *Wnt*6. For example, *Wnt*3 is thought to time cell divisions in mammalian intestinal stem cells, where Wnt6 and Wnt9 are also expressed (Wu et al., 2012). Alternatively, they could have their own independent function, possibly playing a role in the epithelial–mesenchymal transition (EMT) (García-Arrarás et al., 2011). It is thus interesting that Wnt3 can promote EMT in breast cancer cells (Baulies et al., 2020) while in the sea cucumber, it is overexpressed beginning at 1-dpe, which is slightly earlier than the EMT timeline. Nonetheless, functional studies are required to be certain about the effects of these Wnts during early-stage regeneration.

Wnt Signaling During Late-Stage Regeneration

Late-stage regeneration is dominated by the formation of the lumen whereby at 14-dpe, luminal epithelial cells from the esophagus and cloaca proliferate and migrate into the intestinal rudiment (Zheng et al., 2018). At 21-dpe, proliferation continues, causing the intestine to elongate and grow in overall size. The *Wnt* genes most likely associated with these stages are *Wnt4b*, *Wnt5*, and *WntA*, all of which are uniquely overexpressed in late-stage but not in early-stage regeneration. The emergence of these *Wnts* could be a result of the mesenchymal cells that surround the luminal epithelium. This is seen in mammals where mesenchymal cells beneath the basal lamina of the luminal epithelium uniquely secrete Wnt4 and Wnt5 (Flanagan et al., 2015). However, the Wnts associated with the formation of the rudiment (Wnt3, Wnt6 and Wnt9) may also contribute to intestinal growth and elongation, since their upregulation persists during late-stage regeneration.

It is also interesting that several *Fzd* genes are differentially expressed during late-stage regeneration, namely *Fzd1*, *Fzd4* and *Fzd10*. The high expression of *Fzd* genes is not fully

unexpected since, in mammals, these receptors have been closely associated with the regeneration of the luminal epithelium and with stem cell modulation (Mashanov et al., 2010). Additionally, another interesting finding from this study is the different patterns of gene expression between the anterior and posterior intestine. A similar finding had already been documented in *H. glaberrima*, i.e., differences in the spatial or temporal expression of genes and of cell proliferation and apoptotic events (Kemp et al., 2007). Additional examples of spatial differences in Wnt expression can be found in other tissues and other species, mainly during embryonic development. For example, in mouse embryos, *Fzd10* is found in the most posterior region of the epiblast and in the primitive streak but not in mesoderm migrating laterally and anteriorly (Janssen et al., 2021). In some spiders, *Wnt6* expression is not present posteriorly during embryonic patterning (Andre et al., 2015). However, the most interesting of these is *Wnt5*, as it is expressed at the caudal end of the embryo during gastrulation and eventually in the distal portion of structures that extend from the primary body axis (Sato et al., 2011).

The overexpression of *Wnt*A during late-stage regeneration is also of particular interest. This *Wnt* is not found in vertebrates and is thought to have been lost during evolution. However, it is present in echinoderms, and its upregulation during late-stage intestinal regeneration has been confirmed in *A. japonicus* (Li et al., 2017). It is thus easy to speculate that *Wnt*A overexpression might be associated with the amazing regenerative abilities of echinoderms.

Wnt Signaling Pathways

We have also used DGE analyses to elucidate the dynamics of Wnt signaling during holothurian intestinal regeneration. As mentioned, Wnt signaling can occur via the canonical Wnt/β-catenin pathway and the noncanonical Wnt/PCP pathway. These data suggest that the cellular proliferation observed in early-stage regeneration occurs via the Wnt/β-catenin pathway, while luminal epithelial events in late-stage regeneration occur via the Wnt/PCP pathway. This proposition is supported by comparative data, e.g., the Wnts that are upregulated during early-stage regeneration have also been shown to be highly expressed in vertebrate paneth cells (Farin et al., 2012; Gregorieff et al., 2005; Van Es et al., 2005; Krawetz & Kelly, 2008) and are known to activate the Wnt/β-catenin pathway (Sato et al., 2011; Jin et al., 2020; Dush & Nascone-Yoder, 2019). Additional evidence comes from the downregulation of *Kremen* and *Groucho* in the early-stage regeneration. Kremen normally prevents Wnt from binding to Fzd, thus disabling the signaling cascade, while Groucho acts as a repressor of the LEF/TCF. Other evidence comes from this lab, where pharmacological experiments targeting the Wnt/β-catenin pathway both in vivo and in vitro modulate cell proliferation in early intestinal regenerate (Bello et al., 2020). However, the most convincing evidence is the fact that RNAi for β-catenin decreases cellular proliferation in the early regenerating intestines in vitro (Alicea-Delgado et al., 2021).

As for late-stage regeneration, the Wnt/PCP pathway is most likely involved in the regeneration of the luminal epithelium. This is evidenced by a few observations: first, the upregulation of *Vangl*, which is known to facilitate gut elongation and lumen formation (Dush et al., 2019);second, the upregulation of *Wnt5*, which can operate through the Wnt/PCP pathway to elongate the posterior gut, particularly when taking into account that, in mice, Wnt5

has been shown to work together with Vangl to orient cell division along the rostrocaudal axis to increase fore-stomach length (Matsuyama et al., 2009); third, *Fzd4*, which is associated with the Wnt/PCP pathway, was highly upregulated in late-stage regeneration, while it was under-expressed in early-stage regeneration; finally, the upregulation of *Kremen1*, which is an inhibitor of the Wnt/β-catenin pathway, suggests crosstalk between Wnt signaling pathways. This would also suggest that Wnt/PCP regulates Wnt/β-catenin during late-stage regeneration, so that it can impose its effects on cellular processes.

We recognize that most DGE-based discussions are purely correlational, and further experimentation is necessary to conclude that specific genes have or lack a specified function. Nonetheless, our analyses can be viewed as a series of hypotheses that can be tested in future investigations.

In conclusion, the work presented here comprises an extensive analysis of sequence information ranging from the characterization of the *Wnt* genes found in the genome of our model organism, *H. glaberrima*, to a comparative expression analysis in normal and regenerating intestines. The results provide important insights into the molecular bases of intestinal regeneration and the role that Wnt signaling mechanisms might be playing in the process. They also provide investigators with a point of departure for molecular analyses, not only of regenerative organogenesis, but also of molecular evolution, signaling pathways, and organ homeostasis.

Materials and Methods

Differential Gene Expression (DGE)

To assess DGEs of the candidate genes, we utilized previously available data from 1- and 3-dpe timepoints deposited at NCBI (ID: PRJNA660762) (Quispe-Parra et al., 2021), along with additional sequenced samples for stages 12-hpe, 3-, 7-, 14, and 21-dpe and normal intestine, as mentioned before. For the differential expression analysis, 14-dpe samples were separated based on the three isolated regions: the anterior and posterior portions where the lumen had already formed and the middle section where no lumen was present. Moreover, distinct tissue samples were utilized as controls, depending on the stage. For instance, samples where no luminal epithelium was present were compared to the mesentery tissue of non-eviscerated sea cucumbers, whereas those with luminal epithelium were compared to intestinal tissue of non-eviscerated sea cucumber.

