Spring 2022

Differential Gene Expression Analysis of Zebrafish Embryos Exposed to Simulated Microgravity and Insights into Cellular Effects

Nicholas Lien
San Jose State University

Differential Gene Expression Analysis of Zebrafish Embryos Exposed to Simulated Microgravity and Insights into Cellular Effects

A Project

Presented to the

Department of Computer Science

San Jose State University

In Partial Fulfillment

Of the Requirements for the Degree

By

Nicholas Lien

May, 2022

ABSTRACT

Spaceflight consists of many dangers which adversely affects the health of astronauts through hazards such as microgravity and cosmic radiation. One area that is still poorly understood is how spaceflight impacts human reproductive health. This study aims to shed insight into how microgravity may impact the development of embryos. Differential gene expression analysis was performed via Jupyter Notebook and SLURM scripts and run on SJSU's HPC server as a method of implementing NASA GeneLab's RNA-Seq Consensus Pipeline. Data for this project utilized RNA-Seq files for early-stage embryonic zebrafish (*Danio rerio*), stored under GLDS-373. Gene Set Enrichment Analysis was performed to gain a clearer understanding of which types of genes are impacted by microgravity, and to provide greater statistical significance to the differential gene expression results. The findings in this study found that there was a relationship between microgravity and upregulation in genes related to cell proliferation, differentiation, and development. However more studies are required before a mechanism can be identified to explain these observations and risks mitigated for future astronauts and their children.

Keywords: Differential gene expression, gene enrichment analysis, microgravity, embryogenesis

ACKNOWLEDGEMENTS

I now thank the following people for their role in supporting me during the completion of this project.

- My advisor Dr. Philip Heller: Thank you for allowing me to be a part of your lab, and for introducing me to this project. You provided so much advice, support, and guidance. I truly would not have completed this project without you.

- Dr. Cleber Ouverney and Dr. Wendy Lee: Thank you for agreeing to be a part of my committee. Both of you provided invaluable advice in tackling this project, and your teachings helped make me a better Bioinformatician.

- Dr. Amanda Saravia-Butler and Dr. Lauren Sanders: Thank you for providing guiding and insight into the development of the GeneLab RNA-Seq Consensus Pipeline.

- The late Steven Boring: Thank you for all your work in maintaining the SJSU HPC, and for setting up and installing all the programs required by the GeneLab RNA-Seq Consensus Pipeline.

- My fellow M.S. Bioinformatics peers, who helped make me feel welcome in the Bioinformatics community and who helped me through difficult times.

- My parents, who worked tirelessly to raise and support me throughout my life and especially during my education. I would not be where I am now without them.

TABLE OF CONTENTS

TABLE OF FIGURES

INDEX OF TABLES

I. INTRODUCTION

A. *Human Interest in Space Exploration and the Challenges Involved*

Humans have always had a desire to explore the unknown, but it is only until recently that humanity gained the technology required to send humans out into the vastness of space. Only in the last century have we seen a surge in technical innovation that allowed the first man to set foot onto the moon, and probes out to each of our neighboring planets. Today, the US's endeavors in space exploration are focused on first returning humans to the moon with the Artemis Project, and the establishment of a moon base which will allow for future missions to other celestial bodies including Mars[1]. The Artemis Project was first announced in March of 2019 by then Vice President Mike Pence with the goal to send the first woman to the moon by 2024. Broadly, the program will consist of two phases. Phase one will utilize existing as well as brand new technologies to allow a crewed space craft to make a landing on the lunar south pole. In phase two more humans will be sent to the moon to develop a stronger presence, allowing for the development of additional lunar projects and preparation for future missions to Mars and beyond.

Space itself poses numerous hazards that astronauts become exposed to during spaceflight, which consist of: space radiation, altered gravity (or lack thereof), isolation/confinement, distance from Earth, and hostile/closed environments [2]. Outside of Earth's protective magnetosphere, astronauts are constantly bombarded with galactic cosmic rays and high charge, high energy ions. Damage from these types of radiation can result in DNA damage, development of cancerous tumors, and degeneration of various tissues. The small size of space crafts, small crews, and large distance from Earth result in small, closed environments

and long months spent in isolation. These can have negative behavioral impacts such as an increase in stress and decreases in performance.

As announced by then Vice President Mike Pence, one of the goals of the Artemis Project is to send the first woman to the moon [1]. This signals the inclusion of more female astronauts, and protecting their health includes protecting their reproductive health. As humans set their sights on Mars and other long-distance celestial bodies in space, they will inevitably spend more and more time exposed to the hazards of space. We must therefore consider not only the immediate damage to astronauts but also study how these environments may affect our ability to reproduce outside of Earth. Understanding how our reproductive functions are disrupted by hazards such as microgravity and radiation exposure will be especially important as humans seek to colonize new space frontiers [3]. Without this knowledge, it will be difficult to mitigate space-related health risks to both potential mothers and their children.

The research presented here aims to shed light on astronaut reproductive health by examining how microgravity may change the expression of genes in early embryos. It will do so by examining mRNA data collected from early-stage zebrafish (*Danio rerio*) embryos exposed to either normal gravity or simulated microgravity via a rotary system (explained in more detail in section C). By analyzing which genes are significantly upregulated or downregulated, a few target genes may be identified for future research and some early conclusions can be made on the effects of microgravity on early-stage embryos.

B. *A Brief Review of Early Embryogenesis*

Starting with fertilization, an embryo (called a zygote at this stage) begins with a series of cleavages which double the number of cells that make up the early embryo [3,4]. After the third

cleavage, the cells begin to compact tightly together, a process mediated by maternal E-cadherin [3,5]. At this stage, the cells that make up the embryo (now called a morula) undergo the first fate decision where the cells in the outer layer commit to giving rise to trophectoderm epithelium and the inner cells form the inner cell mass [3,6]. The next major change is called cavitation, where a fluid cavity forms in the inner region and the inner cells undergo a second fate decision. Here the embryo is called a blastocyst and cells in contact with the cavity give rise to the primitive endoderm while the cells not in contact form the epiblast [3].

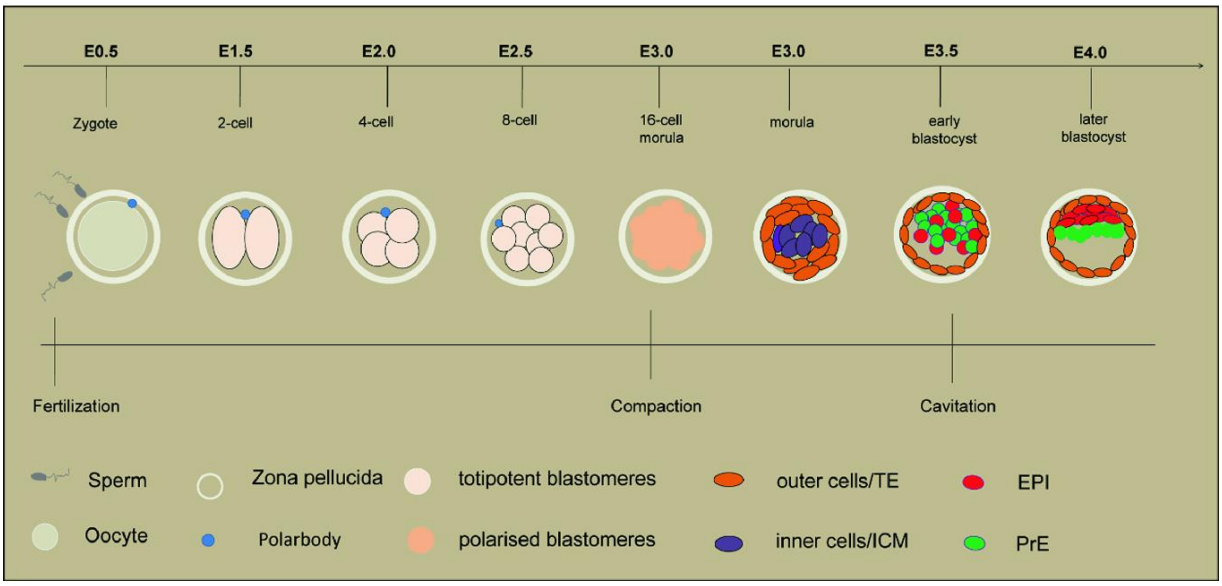

Figure 1. Diagram of the Stages of Early Embryogenesis from [3]

C. *Effects of Microgravity on Early Embryogenesis*

Microgravity seems to have detrimental effects upon early embryos, with the severity of the effects varying with time of exposure. It has been observed that mouse and zebrafish (*Danio rerio*) zygotes exposed early to microgravity conditions tend to have high mortality rates, failing to produce viable offspring [7,8]. Zygotes exposed to microgravity conditions later in development seem to adapt more easily, shown in lower mortality rates among later stage

zebrafish zygotes and middle and late-stage pregnant mice producing viable offspring. Mouse and Zebrafish zygotes also exhibited slower growth rates and longer development time between stages in microgravity conditions [7,9-11]. Furthermore, many surviving embryos exhibited many growth deformities [7,10]. Mouse blastocysts exposed to microgravity early in development were observed to contain fewer cells than control blastocysts, and many blastocyst cells failed to differentiate into the proper cell lines [10]. Zebrafish embryos also exhibited several deformities in the circulatory system, otolith, and brain beam structures [7]. These observations would suggest that embryos are sensitive to microgravity soon after fertilization. Underlying molecular mechanisms are still poorly understood, though many studies have offered suggestions for what has been observed.

D. *NASA Genelab RNA-Seq Consensus Pipeline*

The NASA Genelab is a public database where omics data from spaceflight-related experiments can be hosted and accessed by other scientists [12]. The database was created as part of a project to integrate the wide variety in spaceflight experimental conditions such that it can be presented to the public in a fully standardized format. A second goal is to create a standardized procedure by which RNA-Seq data (data generated by sequencing an organism's mRNA) can be processed. A standardized data processing pipeline has several benefits, including consistency in the processing methods and in final processed data, which allows easier cross-analysis of future studies. Another benefit of fully standardized data with a standardized method of analysis is that it can lead to bulk analysis, where multiple data sets can be analyzed in a single run. Doing so will save time and computational resources, as well as giving the results increased statistical strength. The RNA-Seq Consensus Pipeline (RCP) consists of three main steps: Data Preprocessing, Read Mapping and Sample Quantification, and Differential Gene

Expression Analysis (Figure 3). These will be explained in more detail in the Methods section of this report.