DGE was performed based on previously reported protocols (Quispe-Parra et al., 2021). Reads were initially trimmed using Trimmomatic v.0.39 (Bolger et al. 2014) (ILLUMINACLIP:{}:2:40:15 LEADING:20 TRAILING:20 SLIDINGWINDOW:4:15 MINLEN:35) and then quantified using Salmon v.0.8.2 (Patro et al., 2017) with default parameters and the dumpEq flag that produces equivalence classes. Since the focus of this study was not to analyze the complete expression profile of all the stages, but rather, to determine the expression of candidate genes, we utilized previously assembled transcriptome deposited at NCBI (ID: GIVL00000000.1). Equivalence classes generated with Salmon were then passed through Corset v.1.09 (Davidson & Oshlack, 2014) with default parameters to hierarchically cluster contigs by

sequence similarity and expression class. Then, DESeq2 (Love et al., 2014) was used to perform differential expression analyses based on Corset clusters from the read counts generated. Samples with less than 30 read counts were filtered from the biological replicates of each sample. To identify the candidate genes on the transcriptome, we used the characterized *Wnt* genes sequences and annotated the transcriptome using UniProt database and predictions from *S. purpuratus* (NCBI ID: GCF_000002235.5). The complete script of the differential expression analysis can be found in the following GitHub: https://github.com/devneurolab/HgWnt2023.

Wnt Characterization and Manual Annotation

Intestine transcriptomic data were first utilized for mapping all the potential Wnt sequences using amino acid sequences from *Strongylocentrotus purpuratus* and *Eupentacta fraudatrix,* obtained from EchinoBase and NCBI. Sequences from these species were utilized as queries against the sea cucumber peptide transcriptomic database, available at NCBI (BioProject: PRJNA660762) (Quispe-Parra et al., 2021). The ORF of the obtained sequences were further characterized by BLASTp against the NCBI RefSeq database. The resulting *Wnt* transcripts were then used to assess if further genes from the *Wnt* family were present in the genome. Additionally, these were manually annotated in the draft genome (Medina-Feliciano et al., 2021) of the sea cucumber, available at NCBI (ID: GCA_009936505.2), to confirm that all were distinct genes. For the manual annotation, the sequences obtained from the transcriptome were mapped against the draft genome using BLASTn and tBLASTn for nucleotide and amino acid sequences, respectively. We used both BLAST results to manually identify the exons of all the *Wnt* genes identified. Based on the coordinates obtained, we extracted the sequences for each exon of all the *Wnt* genes using the getFasta feature of BedTools v.2.30.0. Exons of each Wnt were then joined together using a custom Perl script described in a previous study (Medina-Felician et al., 2021).

Phylogenetic Analysis of Wnt Genes

The *Wnt* gene sequences obtained from the manual genome annotation were used to perform a phylogenetic analysis. Given that this annotation was done utilizing a draft genome assembly, we wanted to confirm that our characterization has been done properly and that each *Wnt* was grouped with those of other echinoderm species. This would also allow us to assess the conservation of the *Wnt* gene family with other echinoderms. Here, we followed protocols previously employed (Medina-Feliciano et al., 2021), starting by using MAFFT v.7.312 (Katoh & Standley, 2013) to perform a multiple sequence alignment of *Wnt* genes from *S. purpuratus, Patiria miniata, Lytechinus variegatus, Apostichopus japonicus,* and *E. fraudatrix.* All of these were obtained from NCBI or EchinoBase (Table 9). Aligned sequences were utilized to create a phylogenetic tree by comparative analysis using IQTree v.2.0.3 (Nguyen et al., 2015) and RAXmL v.8.2.12 (Stamatakis, 2014). The complete methods and code for our phylogenetic analysis can be found in the Github repository of this project: https://github.com/devneurolab/HgWnt2023.

Chapter 4: General Conclusions

Significance of Thesis

The combined results of the presented studies have broad implications not only in regenerative biology and medicine, but also local impactions within our lab. To the best of our knowledge this is the first time that ribosome biogenesis has been implicated in intestinal regeneration. This study also reinforces the notion that the regenerating rudiment and luminal epithelium displays different regenerative molecular mechanisms, although undergoing similar cellular processes such as differentiation. Similarly, anterior and posterior regenerates were shown to display differences in gene expression and potential differential activation of highly conserved signaling pathways. Together these results advance our understanding of the molecular mechanisms of regeneration. Furthermore, the collected RNA-seq from our experiments have extended the transcriptomic profile of intestinal regeneration by providing data of both the regenerating rudiment and luminal epithelium. This data has now been compiled and is openly available on our lab website: https://blastkit.hpcf.upr.edu/hglaberrima-v1/. Future investigators may now use this transcriptomic data to plan experiments in a hypothesis-driven approach.

Future Directions

The scope of this research can be expanded in several ways. First, further studies are warranted on the molecular pathways involved in regeneration from this study. For example, ribosome biogenesis could be functionally explored by using chemical inhibitors to inject into regenerating sea cucumbers then perform assays on proliferation and differentiation and using BrdU and rhodamine-labelled phalloidin, respectively. Protocols have already been established in the lab for these assays (Mashanov & García-Arrarás, 2011; Miguel-Ruiz & García-Arrarás, 2007). Or studies on Wnt signaling pathway can be conducted such as selecting a specific Wnt and performing an RNAi experiment on explants to then perform downstream assays. The protocol for this has already been established. This could also be done with a vivo RNAi protocol, which other labs studying echinoderm intestinal regeneration have claimed to do. Moreover, if future analyses were confined to the computational biology and bioinformatic realm, then perhaps an ATAC-Seq experiment would be in order. The timepoints fort his experiment would align with the RNAs-seq data so that comparisons could be examined. This way gene expression could be compared with chromatin accessibility. However, this would be reliant on having a well-annotated and well-constructed genome, which is currently underway in our lab.

References

Alicea-Delgado, M., & García-Arrarás, J. E. (2021). Wnt/β-catenin signaling pathway regulates cell proliferation but not muscle dedifferentiation nor apoptosis during sea cucumber intestinal regeneration. *Developmental biology*, *480*, 105-113.

Amano, M., Nakayama, M., & Kaibuchi, K. (2010). Rho-kinase/ROCK: a key regulator of the cytoskeleton and cell polarity. *Cytoskeleton*, *67*(9), 545-554.

Andre, P., Song, H., Kim, W., Kispert, A., & Yang, Y. (2015). Wnt5a and Wnt11 regulate mammalian anterior-posterior axis elongation. *Development*, *142*(8), 1516-1527.

Andrews, S. (2010). FastQC: a quality control tool for high throughput sequence data.