E. *GLDS-373*

This project utilizes data from an immunological study, and is stored in the NASA Genelab database under Genelab Dataset 373 (GLDS-373) [13]. In the study, wild-type AB Zebrafish (*Danio rerio*) 3-12 months of age were kept in standard laboratory conditions and allowed to breed. Of the resulting embryos, one-cell stage embryos were selected and split into four groups. Two groups each were microinjected with either 2 nL poly I:C (a double-stranded RNA analog designed to mimic viral RNA and stimulate an immunological response) or microinjection buffer (mock buffer control). Then one poly I:C group and one mock buffer group was then subjected to either normal gravity (1G) in a standard cell culture dish or simulated microgravity (uG) in a rotary cell culture system. The resulting groups consist of the following conditions: normal gravity and mock buffer controls, simulated microgravity condition and mock buffer control, normal gravity control and poly I:C condition, and simulated microgravity and poly I:C conditions. These groups are described in Table I. Each group consists of three samples (biological replicates). Microgravity was simulated by rotating samples at 12 rpm (rotations per minute) to accelerate the samples to 4×10^{-3} to 7.2×10^{-3} g. 12 hours after fertilization, RNA was collected from each sample using TRIzol reagent and quantified via a NanoDrop ND-1000 instrument and agarose gel electrophoresis. RNA libraries were prepared using a KAPA Stranded RNA-Seq Library Prep Kit. An Illumina HiSeq 4000 instrument was then used to sequence the RNA libraries, producing paired end 150 bp reads. These raw reads are utilized in this project, and can be found within the NASA Genelab Data Repository under the label GLDS-373.

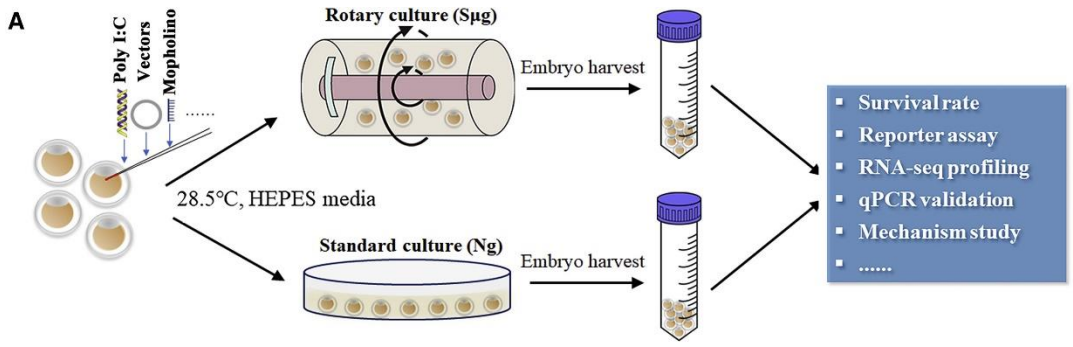

Figure 2. Illustration of the Experimental Procedure from [13]

II. METHODS

A. High Performance Computing Cluster

This project was run on San Jose State University's College of Science High Performance Computing (HPC) server. An HPC is a computing system consisting of strong multi-core servers and high amounts of memory (124GB RAM per compute node) enabling many computationally demanding tasks to be run efficiently [14, 15]. Many of the programs used in the RCP are computationally intensive and require large amounts of memory, making the HPC vital for completion of the tasks within the RCP. The College of Science's HPC is Linux-based, requiring jobs be written in a bash script and submitted via a SLURM job scheduler.

This project also utilizes JupyterHub to access the HPC servers and document the code written for the RCP in python notebooks. This method of implementation allows for the introduction of the RCP to new users without burdening them with prior installation of other managing programs [16]. JupyterHub also supports other coding languages enabling users to work with programs in other environments such as RStudio. Written code utilized in this project can be found at: https://github.com/NL-95/GLDS-373_Differential_Analysis_Files

B. RNA-Seq Consensus Pipeline

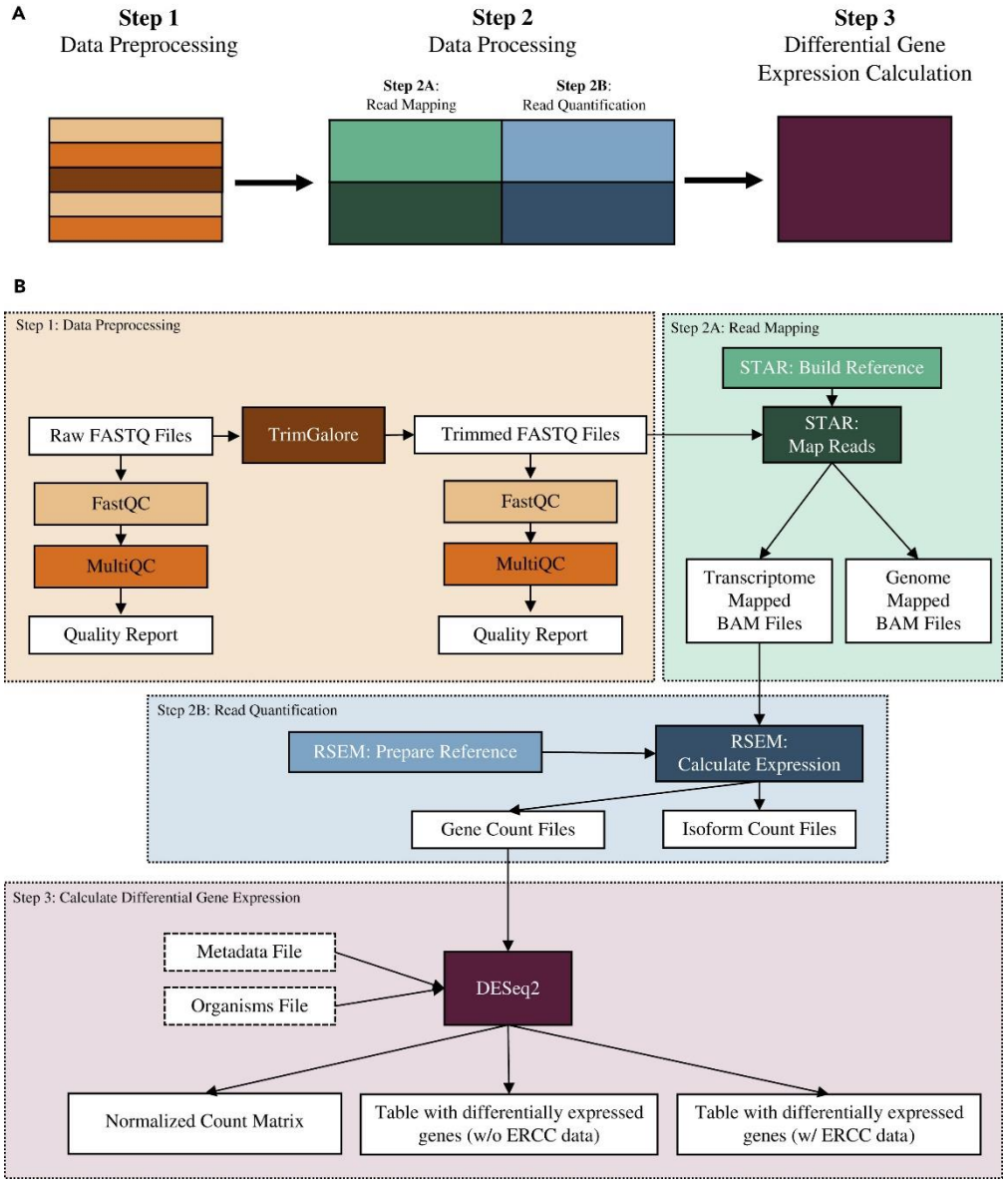

Figure 3. Diagram of the NASA RNA-Seq Consensus Pipeline from [12]

1) *Data Preprocessing:*

a) Obtaining Raw RNA-Seq Data: The first step in the RCP is to download FastQ files for GLDS-373 from the GeneLab data repository at https://genelab-data.ndc.nasa.gov/genelab/accession/GLDS-373/. GLDS-373 contains forward and reverse reads for all 12 samples (4 groups in total). Within the dataset, this results in a total of 24 fastq.gz files, with a forward read (suffix _1) and a reverse read (suffix _2) for each sample, listed in Table I. This project did not analyze the poly I:C condition, so only data related to the six samples corresponding to mock buffer were utilized. For the samples utilized in this project, they correspond to samples 15-17 (normal gravity and mock buffer controls) and samples 21-23 (simulated microgravity condition and mock buffer control). In total, 12 fastq.gz files were downloaded corresponding to forward and reverse reads for samples 15-17 and 21-23. These files are: SRR11185415_1 - SRR11185417_2, and SRR11185421_1 - SRR11185423_2.

Table I

Sample Groups and Treatment Conditions

Sample	Associated Files	Microgravity Simulation	Treatment Condition
Ng-1	GLDS-373_rna-seq_SRR11185415_1.fastq.gz GLDS-373_rna-seq_SRR11185415_2.fastq.gz	1G	Mock buffer
Ng-2	GLDS-373_rna-seq_SRR11185416_1.fastq.gz GLDS-373_rna-seq_SRR11185416_2.fastq.gz	1G	Mock buffer
Ng-3	GLDS-373_rna-seq_SRR11185417_1.fastq.gz GLDS-373_rna-seq_SRR11185417_2.fastq.gz	1G	Mock buffer
Ng-pIC-1	GLDS-373_rna-seq_SRR11185418_1.fastq.gz GLDS-373_rna-seq_SRR11185418_2.fastq.gz	1G	Poly I:C
Ng-pIC-2	GLDS-373_rna-seq_SRR11185419_1.fastq.gz GLDS-373_rna-seq_SRR11185419_2.fastq.gz	1G	Poly I:C
Ng-pIC-3	GLDS-373_rna-seq_SRR11185420_1.fastq.gz GLDS-373_rna-seq_SRR11185420_2.fastq.gz	1G	Poly I:C
Smg-1	GLDS-373_rna-seq_SRR11185421_1.fastq.gz GLDS-373_rna-seq_SRR11185421_2.fastq.gz	uG with rotary cell culture system	Mock buffer

Smg-2	GLDS-373_rna-seq_SRR11185422_1.fastq.gz GLDS-373_rna-seq_SRR11185422_2.fastq.gz	uG with rotary cell culture system	Mock buffer
Smg-3	GLDS-373_rna-seq_SRR11185423_1.fastq.gz GLDS-373_rna-seq_SRR11185423_2.fastq.gz	uG with rotary cell culture system	Mock buffer
Smg-pIC-1	GLDS-373_rna-seq_SRR11185424_1.fastq.gz GLDS-373_rna-seq_SRR11185424_2.fastq.gz	uG with rotary cell culture system	Poly I:C
Smg-pIC-2	GLDS-373_rna-seq_SRR11185425_1.fastq.gz GLDS-373_rna-seq_SRR11185425_2.fastq.gz	uG with rotary cell culture system	Poly I:C
Smg-pIC-3	GLDS-373_rna-seq_SRR11185426_1.fastq.gz GLDS-373_rna-seq_SRR11185426_2.fastq.gz	uG with rotary cell culture system	Poly I:C

b) Quality Control: Quality of the raw reads was assessed using FastQC version 0.11.9 which were then summarized into a single report by MultiQC version 1.11 [17, 18]. Removal of low-quality reads (identified as bases with a phred score 20 or below) and adapter sequences was performed using TrimGalore version 0.6.6 [19]. After their removal, the reads were again assessed and summarized using FastQC and MultiQC to check for improvement in quality scores and assess if further trimming is needed.

2) *Read Mapping and Quantification:* Next, the trimmed reads were aligned and mapped to the *Danio rerio* reference genome using the Spliced Transcripts Alignment to a Reference (STAR) tool, version 2.7.7 [20]. STAR requires a reference genome and an annotated Gene Transfer File (GTF), which were both obtained from the Ensembl database at https://ftp.ensembl.org/pub/, public release version 101. The reports for each STAR alignment were summarized using MultiQC.

To quantify the mapped reads by gene, the RNA-Seq by Expectation-Maximization (RSEM) tool version 1.3.1 was used [21]. RSEM produces a table of unnormalized read counts, organized by sample and by gene. To account for the stranded nature of the RNA-Seq, the "--strandedness reverse" option was utilized as suggested in Illumina's TruSeq Stranded protocols.

3) *Differential Gene Expression:* The final step in the RCP is to normalize the counts and calculate which genes are differentially expressed between samples. Unlike the previous steps which utilized Linux-based tools, the differential gene expression is done using R's DESeq2 version 1.34.0 R package on R version 4.1.1 [22].

a) *Count Normalization:* The first step utilizes the **estimateSizeFactors()** function, which normalizes the differences in read depth. A size factor is calculated for each gene by dividing the median ratio of all gene counts of a specific gene by the geometric mean of that gene across all samples. Then the raw counts from each sample are divided by the sample-specific size factor for each gene.

Next gene dispersions are calculated using the **estimateDispersions()** function. The variance of each gene's expression across all samples is compared to the mean of that gene's expression. This information can then be plotted in a scatter plot to examine if there is any large variation in the dataset.

b) *Calculation and Ranking of Differential Gene Expression:* Hypothesis testing is done using the **nbinomWaldTest()** function. The goal of hypothesis testing is to calculate how likely the expression of a gene is when compared across two conditions; in this dataset, normal gravity and simulated microgravity. This was done by fitting a negative binomial model to the gene expression data and performing a Wald test to calculate a p-value for each gene between the two conditions.

The last step is to account for the multiple testing problem, which is the increasingly high rate of false positive results that arise from performing an increasingly large number of hypothesis tests using a set False Discovery Rate (FDR). P-values are adjusted using the

Benjamini and Hochberg (BH) method by ranking each gene by p-value, calculating the BH critical value, and comparing each p-value to the BH critical value [23]. The gene with the largest p-value within the threshold (in this case 0.05) that is also smaller than it's BH critical value is set as an adjusted threshold, and all genes ranked higher (smaller p-values) are considered significant and differentially expressed. These normalized counts are then annotated using Ensembl IDs and gene names and saved as a .txt file and .csv file.

C. Gene-Set Enrichment Analysis:

1) *DAVID:* The Database for Annotation, Visualization and Integrated Discovery (DAVID) version 6.8 was used to examine the relationship between genes of interest (those with adjusted p-values below 0.05) [24, 25]. The tool analyzes a given list of genes, for example as by Ensembl ID, and cluster them into gene sets based on the similarity of their functions. The purpose of this analysis is to gain insight into what types of genes are upregulated or downregulated and to look for possible patterns in expression between two experimental conditions. To use the DAVID tool, the annotated file containing the differential gene expression information was sorted by adjusted p-value, and filtered for only those genes with an adjusted p-value within the 0.05 threshold. Then the Ensembl IDs from each gene was saved into a .txt file, and split so that each column had at most 3000 genes. In the DAVID tool, the multi-list file was uploaded with the "Multi-List File" option checked, the Identifier option set to "ENZEMBL_GENE_ID", List Type option set to "Gene List", and Species option set to "*Danio rerio*". The "Functional Annotation Tool" was utilized to perform functional annotation clustering. Within the Functional Annotation Clustering window, all default settings were utilized but the display was modified to include the FDR values in addition to the Benjamini values.

2) *GSEA:* Gene Set Enrichment Analysis (GSEA) version 4.2.3 was also used to analyze the genes that were differentially expressed and calculate which gene sets were statistically significant [26, 27].

To utilize the GSEAPreranked tool, several files need to first be prepared: a rank list, and a gene set database. To make the rank list, the normalized differential expression data was first filtered to exclude all genes that did not fall within the 0.05 adjusted p-value significance threshold. Then the genes were ordered by decreasing fold change. The gene symbols in the Symbols column was capitalized using Microsoft Word, and the Symbols column and Log2fc column (representing the fold change) were saved in a .txt file. Finally, a copy was made with the file type changed to .rnk. For the gene set database, two gene sets were utilized in two separate gene analyses: the built-in [Hallmarks] gene set provided by GSEA, and a *Danio rerio* specific gene set found at http://ge-lab.org/gskb/. For tool settings, the "Collapse/Remap to gene symbols" setting was changed to "No_collapse" and the "Min size: exclude smaller sets" setting was changed to 5. All other options utilized the default settings.

III. RESULTS

A. Quality Control Metrics

Figure 4. Mean Quality Score Comparison between Untrimmed Reads (a) and Trimmed Reads (b)

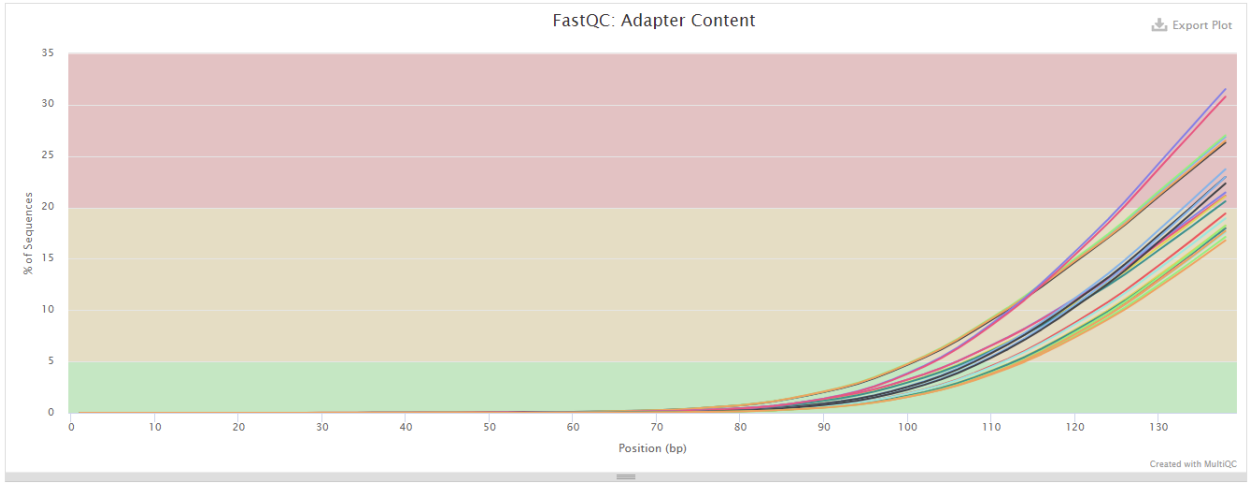

Figure 5. Adapter Content Score of Untrimmed Reads

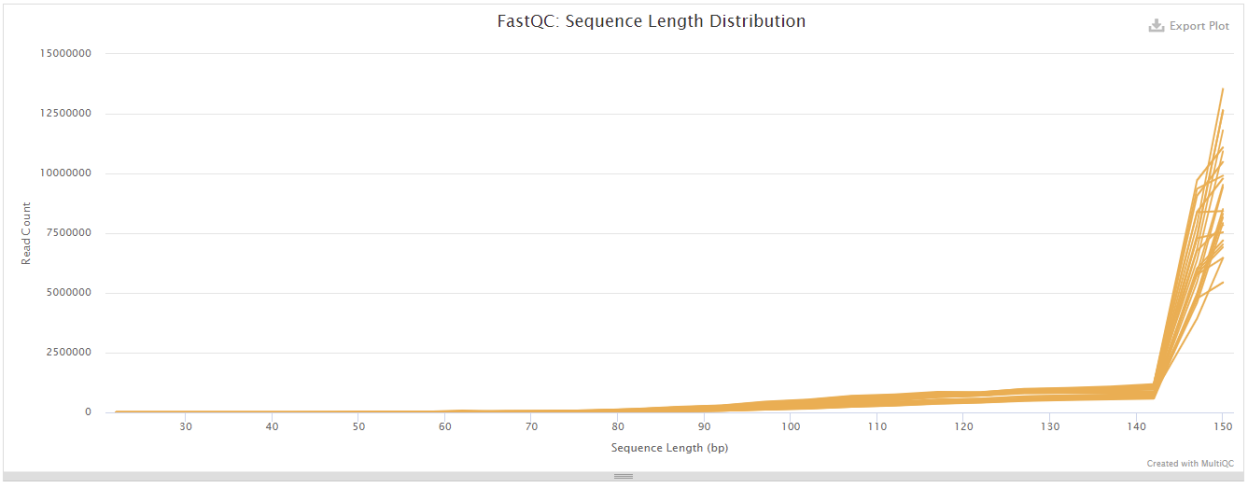

Figure 6. Sequence Length Distribution Plot for Trimmed Reads

Three metrics are shown in Figures 4-6: the average phred score of each base pair for each read by position, the adapter content score, and the sequence length distribution plot. The untrimmed reads show that all base pairs have a phred score of 30 or higher. This indicates that the probability of an incorrect read at a specific base is 1 in 1000 (99.9% accuracy) [28]. In terms of the adapter content score, adapter content was detected in all read samples which will need to be removed. Based on the phred scores, because all reads in all samples scored above 30 the decision was made to only remove adapter sequences. After trimming, we can see in Figure 4. that there is a slight improvement to quality at the end of the read where the adapters were removed and no adapter sequences were detected in the trimmed reads. We can also see in Figure 6 that the removal of the adapter sequences resulted in the sequence length distribution changing from a uniform 150 bp before trimming, to varying lengths between 140-150 bp after trimming.