Atchison, F. W., Capel, B., & Means, A. R. (2003). Pin1 regulates the timing of mammalian primordial germ cell proliferation. *Development (Cambridge, England)*, *130*(15), 3579–3586.

Bahrami, S., & Drabløs, F. (2016). Gene regulation in the immediate-early response process. *Advances in biological regulation*, *62*, 37-49.

Baulies, A., Angelis, N., & Li, V. S. (2020). Hallmarks of intestinal stem cells. *Development*, *147*(15), dev182675.

Beck F. (2002). Homeobox genes in gut development. *Gut*, *51*(3), 450–454. https://doi.org/10.1136/gut.51.3.450

Bello, S. A., Torres-Gutiérrez, V., Rodríguez-Flores, E. J., Toledo-Román, E. J., Rodríguez, N., Díaz-Díaz, L. M., ... & García-Arrarás, J. E. (2020). Insights into intestinal regeneration signaling mechanisms. *Developmental biology*, *458*(1), 12-31.

Benhaj, K., Akcali, K. C., & Ozturk, M. (2006). Redundant expression of canonical Wnt ligands in human breast cancer cell lines. *Oncology reports*, *15*(3), 701-707.

Bitar, K. N., & Raghavan, S. (2012). Intestinal tissue engineering: current concepts and future vision of regenerative medicine in the gut. *Neurogastroenterology and motility : the official journal of the European Gastrointestinal Motility Society*, *24*(1), 7–19. https://doi.org/10.1111/j.1365-2982.2011.01843.x

Blanco, G., Coulton, G. R., Biggin, A., Grainge, C., Moss, J., Barrett, M., ... & Brown, S. D. (2001). The kyphoscoliosis (ky) mouse is deficient in hypertrophic responses and is caused by a mutation in a novel muscle-specific protein. *Human molecular genetics*, *10*(1), 9-16.

Boeckmann, B., Bairoch, A., Apweiler, R., Blatter, M. C., Estreicher, A., Gasteiger, E., ... & Schneider, M. (2003). The SWISS-PROT protein knowledgebase and its supplement TrEMBL in 2003. *Nucleic acids research*, *31*(1), 365-370.

Bolger, A. M., Lohse, M., & Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*, *30*(15), 2114-2120.

Boyko, A. V., Girich, A. S., Tkacheva, E. S., & Dolmatov, I. Y. (2020). The Eupentacta fraudatrix transcriptome provides insights into regulation of cell transdifferentiation. *Scientific Reports*, *10*(1), 1522.

Brodrick, J. W., Geokas, M. C., & Largman, C. (1976). Human carboxypeptidase B. II. Purification of the enzyme from pancreatic tissue and comparison with the enzymes present in pancreatic secretion. *Biochimica et Biophysica Acta (BBA)-Enzymology*, *452*(2), 468-481.

Buchon, N., Osman, D., David, F. P., Fang, H. Y., Boquete, J. P., Deplancke, B., & Lemaitre, B. (2013). Morphological and molecular characterization of adult midgut compartmentalization in Drosophila. *Cell reports*, *3*(5), 1725-1738.

Campbell, K. J., & White, R. J. (2014). MYC regulation of cell growth through control of transcription by RNA polymerases I and III. *Cold Spring Harbor perspectives in medicine*, *4*(5), a018408.

Candelaria, A. G., Murray, G., File, S. K., & García-Arrarás, J. E. (2006). Contribution of mesenterial muscle dedifferentiation to intestine regeneration in the sea cucumber Holothuria glaberrima. *Cell and tissue research*, *325*(1), 55-65.

Chaillou, T. (2019). Ribosome specialization and its potential role in the control of protein translation and skeletal muscle size. *Journal of Applied Physiology*.

Chen, Y., Banda, M., Speyer, C. L., Smith, J. S., Rabson, A. B., & Gorski, D. H. (2010). Regulation of the expression and activity of the antiangiogenic homeobox gene GAX/MEOX2 by ZEB2 and microRNA-221. *Molecular and cellular biology*, *30*(15), 3902-3913.

Choudhury, P., Hackert, P., Memet, I., Sloan, K. E., & Bohnsack, M. T. (2019). The human RNA helicase DHX37 is required for release of the U3 snoRNP from pre-ribosomal particles. *RNA biology*, *16*(1), 54-68.

Coates, P. J., Nenutil, R., McGregor, A., Picksley, S. M., Crouch, D. H., Hall, P. A., & Wright, E. G. (2001). Mammalian prohibitin proteins respond to mitochondrial stress and decrease during cellular senescence. *Experimental cell research*, *265*(2), 262-273.

Conway, J. R., Lex, A., & Gehlenborg, N. (2017). UpSetR: an R package for the visualization of intersecting sets and their properties. *Bioinformatics*.

Costa-Silva, J., Domingues, D., & Lopes, F. M. (2017). RNA-Seq differential expression analysis: An extended review and a software tool. *PloS one*, *12*(12), e0190152.

Crepin, T., Shalak, V. F., Yaremchuk, A. D., Vlasenko, D. O., McCarthy, A., Negrutskii, B. S., ... & El'skaya, A. V. (2014). Mammalian translation elongation factor eEF1A2: X-ray structure and new features of GDP/GTP exchange mechanism in higher eukaryotes. *Nucleic acids research*, *42*(20), 12939-12948.

Crusoe, M. R., Alameldin, H. F., Awad, S., Boucher, E., Caldwell, A., Cartwright, R., ... & Schwarz, E. (2015). The khmer software package: enabling efficient nucleotide sequence analysis [version 1; referees: 2 approved, 1 approved with reservations].

Davidson, N. M., & Oshlack, A. (2014). Corset: enabling differential gene expression analysis for de novo assembled transcriptomes. *Genome biology*, *15*(7), 1-14.

Dode, L., Andersen, J. P., Raeymaekers, L., Missiaen, L., Vilsen, B., & Wuytack, F. (2005). Functional comparison between secretory pathway Ca2+/Mn2+-ATPase (SPCA) 1 and sarcoplasmic reticulum Ca2+-ATPase (SERCA) 1 isoforms by steady-state and transient kinetic analyses. *Journal of Biological Chemistry*, *280*(47), 39124-39134.

Donato-Santana, C., Rojas-Cartagena, C., & Garcia-Arraras, J. E. (2008). Apoptosis and Surivivin gene expression during Holothuria glaberrima intestinal regeneration.

Dong, X., You, Y., & Wu, J. Q. (2016). Building an RNA sequencing transcriptome of the central nervous system. *The Neuroscientist*, *22*(6), 579-592.

Dush, M. K., & Nascone-Yoder, N. M. (2019). Vangl2 coordinates cell rearrangements during gut elongation. *Developmental Dynamics*, *248*(7), 569-582.

Farin, H. F., Van Es, J. H., & Clevers, H. (2012). Redundant sources of Wnt regulate intestinal stem cells and promote formation of Paneth cells. *Gastroenterology*, *143*(6), 1518-1529.