B. STAR Alignment

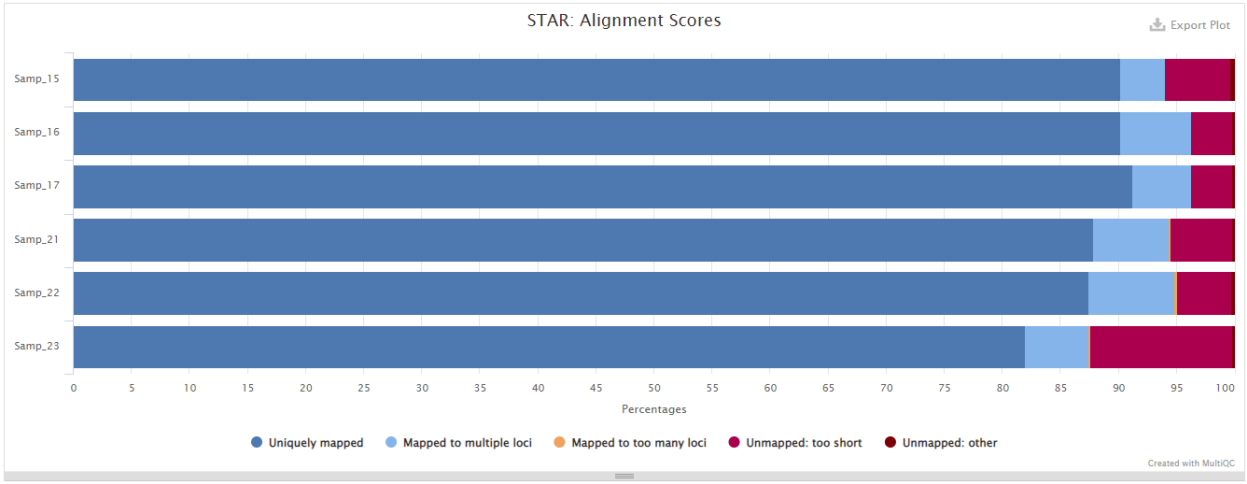

Figure 7. STAR Alignment Scores for All Samples

In Figure 7, the alignment scores for each sample's read against the *Danio rerio* reference genome is shown. Samples 15-17 correspond to the normal gravity control, and samples 21-23 correspond to the simulated microgravity condition. Between 82%-91.2% of all reads were able to be aligned to a unique gene. 3.9%-7.4% of the reads were found to align to multiple loci in the genome, and 3.5%-12.2% of the reads were too short to be accurately aligned to the reference genome. Of the samples, sample 23 showed the worst mapping score, and sample 17 showed the best mapping score.

C. Differential Gene Expression

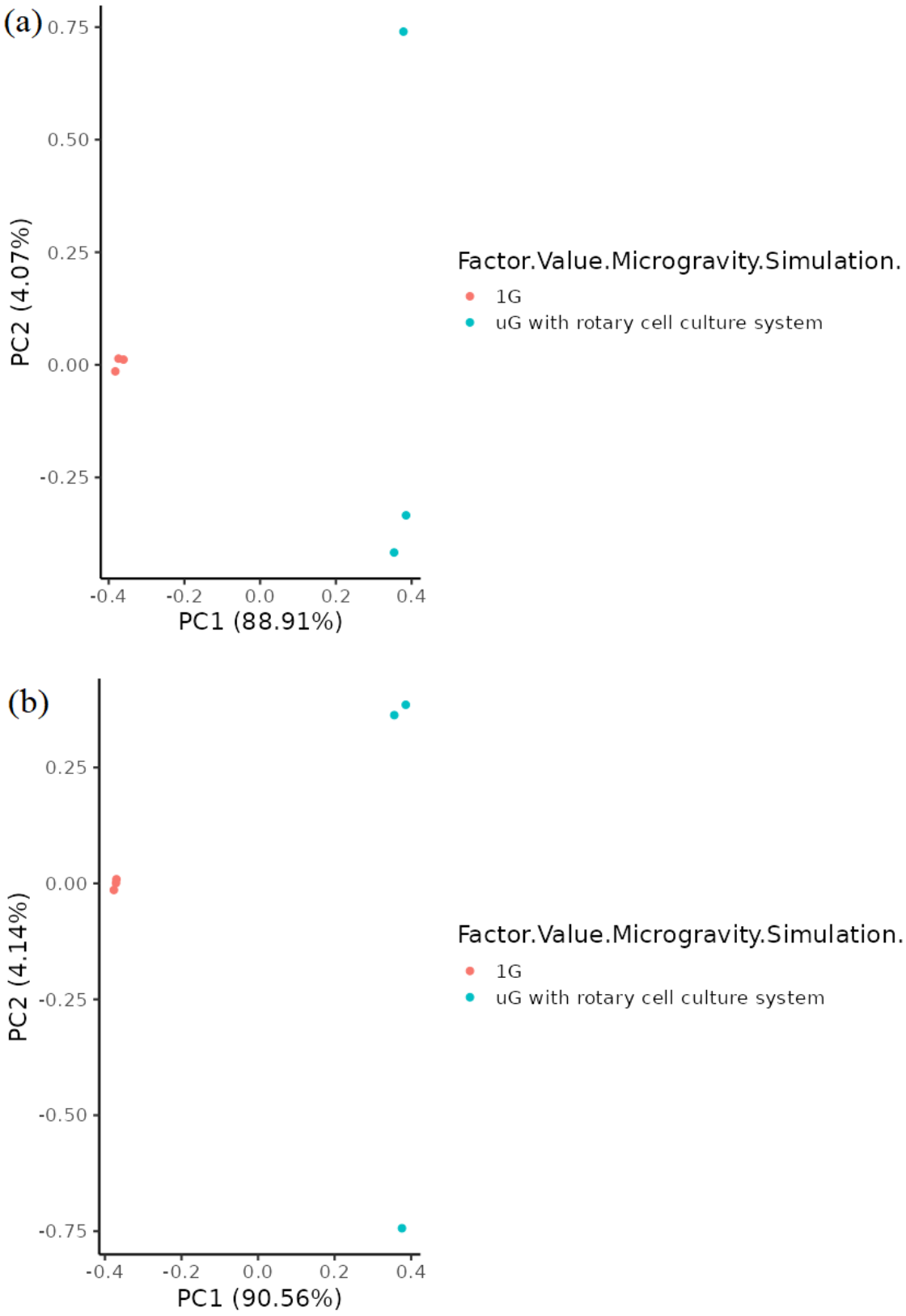

Figure 8. PCA Plots for Unnormalized Counts (a) and Normalized Counts (b)

Principal Analysis Component (PCA) plots simplify understanding of variance across samples and conditions. By employing dimension reduction, many different factors can be reduced into a 2D plot that is easy to grasp visually. In Figure 8, we can see that overall the variance of the samples before and after normalization show mostly the same patterns. The control sample (normal gravity) show little variance and tend to cluster together, shown in Figure 8. in orange. With the simulated microgravity group, one of the three samples shows a large difference in variance compared to the other two. This is shown in Figure 8 where one cyan dot is separate from the other two, either as a very positive PC2 value (before normalization) or a very negative PC2 value (after normalization). The other two samples show little variance between the two, shown by how these two cyan dots cluster together.

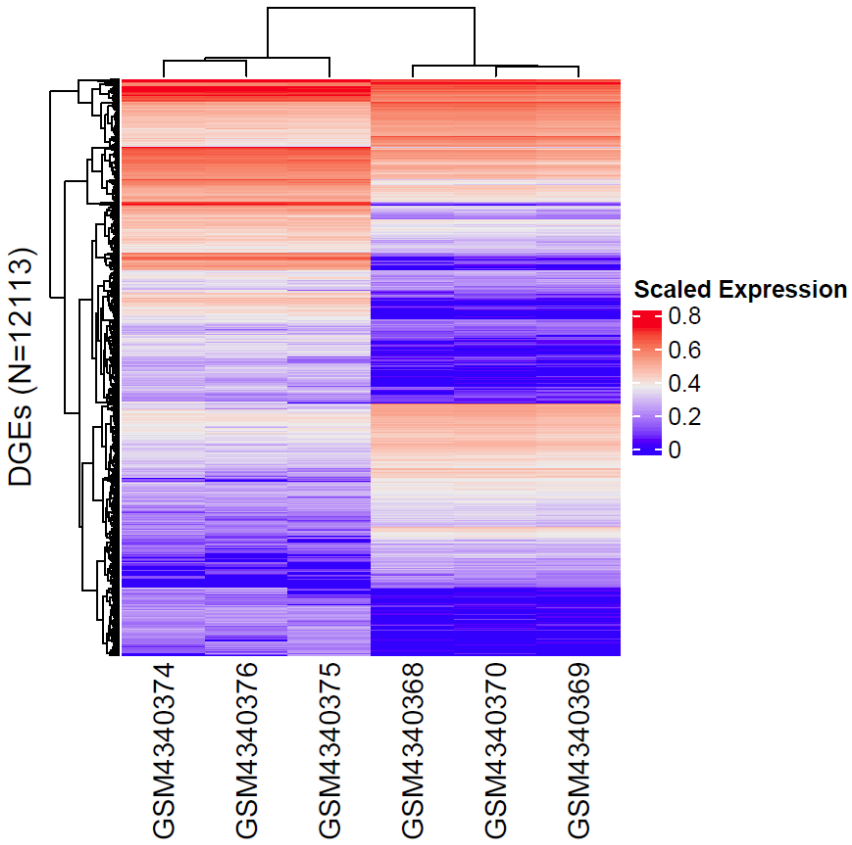

Figure 9. Normal Gravity Control Samples (Left) vs Simulated Microgravity Samples (Right) DEG Heatmap

In Figure 9, a clustered heatmap illustrates the difference in expression between genes of the three samples in the normal gravity control group on the left and the three samples in the simulated microgravity group on the right. Blue represents less expression while red represents more expression. Looking at the heatmap, we can see that there are clear differences between the two groups: there are large clusters of low expression genes in the simulated microgravity group that is not present in the control group, and there are a few regions where genes are slightly more expressed in the simulated microgravity group compared to the control normal gravity group.