Finn, R. D., Bateman, A., Clements, J., Coggill, P., Eberhardt, R. Y., Eddy, S. R., ... & Punta, M. (2014). Pfam: the protein families database. *Nucleic acids research*, *42*(D1), D222-D230.

Fiorentini, F., Romero, E., Fraaije, M. W., Faber, K., Hall, M., & Mattevi, A. (2017). Baeyer–Villiger monooxygenase FMO5 as entry point in drug metabolism. *ACS chemical biology*, *12*(9), 2379-2387.

Flanagan, D. J., Phesse, T. J., Barker, N., Schwab, R. H., Amin, N., Malaterre, J., ... & Vincan, E. (2015). Frizzled7 functions as a Wnt receptor in intestinal epithelial Lgr5+ stem cells. *Stem cell reports*, *4*(5), 759-767.

Gamer, L. W., & Wright, C. V. (1993). Murine Cdx-4 bears striking similarities to the Drosophila caudal gene in its homeodomain sequence and early expression pattern. *Mechanisms of development*, *43*(1), 71-81.

Garber, M., Grabherr, M. G., Guttman, M., & Trapnell, C. (2011). Computational methods for transcriptome annotation and quantification using RNA-seq. *Nature methods*, *8*(6), 469-477.

García-Arrarás, J. E., Bello, S. A., & Malavez, S. (2019). The mesentery as the epicenter for intestinal regeneration. In *Seminars in cell & developmental biology* (Vol. 92, pp. 45-54). Academic Press.

García-Arrarás, J. E., Valentín-Tirado, G., Flores, J. E., Rosa, R. J., Rivera-Cruz, A., Miguel-Ruiz, S., ... & Tossas, K. (2011). Cell dedifferentiation and epithelial to mesenchymal transitions during intestinal regeneration in H. glaberrima. *BMC Developmental Biology*, *11*(1), 1-18.

García-Arrarás, J. E., Estrada-Rodgers, L., Santiago, R., Torres, I. I., Díaz-Miranda, L., & Torres-Avillán, I. (1998). Cellular mechanisms of intestine regeneration in the sea cucumber, Holothuria glaberrima Selenka (Holothuroidea: Echinodermata). *Journal of Experimental Zoology*, *281*(4), 288-304.

Genuth, N. R., & Barna, M. (2018). The discovery of ribosome heterogeneity and its implications for gene regulation and organismal life. *Molecular cell*, *71*(3), 364-374.

Girich, A. S., & Boyko, A. V. (2019). Wnt and Frizzled Genes in Echinoderms. *Russian Journal of Marine Biology*, *45*, 302-312.

Girich, A. S., Isaeva, M. P., & Dolmatov, I. Y. (2017). Wnt and frizzled expression during regeneration of internal organs in the holothurian Eupentacta fraudatrix. *Wound Repair and Regeneration*, *25*(5), 828-835.

Golenia, G., Gatie, M. I., & Kelly, G. M. (2017). Frizzled gene expression and negative regulation of canonical WNT–β-catenin signaling in mouse F9 teratocarcinoma cells. *Biochemistry and Cell Biology*, *95*(2), 251-262.

Gonçalves, A. M., Pereira-Santos, A. R., Esteves, A. R., Cardoso, S. M., & Empadinhas, N. (2021). The Mitochondrial Ribosome: A World of Opportunities for Mitochondrial Dysfunction Toward Parkinson's Disease. *Antioxidants & Redox Signaling*, *34*(8), 694-711.

Gong, C., Ai, J., Fan, Y., Gao, J., Liu, W., Feng, Q., ... & Wu, L. (2019). NCAPG promotes the proliferation of hepatocellular carcinoma through PI3K/AKT signaling. *OncoTargets and therapy*, *12*, 8537.

Gottlieb, E., & Steitz, J. A. (1989). Function of the mammalian La protein: evidence for its action in transcription termination by RNA polymerase III. *The EMBO journal*, *8*(3), 851-861.

Grabherr, M. G., Haas, B. J., Yassour, M., Levin, J. Z., Thompson, D. A., Amit, I., ... & Regev, A. (2011). Trinity: reconstructing a full-length transcriptome without a genome from RNA-Seq data. *Nature biotechnology*, *29*(7), 644.

Gregorieff, A., Pinto, D., Begthel, H., Destrée, O., Kielman, M., & Clevers, H. (2005). Expression pattern of Wnt signaling components in the adult intestine. *Gastroenterology*, *129*(2), 626-638.

He, T. C., Sparks, A. B., Rago, C., Hermeking, H., Zawel, L., Da Costa, L. T., ... & Kinzler, K. W. (1998). Identification of c-MYC as a target of the APC pathway. *Science*, *281*(5382), 1509-1512.

Holdt, L. M., Stahringer, A., Sass, K., Pichler, G., Kulak, N. A., Wilfert, W., ... & Teupser, D. (2016). Circular non-coding RNA ANRIL modulates ribosomal RNA maturation and atherosclerosis in humans. *Nature communications*, *7*(1), 12429.

Hölzel, M., Rohrmoser, M., Schlee, M., Grimm, T., Harasim, T., Malamoussi, A., ... & Eick, D. (2005). Mammalian WDR12 is a novel member of the Pes1–Bop1 complex and is required for ribosome biogenesis and cell proliferation. *The Journal of cell biology*, *170*(3), 367-378.

Hou, P., Meng, S., Li, M., Lin, T., Chu, S., Li, Z., ... & Bai, J. (2021). LINC00460/DHX9/IGF2BP2 complex promotes colorectal cancer proliferation and metastasis by mediating HMGA1 mRNA stability depending on m6A modification. *Journal of Experimental & Clinical Cancer Research*, *40*(1), 1-18.

Huang, S. Q., & García-Arrarás, J. E. (2013). Characterization of the intestinal tract neuroendocrine cells in normal and regenerating Holothuria glaberrima.

Iadevaia, V., Liu, R., & Proud, C. G. (2014, December). mTORC1 signaling controls multiple steps in ribosome biogenesis. In *Seminars in cell & developmental biology* (Vol. 36, pp. 113-120). Academic Press.

Ito, S., Horikawa, S., Suzuki, T., Kawauchi, H., Tanaka, Y., Suzuki, T., & Suzuki, T. (2014). Human NAT10 is an ATP-dependent RNA acetyltransferase responsible for N4-acetylcytidine formation in 18 S ribosomal RNA (rRNA). *Journal of Biological Chemistry*, *289*(52), 35724-35730.

Janssen, R., Pechmann, M., & Turetzek, N. (2021). A chelicerate Wnt gene expression atlas: novel insights into the complexity of arthropod Wnt-patterning. *EvoDevo*, *12*, 1-27.

Jiao, L., Liu, Y., Yu, X. Y., Pan, X., Zhang, Y., Tu, J., ... & Li, Y. (2023). Ribosome biogenesis in disease: new players and therapeutic targets. *Signal Transduction and Targeted Therapy*, *8*(1), 15.