D. Gene Set Enrichment Analysis

1) *DAVID:*

TABLE II

Top 5 DAVID Annotation Clusters for Normal Gravity vs. Simulated Microgravity

Category	**Term**	**FDR**
Cluster 1, Enrichment Score 31.29		
UP_KEYWORDS	Ribonucleoprotein	4.65E-42
GOTERM_CC_DIRECT	Intracellular ribonucleoprotein complex	7.44E-42
UP_KEYWORDS	Ribosomal protein	3.59E-33
GOTERM_CC_DIRECT	Ribosome	7.43E-32
KEGG_PATHWAY	Ribosome	8.16E-32
GOTERM_BP_DIRECT	Translation	7.43E-24
GOTERM_MF_DIRECT	Structural constituent of ribosome	1.46E-22
GOTERM_CC_DIRECT	Cytosolic large ribosomal subunit	4.20E-22
GOTERM_CC_DIRECT	Cytosolic small ribosomal subunit	1.12E-16
Cluster 2, Enrichment Score 20.92		
INTERPRO	RNA recognition motif domain	3.90E-19
INTERPRO	Nucleotide-binding, alpha-beta plait	4.42E-18
SMART	RNA Recognition Motif	2.00E-18

Cluster 3, Enrichment Score 11.46		
GOTERM_MF_DIRECT	Structural molecule activity	3.47E-11
UP_KEYWORDS	Intermediate filament	4.71E-12
INTERPRO	Intermediate filament protein	2.50E-10
GOTERM_CC_DIRECT	Intermediate filament	9.79E-11
SMART	SM01391	1.78E-10
INTERPRO	Keratin, type I	1.87E-08
INTERPRO	Intermediate filament protein, conserved site	1.70E-05
Cluster 4, Enrichment Score 10.48		
GOTERM_MF_DIRECT	Nucleic acid binding	9.24E-21
INTERPRO	Zinc finger, C2H2-like	9.94E-08
INTERPRO	Zinc finger C2H2-type/integrase DNA-binding domain	1.06E-07
INTERPRO	Zinc finger, C2H2	1.06E-07
SMART	ZnF_C2H2	2.27E-06
GOTERM_MF_DIRECT	Metal ion binding	0.394326253
Cluster 5, Enrichment Score 10.41		
UP_KEYWORDS	Viral nucleoprotein	7.57E-10
GOTERM_CC_DIRECT	Viral nucleocapsid	6.22E-10
GOTERM_CC_DIRECT	Virion	5.14E-09
UP_KEYWORDS	Virion	2.46E-08

In Table II, the top five DAVID annotation clusters are shown for the normal gravity control vs. the simulated microgravity condition, ordered by enrichment scores. The first cluster has an enrichment score of 31.29 and includes several ribosomal genes. Cluster two has an enrichment score of 20.92 and includes genes related to the recognition and binding of RNA/DNA. Cluster three has an enrichment score of 11.46 and includes genes related with filament proteins, components of the cytoskeleton. Cluster four has an enrichment score of 10.48 and includes genes related to the zinc finger DNA-binding motif. Cluster five has an enrichment score of 10.41 and includes several genes related to viruses and the viral envelope. All genes in all five clusters show high enrichment scores and low FDR values (below 0.05) which suggest

that not only are these results of significant interest, but also that these results are likely not a false discovery.

2) *GSEA:*

TABLE III

GSEA Top 9 Upregulated Hallmark Genes

Name	**ES**	**NES**	**Nom p-val**	**FDR q-val**
HALLMARK_COAGULATION	0.5959	1.7589	0	0.0098
HALLMARK_PANCREAS_BETA_CELLS	0.6155	1.6006	0.0106	0.0783
HALLMARK_KRAS_SIGNALING_DN	0.5067	1.5523	0.0073	0.0960
HALLMARK_ANGIOGENESIS	0.7693	1.5487	0.0137	0.0751
HALLMARK_ESTROGEN_RESPONSE_EARLY	0.4826	1.5370	0.0051	0.0692
HALLMARK_MYOGENESIS	0.4889	1.4908	0.0114	0.1048
HALLMARK_NOTCH_SIGNALING	0.6294	1.4908	0.0528	0.1118
HALLMARK_EPITHELIAL_MESENCHYMAL_TRANSITION	0.4655	1.4514	0.0124	0.1250
HALLMARK_APICAL_SURFACE	0.6410	1.3857	0.0787	0.2200

For the hallmark gene set's gene analysis, there were a total of 42 out of 50 gene sets that were upregulated in simulated microgravity compared to normal gravity. However only 9 of these gene sets fell within the chosen significance thresholds (nominal p-val below 0.05, FDR below 0.25), shown in Table III. Of the 8 gene sets that were downregulated, none fell below the FDR threshold of 0.25. The *Danio rerio* gene set's gene analysis found 606 out of 840 gene sets that were upregulated in simulated microgravity compared to normal gravity. However only 158 of these gene sets fell within the significance thresholds (nominal p-val below 0.05, FDR below 0.25), shown in the Appendix section, Table IV. Of the 234 gene sets that were downregulated, none fell within the chosen FDR threshold of 0.25.

IV. DISCUSSION

A. RCP Implementation

Implementation of the RCP via Jupyter Notebook is a good way to learn the tools and processes which make up the RCP, but several challenges call for a more streamlined approach. In the case for this project, the RCP was already developed in Jupyter notebook, and adapted to process the GLDS-373 dataset. One important factor to take note of is that the HPC at SJSU requires a SLURM script to properly allocate resources and efficiently perform computational tasks. This necessitates most jobs be written in a separate SLURM script that is not included within the Jupyter notebook; instead, the notebook only calls the script to run each step of the RCP. Care must be taken to ensure the pipeline is able to run smoothly, as several steps produce output files that are required as inputs to subsequent steps in the RCP. Additionally, the HPC will not alert the user to failed jobs, necessitating frequent manual checks for errors or unintended outputs. This requires significant user time and effort, which runs counter to the goal of a smooth automated pipeline. Future efforts might include background software processes that monitor for errors and alert the user in a timely manner via email or text message. They may even include a method of automation to run new scripts after a previous step has completed.

B. Gene Enrichment Analysis

1) *DAVID Annotated Gene Clusters:* Cluster one is related to ribosomal functions, and clusters two and four are related to the binding of DNA/RNA. Ribosomes are vital to translation, where they synthesize new proteins from messenger RNA (mRNA). Once a ribosome complex binds to a strand of mRNA, small tRNAs containing complimentary anti-codons and a single amino acid bind to the mRNA and consecutive tRNAs chain amino acids together into a protein

chain [29]. Zinc finger proteins, and specifically the C_2H_2 finger proteins play important roles in the binding of DNA or RNA segments [30]. Together, clusters one, two, and four suggest that microgravity affects the binding of DNA/RNA and overall translation.

Cluster three involves genes related to intermediate filament proteins (IF-proteins). IF-proteins are proteins that form many of the filaments that assemble into more complex cytoskeletal structures [31]. These proteins are also involved in organ development and tissue differentiation, suggesting that microgravity can affect the tissue differentiation of an early embryo. Cluster five was surprising since it involved many genes related to virions. This suggests that there may be some contamination in the RNA collected from the embryo samples.

2) *GSEA:* In the GSEA analysis, nine hallmark gene sets were found to be upregulated and of significant interest while still falling within the chosen 0.25 FDR threshold. These nine gene sets include genes related to coagulation, pancreas beta cells, angiogenesis, early estrogen response, kras signaling, myogenesis, notch signaling, apical surface and genes involved in epithelial and mesenchymal differentiation. None of these gene sets matches directly with the clusters found in the DAVID analysis.

a) *Hallmark Gene Set Analysis:* Kras, part of the Ras protein family, is a signaling protein that mediates nuclear transcription factors with extracellular signals [32]. When activated, Ras signaling can induce gene expression in genes related to cell proliferation, differentiation, and apoptosis. Angiogenesis genes are upregulated as well, which supports the idea that cells are rapidly proliferating. Angiogenesis refers to the process by which new blood vessels form from existing vessels [33]. As blood vessels provide the necessary nutrients for cell growth, if cells are rapidly proliferating new blood vessels would be needed to maintain cell development. Coagulation genes are also expressed early in development [34]. During

embryonic development, coagulation genes perform functions related to cellular proliferation and differentiation. For example, the absence of coagulation proteins tissue factor (TF), TF pathway inhibitor (TFPI), and prothrombin is lethal for developing mouse embryos.

Two of the results: notch signaling and epithelial-mesenchymal transition (EMT) are directly related to processes involving cell fate and differentiation. Notch signaling in embryos plays key roles in the development of embryos where embryonic stem cells differentiate into more specialized cell types [35]. For example, the knockout of the notch1 receptor in mouse embryos is embryonically lethal. In comparison, knockouts of the notch2 and notch3 receptors does not result in embryo death but several deformities do arise, indicating that notch signaling is a key factor in proper embryonic development. The epithelial-mesenchymal transition is a process by which epithelial cells transition into mesenchymal cells [36]. In embryogenesis, the EMT is activated in the gastrulation phase where it mediates the formation and differentiation of the mesoderm and endoderm layers from the mesendoderm layer. This is the first step in determining cell fate, where cells from each layer will undergo further differentiation into more specialized cell types and tissues. EMT is also responsible for the transition of epithelial cells in the neuroectoderm into migratory neural crest cells. These neural crest cells have been observed to migrate to other parts of the embryo, where they then undergo further differentiation to give rise to other embryonic structures. The upregulation of genes involved in notch signaling and EMT suggest that microgravity may stimulate genes associated to cell differentiation, which can affect the proper development of embryos.

Several of the results are related to the development of embryonic tissues and structures. Estrogen is a hormone that helps control the development of reproductive organs and sex differentiation in early embryos [37]. As genes related to responding to estrogen are upregulated

in stimulated microgravity, this could mean that the embryo is more sensitive to estrogen or that tissue and organ development is stimulated in low gravity conditions. Genes related to the development of pancreatic beta cells, myogenesis (the development of skeletal muscle), and those encoding proteins found on the apical surface of cells (responsible for cell polarity) are also upregulated, indicating that microgravity may stimulate early development of various structures.

b) *Danio rerio Gene Set Analysis:* Of the 129 gene sets that were found to be of significant interest and fell within the FDR threshold, most were found to be involved in cell differentiation and embryo development. Examples include gene sets involved in the development of key organs such as the liver and brain, gene sets involved in key stages of the embryonic development process such as mesoderm development, determination of cell fate and left/right parts of the body, and several important development-related receptor signaling pathways such as Wnt and G-protein signaling. While nothing abnormal stands out at first glance, it is interesting that so many development-related genes are upregulated in microgravity conditions. It is currently unknown what effects upregulation of so many embryonic developmental genes may have. Future studies may individually examine the specific morphological and physiological effects that may occur when one or a small number of these developmental genes are upregulated in early-stage embryos.

V. CONCLUSION

This project aimed to provide an efficient implementation of the NASA Genelab RCP to analyze data from the GLDS-373 dataset, which could provide insight into how microgravity conditions may affect the development of early embryos. Overall, the use of Jupyter Notebook to

implement the pipeline provided a good start but there is room for a more streamlined approach to automate more of the process and save user time and effort.

This project utilized mRNA data collected from zebrafish embryos subjected to simulated microgravity via a rotary system. While this is not true microgravity, there was still a clear difference in expression patterns between the normal gravity control and the simulated gravity groups. This shows that the rotary system can simulate microgravity to some degree and an analysis of differentially expressed genes can provide profound insights into the role microgravity can play upon early-stage embryos.

In terms of the results gained, two gene enrichment analysis methods were used: DAVID and GSEA. What is disappointing is that there were no downregulated gene sets that passed both the significance threshold of 0.05 for p-values, and the FDR threshold of 0.25. For both DAVID and GSEA, in this project nearly all options utilized the default settings. For future analyses, these default settings can be changed to see if more results will fall within the significance and FDR thresholds. The thresholds themselves can also be changed to be less stringent, to allow more results through.

From the results, it was found that microgravity does result in several changes to expression patterns. Of particular interest would be genes related to DNA transcription and translation, and genes that regulate early cell proliferation, cell differentiation, and early development of embryonic tissues and organs. These genes were found to be upregulated in embryos exposed to simulated microgravity. The exact impact that microgravity may have on embryonic development is still not well understood, and more research is still required before a mechanism to explain these observations can be identified. Hopefully the findings from this

study can provide some insight and help guide future studies. Most importantly, these findings may help in protecting the health of future astronauts and their children.

REFERENCES

[1] M. Smith et al., “The artemis program: An overview of nasa’s activities to return humans to the moon,” 2020, pp. 1–10.

[2] Z. S. Patel et al., “Red risks for a journey to the red planet: The highest priority human health risks for a mission to Mars,” vol. 6, no. 1, pp. 1–13, 2020.

[3] X. Lei, Y. Cao, Y. Zhang, and E. Duan, “Advances of mammalian reproduction and embryonic development under microgravity,” pp. 281–315, 2019.

[4] J. Abramczuk and W. Sawicki, “Pronuclear synthesis of DNA in fertilized and parthenogenetically activated mouse eggs: a cytophotometric study,” vol. 92, no. 2, pp. 361–371, 1975.

[5] L. Larue, M. Ohsugi, J. Hirchenhain, and R. Kemler, “E-cadherin null mutant embryos fail to form a trophectoderm epithelium,” vol. 91, no. 17, pp. 8263–8267, 1994.

[6] A. K. Tarkowski and J. Wróblewska, “Development of blastomeres of mouse eggs isolated at the 4-and 8-cell stage,” 1967.

[7] L. Li, N. Gu, H. Dong, B. Li, and T. Kenneth, “Analysis of the effects of acoustic levitation to simulate the microgravity environment on the development of early zebrafish embryos,” vol. 10, no. 72, pp. 44593–44600, 2020.

[8] B. Mishra and U. Luderer, “Reproductive hazards of space travel in women and men,” vol. 15, no. 12, pp. 713–730, 2019.

[9] S. Wakayama, Y. Kawahara, C. Li, K. Yamagata, L. Yuge, and T. Wakayama, “Detrimental effects of microgravity on mouse preimplantation development in vitro,” vol. 4, no. 8, p. e6753, 2009.

[10] X. Lei et al., “Development of mouse preimplantation embryos in space,” vol. 7, no. 9, pp. 1437–1446, 2020.

[11] Y. Kojima, S. Sasaki, Y. Kubota, T. Ikeuchi, Y. Hayashi, and K. Kohri, “Effects of simulated microgravity on mammalian fertilization and preimplantation embryonic development in vitro,” vol. 74, no. 6, pp. 1142–1147, 2000.

[12] E. G. Overbey et al., “NASA GeneLab RNA-seq consensus pipeline: standardized processing of short-read RNA-seq data,” vol. 24, no. 4, p. 102361, 2021.

[13] L. Zhu et al., “Attenuation of Antiviral Immune Response Caused by Perturbation of TRIM25-Mediated RIG-I Activation under Simulated Microgravity,” vol. 34, no. 1, p. 108600, 2021.

[14] “SJSU’s College of Engineering’s High Performance Computing Cluster Details.” http://coe-hpc-web.sjsu.edu/ (accessed Apr. 11, 2022).

[15] “SJSU’s College of Science’s High Performance Computing Cluster.” https://www.sjsu.edu/wildfire/facilities/modeling.php (accessed Apr. 11, 2022).

[16] “JupyterHub Brief Introduction.” https://jupyter.org/hub (accessed Apr. 11, 2022).

[17] S. Andrews, “FastQC: a quality control tool for high throughput sequence data,” 2010.

[18] P. Ewels, M. Magnusson, S. Lundin, and M. Käller, “MultiQC: summarize analysis results for multiple tools and samples in a single report,” vol. 32, no. 19, pp. 3047–3048, 2016.

[19] F. Krueger, “Trim galore,” vol. 516, no. 517, 2015.

[20] A. Dobin et al., “STAR: ultrafast universal RNA-seq aligner,” vol. 29, no. 1, pp. 15–21, 2013.

[21] B. Li and C. N. Dewey, “RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome,” vol. 12, no. 1, pp. 1–16, 2011.

[22] M. I. Love, W. Huber, and S. Anders, “Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2,” vol. 15, no. 12, pp. 1–21, 2014.

[23] Y. Benjamini and Y. Hochberg, “Controlling the false discovery rate: a practical and powerful approach to multiple testing,” vol. 57, no. 1, pp. 289–300, 1995.

[24] D. W. Huang, B. T. Sherman, and R. A. Lempicki, “Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources,” vol. 4, no. 1, pp. 44–57, 2009.

[25] D. W. Huang, B. T. Sherman, and R. A. Lempicki, “Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists,” vol. 37, no. 1, pp. 1–13, 2009.

[26] A. Subramanian et al., “Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles,” vol. 102, no. 43, pp. 15545–15550, 2005.

[27] V. K. Mootha et al., “PGC-1α-responsive genes involved in oxidative phosphorylation are coordinately downregulated in human diabetes,” vol. 34, no. 3, pp. 267–273, 2003.

[28] B. Ewing and P. Green, “Base-calling of automated sequencer traces using phred. II. Error probabilities,” vol. 8, no. 3, pp. 186–194, 1998.

[29] R. Green and H. F. Noller, “Ribosomes and translation,” vol. 66, no. 1, pp. 679–716, 1997.

[30] S. Iuchi, “Three classes of C2H2 zinc finger proteins,” vol. 58, no. 4, pp. 625–635, 2001.

[31] K. Albers and E. Fuchs, “The molecular biology of intermediate filament proteins,” vol. 134, pp. 243–279, 1992.

[32] P. Castagnola and W. Giaretti, “Mutant KRAS, chromosomal instability and prognosis in colorectal cancer,” vol. 1756, no. 2, pp. 115–125, 2005.

[33] A. Schuermann, C. S. Helker, and W. Herzog, “Angiogenesis in zebrafish,” 2014, vol. 31, pp. 106–114.

[34] M. J. Manco-Johnson, “Development of hemostasis in the fetus.,” vol. 115, pp. 55–63, 2005.

[35] H. L. Ng, E. Quail, M. N. Cruickshank, and D. Ulgiati, “To Be, or Notch to Be: Mediating Cell Fate from Embryogenesis to Lymphopoiesis,” vol. 11, no. 6, p. 849, 2021.

[36] R. Kalluri and R. A. Weinberg, “The basics of epithelial-mesenchymal transition,” vol. 119, no. 6, pp. 1420–1428, 2009.

[37] M. Bondesson, R. Hao, C.-Y. Lin, C. Williams, and J.-Å. Gustafsson, “Estrogen receptor signaling during vertebrate development,” vol. 1849, no. 2, pp. 142–151, 2015.

APPENDIX

TABLE IV

GSEA Top Danio rerio Upregulated Genes

NAME	ES	NES	NOM p-val	FDR q-val
GO_BP_DR_ANTERIOR_POSTERIOR_PATTERN_SPECIFICATION	0.759 73463	2.197 3498	0	0
GO_CC_DR_EXTRACELLULAR_SPACE	0.703 5055	2.108 2506	0	0
GO_MF_DR_LIPID_BINDING	0.770 6552	2.042 3787	0	0
GO_CC_DR_EXTRACELLULAR_REGION	0.613 14183	2.034 4892	0	0
GO_BP_DR_FIN_REGENERATION	0.749 9587	1.989 9322	0	2.98E-04
GO_BP_DR_OTIC_PLACODE_FORMATION	0.820 1317	1.995 718	0	3.58E-04
GO_CC_DR_MYOSIN_FILAMENT	0.878 67767	1.973 6123	0	5.07E-04
GO_BP_DR_DETERMINATION_LEFT_RIGHT_SYMMETRY	0.648 9403	1.911 9099	0	0.0020 08484
GO_CC_DR_INTERMEDIATE_FILAMENT	0.696 1885	1.885 4526	0	0.0028 20956
GO_BP_DR_PHARYNGEAL_SYSTEM_DEVELOPMENT	0.885 54484	1.892 4813	0	0.0029 42649
GO_BP_DR_HINDBRAIN_DEVELOPMENT	0.679 4562	1.888 1744	0	0.0030 77407
GO_BP_DR_LIVER_DEVELOPMENT	0.682 1263	1.892 543	0	0.0032 6961
GO_BP_DR_DORSAL_VENTRAL_PATTERN_FORMATION	0.607 0417	1.867 3452	0	0.0035 68135
GO_BP_DR_FOREBRAIN_DEVELOPMENT	0.716 9812	1.859 5748	0.0011 52074	0.0041 35086
GO_BP_DR_MESODERM_DEVELOPMENT	0.789 2168	1.851 3047	0	0.0042 85076
GO_BP_DR_SOMITOGENESIS	0.623 9047	1.855 0582	0	0.0043 349
GO_CC_DR_COLLAGEN	0.662 14687	1.839 3908	0	0.0052 8497
GO_MF_DR_STRUCTURAL_MOLECULE_ACTIVITY	0.582	1.833	0	0.0059