Jimenez-Rojo, L., Granchi, Z., Graf, D., & Mitsiadis, T. A. (2012). Stem cell fate determination during development and regeneration of ectodermal organs. *Frontiers in physiology*, *3*, 107.

Jin, Y. R., Han, X. H., Nishimori, K., Ben-Avraham, D., Oh, Y. J., Shim, J. W., & Yoon, J. K. (2020). Canonical WNT/β-Catenin Signaling Activated by WNT9b and RSPO2 cooperation regulates facial morphogenesis in mice. *Frontiers in Cell and Developmental Biology*, *8*, 264.

Kang, W., & Reid, K. B. (2003). DMBT1, a regulator of mucosal homeostasis through the linking of mucosal defense and regeneration?. *FEBS letters*, *540*(1-3), 21-25.

Katoh, K., & Standley, D. M. (2013). MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Molecular biology and evolution*, *30*(4), 772-780.

Kemp, C. R., Willems, E., Wawrzak, D., Hendrickx, M., Agbor Agbor, T., & Leyns, L. (2007). Expression of Frizzled5, Frizzled7, and Frizzled10 during early mouse development and interactions with canonical Wnt signaling. *Developmental dynamics: an official publication of the American Association of Anatomists*, *236*(7), 2011-2019.

Kim, J. S., Crooks, H., Dracheva, T., Nishanian, T. G., Singh, B., Jen, J., & Waldman, T. (2002). Oncogenic β-catenin is required for bone morphogenetic protein 4 expression in human cancer cells. *Cancer research*, *62*(10), 2744-2748.

Komiya, Y., & Habas, R. (2008). Wnt signal transduction pathways. *Organogenesis*, *4*(2), 68-75.

Koren, I., Reem, E., & Kimchi, A. (2010). DAP1, a novel substrate of mTOR, negatively regulates autophagy. *Current biology*, *20*(12), 1093-1098.

Krawetz, R., & Kelly, G. M. (2008). Wnt6 induces the specification and epithelialization of F9 embryonal carcinoma cells to primitive endoderm. *Cellular signalling*, *20*(3), 506-517.

Kupperman, E., An, S., Osborne, N., Waldron, S., & Stainier, D. Y. (2000). A sphingosine-1-phosphate receptor regulates cell migration during vertebrate heart development. *Nature*, *406*(6792), 192-195.

Laneve, P., Gioia, U., Ragno, R., Altieri, F., Di Franco, C., Santini, T., ... & Caffarelli, E. (2008). The tumor marker human placental protein 11 is an endoribonuclease. *Journal of Biological Chemistry*, *283*(50), 34712-34719.

Langfelder, P., & Horvath, S. (2008). WGCNA: an R package for weighted correlation network analysis. *BMC bioinformatics*, *9*(1), 1-13.

Lee, J. E., & Gleeson, J. G. (2011). Cilia in the nervous system: linking cilia function and neurodevelopmental disorders. *Current opinion in neurology*, *24*(2), 98.

Lewis, S. A., Tian, G., Vainberg, I. E., & Cowan, N. J. (1996). Chaperonin-mediated folding of actin and tubulin. *The Journal of cell biology*, *132*(1-2), 1-4.

Li, X., Sun, L., Yang, H., Zhang, L., Miao, T., Xing, L., & Huo, D. (2017). Identification and expression characterization of WntA during intestinal regeneration in the sea cucumber Apostichopus japonicus. *Comparative Biochemistry and Physiology Part B: Biochemistry and Molecular Biology*, *210*, 55-63.

Lin, G., Xu, N., & Xi, R. (2008). Paracrine Wingless signalling controls self-renewal of Drosophila intestinal stem cells. *Nature*, *455*(7216), 1119-1123.

Lin, J., Friesen, M. T., Bocangel, P., Cheung, D., Rawszer, K., & Wigle, J. T. (2005). Characterization of Mesenchyme Homeobox 2 (MEOX2) transcription factor binding to RING finger protein 10. *Molecular and cellular biochemistry*, *275*, 75-84.

Love, M. I., Huber, W., & Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome biology*, *15*(12), 1-21.

Lü, J., Qian, J., Chen, F., Tang, X., Li, C., & Cardoso, W. V. (2005). Differential expression of components of the microRNA machinery during mouse organogenesis. *Biochemical and biophysical research communications*, *334*(2), 319-323.

Madden, T. (2003). The BLAST sequence analysis tool. *The NCBI handbook*.

Mao, J., Fan, S., Ma, W., Fan, P., Wang, B., Zhang, J., ... & Li, L. (2014). Roles of Wnt/β-catenin signaling in the gastric cancer stem cells proliferation and salinomycin treatment. *Cell death & disease*, *5*(1), e1039-e1039.

Mashanov, V. S., & García-Arrarás, J. E. (2011). Gut regeneration in holothurians: a snapshot of recent developments. *The biological bulletin*, *221*(1), 93-109.

Mashanov, V. S., Zueva, O. R., & Garcia-Arraras, J. E. (2012). Expression of Wnt9, TCTP, and Bmp1/Tll in sea cucumber visceral regeneration. *Gene Expression Patterns*, *12*(1-2), 24-35.

Mashanov, V. S., Zueva, O. R., & García-Arrarás, J. E. (2014). Transcriptomic changes during regeneration of the central nervous system in an echinoderm. *BMC genomics*, *15*(1), 1-21.

Mashanov, V. S., Zueva, O. R., Rojas-Catagena, C., & Garcia-Arraras, J. E. (2010). Visceral regeneration in a sea cucumber involves extensive expression of survivin and mortalin homologs in the mesothelium. *BMC Developmental Biology*, *10*, 1-24.

Matsu-Ura, T., Dovzhenok, A., Aihara, E., Rood, J., Le, H., Ren, Y., ... & Hong, C. I. (2016). Intercellular coupling of the cell cycle and circadian clock in adult stem cell culture. *Molecular cell*, *64*(5), 900-912.

Matsuyama, M., Aizawa, S., & Shimono, A. (2009). Sfrp controls apicobasal polarity and oriented cell division in developing gut epithelium. *PLoS genetics*, *5*(3), e1000427.

McCool, M. A., Bryant, C. J., & Baserga, S. J. (2020). MicroRNAs and long non-coding RNAs as novel regulators of ribosome biogenesis. *Biochemical Society Transactions*, *48*(2), 595-612.

McDermaid, A., Monier, B., Zhao, J., Liu, B., & Ma, Q. (2019). Interpretation of differential gene expression results of RNA-seq data: review and integration. *Briefings in bioinformatics*, *20*(6), 2044-2054.

Medina-Feliciano, J. G., & García-Arrarás, J. E. (2021). Regeneration in echinoderms: Molecular advancements. *Frontiers in Cell and Developmental Biology*, *9*, 768641.