	972	0419		29986
GO_MF_DR_CALCIUM-DEPENDENT_PHOSPHOLIPID_BINDING	0.781 3221	1.821 8936	0	0.0063 801
GO_BP_DR_VASCULATURE_DEVELOPMENT	0.689 35317	1.819 6349	0.0011 45475	0.0064 88467
GO_MF_DR_TRANSMEMBRANE_SIGNALING_RECEPTOR_ACTIVITY	0.648 8264	1.815 6592	0	0.0065 14334
GO_BP_DR_WNT_RECEPTOR_SIGNALING_PATHWAY	0.600 8597	1.826 2619	0	0.0065 47287
GO_MF_DR_RECEPTOR_BINDING	0.619 8701	1.822 2198	0	0.0065 5804
GO_MF_DR_WNT-PROTEIN_BINDING	0.712 53777	1.807 7493	0.0012 13592	0.0066 87156
GO_BP_DR_PANCREAS_DEVELOPMENT	0.697 0568	1.809 3293	0.0011 77856	0.0067 01622
GO_CC_DR_PROTEINACEOUS_EXTRACELLULAR_MATRIX	0.656 13216	1.823 1791	0.0011 12347	0.0067 53042
GO_MF_DR_OXIDOREDUCTASE_PAIRED_DONORS_MOLECULAR_OXYGEN	0.602 75257	1.800 8167	0	0.0068 11359
GO_BP_DR_SURFACE_RECEPTOR_SIGNALING_PATHWAY	0.641 6935	1.797 9599	0.0022 19756	0.0068 79141
GO_MF_DR_PDZ_DOMAIN_BINDING	0.701 53075	1.809 5677	0	0.0069 34248
GO_MF_DR_G-PROTEIN_COUPLED_RECEPTOR_ACTIVITY	0.573 02684	1.800 9905	0	0.0070 54621
GO_BP_DR_RETINAL_GANGLION_AXON_GUIDANCE	0.670 7204	1.789 0759	0	0.0081 79555
GO_MF_DR_WNT-ACTIVATED_RECEPTOR_ACTIVITY	0.712 53777	1.780 3067	0	0.0096 80965
GO_BP_DR_MUSCLE_ORGAN_DEVELOPMENT	0.673 2071	1.771 6115	0.0011 75088	0.0108 1164
GO_BP_DR_TAIL_MORPHOGENESIS	0.780 35253	1.772 3334	0.0026 45503	0.0109 78187
GO_CC_DR_PLASMA_MEMBRANE	0.533 57583	1.763 9987	0	0.0122 3137
GO_BP_DR_BRAIN_DEVELOPMENT	0.579 4218	1.755 5712	0	0.0139 04027
GO_CC_DR_MYOSIN_COMPLEX	0.606 7722	1.749 4786	0.0010 9529	0.0156 16017
GO_BP_DR_MIGRATION_GASTRULATION	0.572 33596	1.737 9303	0	0.0168 38286

GO_BP_DR_STRIATED_MUSCLE_DEVELOPMENT	0.722 43863	1.739 1322	0.0038 51091	0.0168 99023
GO_BP_DR_POSITIVE_REGULATION_MIGRATION	0.907 6989	1.740 235	0	0.0170 28205
GO_MF_DR_GROWTH_FACTOR_ACTIVITY	0.581 78264	1.735 096	0	0.0171 5248
GO_MF_DR_IRON_ION_BINDING	0.545 74573	1.740 7709	0	0.0172 7967
GO_BP_DR_NEGATIVE_REGULATION_PROLIFERATION	0.807 2625	1.741 6197	0.0013 6612	0.0174 06179
GO_BP_DR_HEART_MORPHOGENESIS	0.657 4302	1.728 9448	0.0023 8379	0.0184 95917
GO_BP_DR_NEGATIVE_REGULATION_CANONICAL_WNT_RECEPTOR_SIGNALING_PATHWAY	0.747 20675	1.726 2726	0.0025 64103	0.0188 75003
GO_BP_DR_MULTICELLULAR_ORGANISMAL_DEVELOPMENT	0.511 9817	1.725 1107	0	0.0189 5077
GO_MF_DR_MOTOR_ACTIVITY	0.593 6541	1.720 6435	0.0032 43243	0.0198 1835
GO_BP_DR_GONAD_DEVELOPMENT	0.677 11747	1.714 8366	0.0011 99041	0.0214 15738
GO_MF_DR_PROTEIN_DIMERIZATION_ACTIVITY	0.517 106	1.707 2706	0	0.0224 9533
GO_MF_DR_CALCIUM_ION_BINDING	0.503 6495	1.708 9409	0	0.0227 77012
GO_BP_DR_EMBRYONIC_VISCEROCRANIUM_MORPHOGENESIS	0.699 37146	1.704 8914	0.0048 84005	0.0228 3042
GO_CC_DR_TIGHT_JUNCTION	0.619 4137	1.707 4298	0	0.0228 74095
GO_BP_DR_NEGATIVE_REGULATION_NEUROGENESIS	0.786 4737	1.702 6192	0.0041 66667	0.0232 4214
GO_MF_DR_MONOOXYGENASE_ACTIVITY	0.555 76897	1.698 8525	0	0.0240 6329
GO_MF_DR_EXTRACELLULAR_MATRIX_STRUCTURAL_CONSTITUENT	0.673 0703	1.695 2782	0.0024 0096	0.0240 9276
GO_MF_DR_SEQUENCE-SPECIFIC_DNA_BINDING_TRANSCRIPTION_FACTOR_ACTIVITY	0.500 009	1.697 7614	0	0.0241 27811
GO_BP_DR_NEURON_DIFFERENTIATION	0.720 18355	1.696 1164	0.0063 37136	0.0241 40375
GO_MF_DR_CYSTEINE-TYPE_PEPTIDASE_ACTIVITY	0.556 0248	1.693 0723	0.0010 54852	0.0246 14798
GO_MF_DR_HEME_BINDING	0.543	1.690	0	0.0252

	20323	8098		4286
GO_BP_DR_MYOFIBRIL_ASSEMBLY	0.772 5144	1.689 5777	0.0078 02341	0.0253 88163
GO_BP_DR_SEMICIRCULAR_CANAL_MORPHOGENESIS	0.827 74526	1.684 6633	0.0042 19409	0.0270 02608
GO_BP_DR_NEURAL_TUBE_FORMATION	0.822 5837	1.680 51	0.0028 86003	0.0279 81786
GO_BP_DR_FATE_SPECIFICATION	0.872 06244	1.681 0057	0	0.0282 18564
GO_BP_DR_NOTOCHORD_DEVELOPMENT2	0.637 92515	1.677 2604	0.0023 50176	0.0290 61854
GO_BP_DR_DORSAL_CONVERGENCE	0.837 80414	1.676 033	0	0.0291 75116
GO_BP_DR_HOMOPHILIC_ADHESION	0.584 7218	1.669 9959	0.0021 92983	0.0313 25452
GO_BP_DR_CANONICAL_WNT_RECEPTOR_SIGNALING_PATHWAY	0.605 29643	1.656 801	0.0022 34637	0.0353 59677
GO_MF_DR_SECONDARY_ACTIVE_SULFATE_TRANSMEMBRANE_TRANSPORTER_ACTIVITY	0.807 2892	1.657 3467	0.0071 02273	0.0354 95386
GO_BP_DR_NEGATIVE_REGULATION_CATALYTIC_ACTIVITY	0.751 7826	1.660 1708	0.0039 94674	0.0357 27505
GO_BP_DR_LIPID_TRANSPORT	0.626 88535	1.657 7461	0.0057 60369	0.0359 0663
GO_BP_DR_G-PROTEIN_COUPLED_RECEPTOR_SIGNALING_PATHWAY	0.511 58696	1.658 2955	0	0.0361 1924
GO_BP_DR_RESPONSE_TO_CHEMICAL_STIMULUS	0.573 6075	1.650 0806	0.0032 6087	0.0385 74822
GO_BP_DR_NEGATIVE_DNA-DEPENDENT	0.563 7141	1.648 0579	0.0043 14994	0.0390 10506
GO_BP_DR_ADHESION	0.515 5052	1.643 0173	0.0010 15229	0.0407 27224
GO_BP_DR_MESODERM_FORMATION	0.803 24596	1.643 1787	0.0013 60544	0.0412 1596
GO_BP_DR_ORGAN_MORPHOGENESIS	0.743 6	1.640 719	0.0078 63696	0.0414 3462
GO_BP_DR_PERIPHERAL_NERVOUS_SYSTEM_NEURON_AXONOGENESIS	0.722 37927	1.636 3125	0.0153 06123	0.0431 96235
GO_MF_DR_TRANSPORTER_ACTIVITY	0.518 7727	1.629 08	0.0010 15229	0.0445 19182
GO_BP_DR_SULFATE_TRANSPORT	0.807 2892	1.629 9291	0.0041 60888	0.0445 9714

GO_BP_DR_NEURON_FATE_SPECIFICATION	0.8000691	1.6301423	0.004126547	0.045087617
GO_BP_DR_ADENOHYPOPHYSIS_DEVELOPMENT	0.74508953	1.6314975	0.007968128	0.04545656
GO_BP_DR_NEURON_FATE_COMMITMENT	0.84208596	1.6303514	0.012729844	0.045567803
GO_BP_DR_NEGATIVE_REGULATION_ENDODERMAL_FATE_SPECIFICATION	0.7891458	1.6249622	0.008571428	0.046632845
GO_MF_DR_ACTIN_BINDING	0.50285953	1.623375	0	0.047179356
GO_BP_DR_EPIDERMIS_DEVELOPMENT	0.83421934	1.6212374	0.004511278	0.047712635
GO_MF_DR_SEQUENCE-SPECIFIC_DNA_BINDING	0.47643518	1.6150775	0	0.051326156
GO_MF_DR_SULFATE_TRANSMEMBRANE_TRANSPORTER_ACTIVITY	0.8072892	1.6123066	0.006657779	0.051737722
GO_BP_DR_ECTODERMAL_PLACODE_FORMATION	0.83346033	1.6132418	0.007082153	0.051820412
GO_BP_DR_EMBRYONIC_HEART_TUBE_MORPHOGENESIS	0.70261896	1.607267	0.016927084	0.053405218
GO_CC_DR_TRANSCRIPTION_FACTOR_COMPLEX	0.5171743	1.6078295	0.002047083	0.053633307
GO_BP_DR_FIN_DEVELOPMENT	0.721082	1.6079895	0.008905852	0.054115243
GO_BP_DR_ENDODERM_DEVELOPMENT	0.7052368	1.6030213	0.019430052	0.055858191
GO_BP_DR_KUPFFERS_VESICLE_DEVELOPMENT	0.5999185	1.5990447	0.007945516	0.058333063
GO_BP_DR_DETERMINATION_VENTRAL_IDENTITY	0.65692514	1.5912955	0.006226665	0.062944715
GO_BP_DR_MESODERMAL_MIGRATION	0.7666604	1.5913961	0.009602195	0.063522261
GO_BP_DR_NEGATIVE_REGULATION_ENDOPEPTIDASE_ACTIVITY	0.55006486	1.5895659	0.009708738	0.063633461
GO_MF_DR_HORMONE_ACTIVITY	0.64860964	1.5850556	0.015719468	0.066483475
GO_MF_DR_SERINE-TYPE_ENDOPEPTIDASE_ACTIVITY	0.5208878	1.5799947	0.001042753	0.069755106
GO_BP_DR_CARDIOBLAST_DIFFERENTIATION	0.8100175	1.5785416	0.004267425	0.070337781
GO_BP_DR_MESENDODERM_DEVELOPMENT	0.812	1.576	0.0102	0.0704

	21443	5706	48902	92834
GO_BP_DR_DIENCEPHALON_DEVELOPMENT	0.769 12725	1.576 9774	0.0142 85714	0.0709 04054
GO_BP_DR_SOMITE_SPECIFICATION	0.778 96017	1.574 0733	0.0150 27323	0.0712 3889
GO_CC_DR_EXTRACELLULAR_MATRIX	0.684 92544	1.574 7942	0.0189 63337	0.0713 1008
GO_BP_DR_MIDBRAIN_DEVELOPMENT	0.713 04274	1.572 3867	0.0200 53476	0.0717 6681
GO_BP_DR_DETERMINATION_HEART_LEFT_RIGHT_ASYMMETRY	0.776 74526	1.568 8449	0.0113 31445	0.0740 15945
GO_BP_DR_GROWTH2	0.616 2375	1.567 848	0.0196 07844	0.0741 7935
GO_CC_DR_JUNCTION	0.515 5523	1.561 1818	0.0062 5	0.0793 9391
GO_BP_DR_EXOCRINE_PANCREAS_DEVELOPMENT	0.650 8967	1.560 3876	0.0138 71375	0.0794 9155
GO_BP_DR_HEART_LOOPING	0.539 36064	1.557 6826	0.0097 61388	0.0803 6121
GO_BP_DR_PEPTIDE_CROSS-LINKING	0.679 60715	1.557 9355	0.0233 46303	0.0808 1194
GO_MF_DR_SERINE-TYPE_PEPTIDASE_ACTIVITY	0.530 5699	1.555 8285	0.0107 75862	0.0810 00015
GO_BP_DR_PERIPHERAL_NERVOUS_SYSTEM_DEVELOPMENT	0.694 4112	1.553 8484	0.0164 34893	0.0819 5536
GO_BP_DR_PRONEPHROS_DEVELOPMENT	0.571 7795	1.550 587	0.0181 4059	0.0843 8701
GO_BP_DR_REGULATION_PROTEOLYSIS	0.686 44476	1.548 5213	0.0155 64202	0.0856 0546
GO_CC_DR_RIBONUCLEOPROTEIN_COMPLEX	0.470 01404	1.543 6807	0	0.0861 81976
GO_BP_DR_CELL-SIGNALING	0.773 8752	1.544 8772	0.0147 49263	0.0866 4375
GO_MF_DR_CALCIUM-DEPENDENT_CYSTEINE-TYPE_ENDOPEPTIDASE_ACTIVITY	0.755 4311	1.543 7883	0.0204 63847	0.0867 31285
GO_BP_DR_ENDOCRINE_PANCREAS_DEVELOPMENT	0.647 747	1.545 1759	0.0280 4878	0.0871 13015
GO_MF_DR_SERINE-TYPE_ENDOPEPTIDASE_INHIBITOR_ACTIVITY	0.550 3454	1.545 645	0.0257 27069	0.0874 5969
GO_MF_DR_PROTEIN-GLUTAMINE_GAMMA-GLUTAMYLTRANSFERASE_ACTIVITY	0.679 60715	1.531 0082	0.0317 46034	0.0952 661

GO_BP_DR_POSITIVE_REGULATION_DIVISION	0.800 08644	1.529 157	0.0190 05848	0.0955 64716
GO_BP_DR_OTIC_VESICLE_FORMATION	0.753 4704	1.529 7111	0.0218 28104	0.0957 8175
GO_BP_DR_POSITIVE_REGULATION_NEURON_DIFFERENTIATION	0.760 40804	1.531 1631	0.0100 28653	0.0958 7833
GO_CC_DR_GOLGI_MEMBRANE	0.517 9528	1.528 1621	0.0181 62394	0.0959 3961
GO_BP_DR_NEUROMAST_DEVELOPMENT	0.723 331	1.531 5328	0.0163 2653	0.0962 2943
GO_BP_DR_AXON_GUIDANCE	0.578 8969	1.532 4433	0.0244 75524	0.0962 3938
GO_BP_DR_LEFT_RIGHT_PATTERN_FORMATION	0.747 37537	1.525 8592	0.0214 89972	0.0973 61945
GO_BP_DR_PATTERN_SPECIFICATION_PROCESS	0.579 65326	1.524 0495	0.0252 00458	0.0984 6346
GO_BP_DR_SMOOTHENED_SIGNALING_PATHWAY	0.779 8529	1.518 3779	0.0160 34985	0.1030 0452
GO_BP_DR_MYOBLAST_FUSION	0.778 94545	1.518 772	0.0246 7344	0.1032 86035
GO_MF_DR_METALLOCARBOXYPEPTIDASE_ACTIVITY	0.710 975	1.516 4022	0.0288 06584	0.1042 7034
GO_BP_DR_POSITIVE_FROM_RNA_POLYMERASE_II_PROMOTER	0.549 94595	1.512 2501	0.0213 96397	0.1078 96246
GO_BP_DR_FLOOR_PLATE_FORMATION	0.715 67565	1.510 711	0.0376 0446	0.1086 6287
GO_BP_DR_RETINOIC_ACID_RECEPTOR_SIGNALING_PATHWAY	0.786 1703	1.508 1677	0.0179 10447	0.1108 0528
GO_BP_DR_EMBRYONIC_CAMERA-TYPE_EYE_MORPHOGENESIS	0.732 12725	1.505 6347	0.0359 116	0.1126 5574
GO_BP_DR_PECTORAL_FIN_MORPHOGENESIS	0.730 2532	1.504 125	0.0341 0641	0.1134 5666
GO_MF_DR_TRANSMEMBRANE_TRANSPORTER_ACTIVITY	0.565 9068	1.502 1721	0.0324 44958	0.1150 8129
GO_MF_DR_SODIUMPOTASSIUM-EXCHANGING_ATPASE_ACTIVITY	0.762 4621	1.490 2703	0.0287 7698	0.1282 3692
GO_MF_DR_OXIDOREDUCTASE_SINGLE_DONORS_MOLECULAR_OXYGEN_INCORPORATION_TWO_ATOMS_OXYGEN	0.568 90744	1.489 1564	0.0414 26927	0.1288 0209
GO_BP_DR_MIGRATION	0.672 549	1.484 6255	0.0408 97097	0.1338 8759
GO_BP_DR_RESPONSE_TO_BACTERIUM	0.624	1.479	0.0338	0.1388

	8711	8014	57316	8846
GO_BP_DR_CARDIAC_MUSCLE_TISSUE_DEVELOPMENT	0.779 50597	1.477 4916	0.0202 89855	0.1409 7826
GO_MF_DR_TRANSCRIPTION_REGULATORY_REGION_SEQUENCE-SPECIFIC_DNA_BINDING	0.580 7219	1.472 9912	0.0350 4673	0.1459 943
GO_BP_DR_POSITIVE_REGULATION_PROLIFERATION	0.609 1657	1.471 5387	0.0473 81546	0.1469 1561
GO_MF_DR_ENDOPEPTIDASE_INHIBITOR_ACTIVITY	0.670 57693	1.467 42	0.0364 86488	0.1516 5174
GO_MF_DR_TRANSCRIPTION_FACTOR_BINDING	0.502 4086	1.461 0547	0.0373 6264	0.1590 1224
GO_BP_DR_NERVOUS_SYSTEM_DEVELOPMENT	0.505 36734	1.458 8686	0.0305 34351	0.1610 5488
GO_BP_DR_VISUAL_PERCEPTION	0.648 2679	1.456 2209	0.0490 71617	0.1639 1163
GO_BP_DR_TRANSCRIPTION_DNA-DEPENDENT	0.428 84606	1.454 2688	0	0.1656 263
GO_MF_DR_DNA_BINDING_BENDING	0.525 36464	1.450 737	0.0476 731	0.1689 02
GO_BP_DR_CLOSURE_OPTIC_FISSURE	0.742 2217	1.439 2397	0.0353 30262	0.1782 1065
GO_BP_DR_NOTCH_SIGNALING_PATHWAY	0.503 5914	1.435 0682	0.0408 83977	0.1789 7797
GO_BP_DR_NEURAL_PLATE_MORPHOGENESIS	0.737 99026	1.436 7763	0.0459 7701	0.1797 0671
GO_BP_DR_CARTILAGE_DEVELOPMENT	0.478 72463	1.437 4508	0.0394 0362	0.1799 0687
GO_BP_DR_RESPONSE_TO_STRESS	0.478 99392	1.406 9592	0.0477 707	0.2140 685
GO_BP_DR_LIPID_METABOLIC_PROCESS	0.450 77786	1.397 2254	0.0425 75285	0.2188 0716
GO_MF_DR_ELECTRON_CARRIER_ACTIVITY	0.445 5002	1.395 9094	0.0347 64826	0.2199 0374
GO_BP_DR_PROTEOLYSIS	0.413 2657	1.388 5195	0.0020 02002	0.2296 9209