Medina-Feliciano, J. G., Pirro, S., García-Arrarás, J. E., Mashanov, V., & Ryan, J. F. (2021). Draft genome of the sea cucumber Holothuria glaberrima, a model for the study of regeneration. *Frontiers in Marine Science*, *8*, 603410.

Mentink, R. A., Rella, L., Radaszkiewicz, T. W., Gybel, T., Betist, M. C., Bryja, V., & Korswagen, H. C. (2018). The planar cell polarity protein VANG-1/Vangl negatively regulates Wnt/β-catenin signaling through a Dvl dependent mechanism. *PLoS Genetics*, *14*(12), e1007840.

Miguel-Ruiz, S., José, E., & García-Arrarás, J. E. (2007). Common cellular events occur during wound healing and organ regeneration in the sea cucumber Holothuria glaberrima. *BMC Developmental Biology*, *7*(1), 1-19.

Miller, J. L., deWet, B. J. M., Martinez-Pomares, L., Radcliffe, C. M., Dwek, R. A., Rudd, P. M., & Gordon, S. (2008). The mannose receptor mediates dengue virus infection of macrophages. *PLoS pathogens*, *4*(2), e17.

Miller, J. R. (2001). The wnts. *Genome biology*, *3*, 1-15.

Miyoshi, H., Ajima, R., Luo, C. T., Yamaguchi, T. P., & Stappenbeck, T. S. (2012). Wnt5a potentiates TGF-β signaling to promote colonic crypt regeneration after tissue injury. *Science*, *338*(6103), 108-113.

Mollenhauer, J., Herbertz, S., Holmskov, U., Tolnay, M., Krebs, I., Merlo, A., ... & Poustka, A. (2000). DMBT1 encodes a protein involved in the immune defense and in epithelial differentiation and is highly unstable in cancer. *Cancer research*, *60*(6), 1704-1710.

Molven, A., Njølstad, P. R., & Fjose, A. (1991). Genomic structure and restricted neural expression of the zebrafish wnt-1 (int-1) gene. *The EMBO Journal*, *10*(4), 799-807.

Muro, Y., Yamada, T., Himeno, M., & Sugimoto, K. (1998). cDNA cloning of a novel autoantigen targeted by a minor subset of anti-centromere antibodies. *Clinical & Experimental Immunology*, *111*(2), 372-376.

Murray, G., & García-Arrarás, J. E. (2004). Myogenesis during holothurian intestinal regeneration. *Cell and tissue research*, *318*(3), 515-524.

Murray, G., Gonzalez, T. C., & Garcia-Arraras, J. E. (2001). Myogenesis and visceral muscle regeneration in the echinoderm Holothuria glaberrima. In *FASEB JOURNAL* (Vol. 15, No. 4, pp. A489-A489). 9650 Rockville Pike, Bethesda, MD 20814-3998 USA: Federation Amer Soc Exp Biol.

Nguyen, L. T., Schmidt, H. A., Von Haeseler, A., & Minh, B. Q. (2015). IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Molecular biology and evolution*, *32*(1), 268-274.

Nieves-Ríos, C., Alvarez-Falcón, S., Malavez, S., Rodriguez-Otero, J., & García-Arrarás, J. E. (2020). The nervous system component of the mesentery of the sea cucumber Holothuria glaberrima in normal and regenerating animals. *Cell and tissue research*, *380*(1), 67-77.

Orsolic, I., Jurada, D., Pullen, N., Oren, M., Eliopoulos, A. G., & Volarevic, S. (2016, June). The relationship between the nucleolus and cancer: Current evidence and emerging paradigms. In *Seminars in cancer biology* (Vol. 37, pp. 36-50). Academic Press.

Ortiz-Pineda, P. A., Ramírez-Gómez, F., Pérez-Ortiz, J., González-Díaz, S., Jesús, S. D., Hernández-Pasos, J., ... & García-Arrarás, J. E. (2009). Gene expression profiling of intestinal regeneration in the sea cucumber. *BMC genomics*, *10*(1), 1-21.

Otto, A., Schmidt, C., Luke, G., Allen, S., Valasek, P., Muntoni, F., ... & Patel, K. (2008). Canonical Wnt signalling induces satellite-cell proliferation during adult skeletal muscle regeneration. *Journal of cell science*, *121*(17), 2939-2950.

Ouladan, S., & Gregorieff, A. (2021). Taking a step back: Insights into the mechanisms regulating gut epithelial dedifferentiation. *International Journal of Molecular Sciences*, *22*(13), 7043.

Park, O. H., Ha, H., Lee, Y., Boo, S. H., Kwon, D. H., Song, H. K., & Kim, Y. K. (2019). Endoribonucleolytic cleavage of m6A-containing RNAs by RNase P/MRP complex. *Molecular cell*, *74*(3), 494-507.

Pasten, C. C., Rosa, R., Noya, M., & Garcia-Arraras, J. (2010). Ubiquitin proteasome pathway and organogenesis of Holothuria glaberrima. *Developmental Biology*, *1*(344), 486.

Patro, R., Duggal, G., Love, M. I., Irizarry, R. A., & Kingsford, C. (2017). Salmon provides fast and bias-aware quantification of transcript expression. *Nature methods*, *14*(4), 417-419.

Paytubi, S., Wang, X., Lam, Y. W., Izquierdo, L., Hunter, M. J., Jan, E., ... & Proud, C. G. (2009). ABC50 promotes translation initiation in mammalian cells. *Journal of Biological Chemistry*, *284*(36), 24061-24073.

Pelletier, J., Thomas, G., & Volarević, S. (2018). Ribosome biogenesis in cancer: new players and therapeutic avenues. *Nature Reviews Cancer*, *18*(1), 51-63.

Perochon, J., Carroll, L. R., & Cordero, J. B. (2018). Wnt signalling in intestinal stem cells: lessons from mice and flies. *Genes*, *9*(3), 138.Perochon, J.; Carroll, L.; Cordero, J. Wnt Signalling in Intestinal Stem Cells: Lessons from Mice and Flies. *Genes* 2018, *9*, 138.

Pontén, F., Jirström, K., & Uhlen, M. (2008). The Human Protein Atlas—a tool for pathology. *The Journal of Pathology: A Journal of the Pathological Society of Great Britain and Ireland*, *216*(4), 387-393.

Popay, T. M., Wang, J., Adams, C. M., Howard, G. C., Codreanu, S. G., Sherrod, S. D., ... & Tansey, W. P. (2021). MYC regulates ribosome biogenesis and mitochondrial gene expression programs through its interaction with host cell factor–1. *Elife*, *10*, e60191.

Prakash, V., Carson, B. B., Feenstra, J. M., Dass, R. A., Sekyrova, P., Hoshino, A., ... & Vincent, C. T. (2019). Ribosome biogenesis during cell cycle arrest fuels EMT in development and disease. *Nature communications*, *10*(1), 2110.

Quispe-Parra, D. J., Medina-Feliciano, J. G., Cruz-González, S., Ortiz-Zuazaga, H., & García-Arrarás, J. E. (2021). Transcriptomic analysis of early stages of intestinal regeneration in Holothuria glaberrima. *Scientific reports*, *11*(1), 346. https://doi.org/10.1038/s41598-020-79436-2

Quispe-Parra, D., Valentín, G., & García-Arrarás, J. E. (2020). A roadmap for intestinal regeneration. *International Journal of Developmental Biology*, *65*(4-5-6), 427-437.

Santiago, P., Roig-López, J. L., Santiago, C., & García-Arrarás, J. E. (2000). Serum amyloid A protein in an echinoderm: its primary structure and expression during intestinal regeneration in the sea cucumber Holothuria glaberrima. *Journal of Experimental Zoology*, *288*(4), 335-344.

Sato, T., Van Es, J. H., Snippert, H. J., Stange, D. E., Vries, R. G., Van Den Born, M., ... & Clevers, H. (2011). Paneth cells constitute the niche for Lgr5 stem cells in intestinal crypts. *Nature*, *469*(7330), 415-418.

Savino, T. M., Bastos, R., Jansen, E., & Hernandez-Verdun, D. (1999). The nucleolar antigen Nop52, the human homologue of the yeast ribosomal RNA processing RRP1, is recruited at late stages of nucleologenesis. *Journal of cell science*, *112*(12), 1889-1900.

Shi, D. L. (2022). Planar cell polarity regulators in asymmetric organogenesis during development and disease. *Journal of Genetics and Genomics*.

Shinohara, N., Ohbayashi, I., & Sugiyama, M. (2014). Involvement of rRNA biosynthesis in the regulation of CUC1 gene expression and pre-meristematic cell mound formation during shoot regeneration. *Frontiers in Plant Science*, *5*, 159.

Sigal, M., Logan, C. Y., Kapalczynska, M., Mollenkopf, H. J., Berger, H., Wiedenmann, B., ... & Meyer, T. F. (2017). Stromal R-spondin orchestrates gastric epithelial stem cells and gland homeostasis. *Nature*, *548*(7668), 451-455.

Silva, J. P., Lelianova, V. G., Ermolyuk, Y. S., Vysokov, N., Hitchen, P. G., Berninghausen, O., ... & Ushkaryov, Y. A. (2011). Latrophilin 1 and its endogenous ligand Lasso/teneurin-2 form a high-affinity transsynaptic receptor pair with signaling capabilities. *Proceedings of the National Academy of Sciences*, *108*(29), 12113-12118.

Simão, F. A., Waterhouse, R. M., Ioannidis, P., Kriventseva, E. V., & Zdobnov, E. M. (2015). BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics*, *31*(19), 3210-3212.

Singh, S., Vanden Broeck, A., Miller, L., Chaker-Margot, M., & Klinge, S. (2021). Nucleolar maturation of the human small subunit processome. *Science*, *373*(6560), eabj5338.

Smith-Unna, R., Boursnell, C., Patro, R., Hibberd, J. M., & Kelly, S. (2016). TransRate: reference-free quality assessment of de novo transcriptome assemblies. *Genome research*, *26*(8), 1134-1144.

Sobczak, M., Pitt, A. R., Spickett, C. M., & Robaszkiewicz, A. (2019). PARP1 Co-regulates EP300–BRG1-dependent transcription of genes involved in breast cancer cell proliferation and DNA repair. *Cancers*, *11*(10), 1539.

Soudy, M., Anwar, A. M., Ahmed, E. A., Osama, A., Ezzeldin, S., Mahgoub, S., & Magdeldin, S. (2020). UniprotR: Retrieving and visualizing protein sequence and functional information from Universal Protein Resource (UniProt knowledgebase). *Journal of Proteomics*, *213*, 103613.

Stamatakis, A. (2014). RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics*, *30*(9), 1312-1313.

Strub, G. M., Paillard, M., Liang, J., Gomez, L., Allegood, J. C., Hait, N. C., ... & Spiegel, S. (2011). Sphingosine-1-phosphate produced by sphingosine kinase 2 in mitochondria interacts with prohibitin 2 to regulate complex IV assembly and respiration. *The FASEB Journal*, *25*(2), 600.

Suh, H. N., Kim, M. J., Jung, Y. S., Lien, E. M., Jun, S., & Park, J. I. (2017). Quiescence exit of tert+ stem cells by Wnt/β-catenin is indispensable for intestinal regeneration. *Cell reports*, *21*(9), 2571-2584.

Sun, C., Cao, C., Zhao, T., Guo, H., Fleming, B. C., Owens, B., ... & Wei, L. (2023). A2M inhibits inflammatory mediators of chondrocytes by blocking IL-1β/NF-κB pathway. *Journal of Orthopaedic Research®*, *41*(1), 241-248.

Sun, L. N., Yang, H. S., Chen, M. Y., & Xu, D. X. (2013). Cloning and expression analysis of Wnt6 and Hox6 during intestinal regeneration in the sea cucumber Apostichopus japonicus. *Genet. Mol. Res*, *12*(4), 5321-5334.

Suzek, B. E., Huang, H., McGarvey, P., Mazumder, R., & Wu, C. H. (2007). UniRef: comprehensive and non-redundant UniProt reference clusters. *Bioinformatics*, *23*(10), 1282-1288.

Takeuchi, K., Shimizu, T., Ohishi, N., Seyama, Y., Takaku, F., & Yotsumoto, H. (1989). Purification of human lung angiotensin-converting enzyme by high-performance liquid chromatography: Properties and N-terminal amino acid sequence. *The Journal of Biochemistry*, *106*(3), 442-445.

Tanigawa, G., Orci, L., Amherdt, M., Ravazzola, M., Helms, J. B., & Rothman, J. E. (1993). Hydrolysis of bound GTP by ARF protein triggers uncoating of Golgi-derived COP-coated vesicles. *The Journal of cell biology*, *123*(6 Pt 1), 1365-1371.

Tetsu, O., & McCormick, F. (1999). β-Catenin regulates expression of cyclin D1 in colon carcinoma cells. *Nature*, *398*(6726), 422-426.

Tian, A., Benchabane, H., & Ahmed, Y. (2018). Wingless/Wnt signaling in intestinal development, homeostasis, regeneration and tumorigenesis: a Drosophila perspective. *Journal of Developmental Biology*, *6*(2), 8.

Tian, A., Benchabane, H., Wang, Z., & Ahmed, Y. (2016). Regulation of stem cell proliferation and cell fate specification by Wingless/Wnt signaling gradients enriched at adult intestinal compartment boundaries. *PLoS genetics*, *12*(2), e1005822.

Tossas, K., González-Conty, E., Medina-Ortiz, W., Roig-López, J. L., & García-Arrarás, J. (2004). Tubulin expression and the regeneration of the enteric nervous system in the sea cucumber Holothuria glaberrima. In *Echinoderms: Munchen* (pp. 639-639). CRC Press.

Tossas, K., Qi-Huang, S., Cuyar, E., & García-Arrarás, J. E. (2014). Temporal and spatial analysis of enteric nervous system regeneration in the sea cucumber Holothuria glaberrima. *Regeneration*, *1*(3), 10-26.

Truitt, M. L., & Ruggero, D. (2016). New frontiers in translational control of the cancer genome. *Nature Reviews Cancer*, *16*(5), 288-304.

Turi, Z., Lacey, M., Mistrik, M., & Moudry, P. (2019). Impaired ribosome biogenesis: mechanisms and relevance to cancer and aging. *Aging (Albany NY)*, *11*(8), 2512.

van Es, J. H., Jay, P., Gregorieff, A., van Gijn, M. E., Jonkheer, S., Hatzis, P., ... & Clevers, H. (2005). Wnt signalling induces maturation of Paneth cells in intestinal crypts. *Nature cell biology*, *7*(4), 381-386.

Vaz, F. M., Fouchier, S. W., Ofman, R., Sommer, M., & Wanders, R. J. (2000). Molecular and biochemical characterization of rat γ-trimethylaminobutyraldehyde dehydrogenase and evidence for the involvement of human aldehyde dehydrogenase 9 in carnitine biosynthesis. *Journal of Biological Chemistry*, *275*(10), 7390-7394.

Vergnes, L., Lee, J. M., Chin, R. G., Auwerx, J., & Reue, K. (2013). Diet1 functions in the FGF15/19 enterohepatic signaling axis to modulate bile acid and lipid levels. *Cell metabolism*, *17*(6), 916-928.

Vosburg, C., Reynolds, M., Noel, R., Shippy, T., Hosmani, P. S., Flores-Gonzalez, M., ... & D'Elia, T. (2021). Utilizing a chromosomal-length genome assembly to annotate the Wnt signaling pathway in the Asian citrus psyllid, Diaphorina citri. *Gigabyte*, *2021*, 1-15.

Wang, N., Bosc, C., Ryul Choi, S., Boulan, B., Peris, L., Olieric, N., ... & Huang, H. (2019). Structural basis of tubulin detyrosination by the vasohibin–SVBP enzyme complex. *Nature structural & molecular biology*, *26*(7), 571-582.

Wang, Z., Tian, A., Benchabane, H., Tacchelly-Benites, O., Yang, E., Nojima, H., & Ahmed, Y. (2016). The ADP-ribose polymerase Tankyrase regulates adult intestinal stem cell proliferation during homeostasis in Drosophila. *Development*, *143*(10), 1710-1720.

Wu, X., Chen, T., Huo, D., Yu, Z., Ruan, Y., Cheng, C., ... & Ren, C. (2020). Transcriptomic analysis of sea cucumber (Holothuria leucospilota) coelomocytes revealed the echinoderm cytokine response during immune challenge. *BMC genomics*, *21*, 1-17.

Wu, X., Tu, X., Joeng, K. S., Hilton, M. J., Williams, D. A., & Long, F. (2008). Rac1 activation controls nuclear localization of β-catenin during canonical Wnt signaling. *Cell*, *133*(2), 340-353.

Wu, Y., Ginther, C., Kim, J., Mosher, N., Chung, S., Slamon, D., & Vadgama, J. V. (2012). Expression of Wnt3 Activates Wnt/β-Catenin Pathway and Promotes EMT-like Phenotype in Trastuzumab-Resistant HER2-Overexpressing Breast Cancer CellsWnt3 Promotes EMT and Acquired Resistance to Trastuzumab. *Molecular Cancer Research*, *10*(12), 1597-1606.

Xiao, R., Sun, Y., Ding, J. H., Lin, S., Rose, D. W., Rosenfeld, M. G., ... & Li, X. (2007). Splicing regulator SC35 is essential for genomic stability and cell proliferation during mammalian organogenesis. *Molecular and cellular biology*, *27*(15), 5393-5402.

Xing, Y. H., Yao, R. W., Zhang, Y., Guo, C. J., Jiang, S., Xu, G., ... & Chen, L. L. (2017). SLERT regulates DDX21 rings associated with Pol I transcription. *Cell*, *169*(4), 664-678.

Xue, L., Yi, H., Huang, Z., Shi, Y. B., & Li, W. X. (2011). Global gene expression during the human organogenesis: from transcription profiles to function predictions. *International journal of biological sciences*, *7*(7), 1068-1076.

Yaglova, N. V., Tsomartova, D. A., Obernikhin, S. S., & Nazimova, S. V. (2019). The role of the canonical Wnt-signaling pathway in morphogenesis and regeneration of the adrenal cortex in rats exposed to the endocrine disruptor dichlorodiphenyltrichloroethane during prenatal and postnatal development. *Biology Bulletin*, *46*, 74-81.

Yeoh, S., Pope, B., Mannherz, H. G., & Weeds, A. (2002). Determining the differences in actin binding by human ADF and cofilin. *Journal of molecular biology*, *315*(4), 911-925.

Yu, G., Wang, L. G., Han, Y., & He, Q. Y. (2012). clusterProfiler: an R package for comparing biological themes among gene clusters. *Omics: a journal of integrative biology*, *16*(5), 284-287.

Yu, G., Wang, L. G., Yan, G. R., & He, Q. Y. (2015). DOSE: an R/Bioconductor package for disease ontology semantic and enrichment analysis. *Bioinformatics*, *31*(4), 608-609.

Yuan, J., Gao, Y., Sun, L., Jin, S., Zhang, X., Liu, C., ... & Xiang, J. (2019). Wnt signaling pathway linked to intestinal regeneration via evolutionary patterns and gene expression in the sea cucumber Apostichopus japonicus. *Frontiers in Genetics*, *10*, 112.

Zhang, X., Lin, P. Y., Liakath-Ali, K., & Südhof, T. C. (2022). Teneurins assemble into presynaptic nanoclusters that promote synapse formation via postsynaptic non-teneurin ligands. *Nature Communications*, *13*(1), 2297.

Zhang, X., Sun, L., Yuan, J., Sun, Y., Gao, Y., Zhang, L., ... & Xiang, J. (2017). The sea cucumber genome provides insights into morphological evolution and visceral regeneration. *PLoS Biology*, *15*(10), e2003790.

Zhao, S., Chen, Y., Chen, F., Huang, D., Shi, H., Lo, L. J., ... & Peng, J. (2019). Sas10 controls ribosome biogenesis by stabilizing Mpp10 and delivering the Mpp10–Imp3–Imp4 complex to nucleolus. *Nucleic Acids Research*, *47*(6), 2996-3012.

Zheng, X. L., & Yu, H. G. (2018). Wnt6 contributes tumorigenesis and development of colon cancer via its effects on cell proliferation, apoptosis, cell-cycle and migration. *Oncology Letters*, *16*(1), 1163-1172.