San Jose State University

SJSU ScholarWorks

Fall 12-17-2021

An Examination of Bone Loss During Space Travel with Differential Gene Expression Analysis

Claudia Vo

An Examination of Bone Loss During Space Travel with Differential Gene Expression Analysis

A Project

Presented to

Department of Computer Science

San José State University

In Partial Fulfillment

Of the Requirements for the Degree

By

Claudia Vo

December 2021

ABSTRACT

Spaceflight poses many risks to human health due to the harsh conditions of microgravity, cosmic radiation, and confinement. One of the impacts that spaceflight entails is bone loss, which is a risk to astronauts as future space missions will require long travel durations. To elucidate the role of genetics in spaceflight bone loss, in this study differential gene expression analysis was performed using Nextflow-RCP, an adaptation of NASA Genelab's RNA-Seq Consensus pipeline. The dataset for this project was GLDS-241, which contained samples from mice femoral skin. To gain a comprehensive understanding of the genes involved in bone loss, the results from the differential gene expression analysis were further analyzed using programs specifically for gene enrichment analysis. The findings demonstrated that there are many factors involved in bone loss under microgravity conditions. Altogether, the results from the gene enrichment analysis indicated a relationship between bone loss and glucose metabolism. However, additional studies on the mechanisms involved in bone loss are necessary to reduce bone loss in astronauts and assure their safe travel in space.

Keywords: bone loss, differential gene expression, Nextflow, gene enrichment analysis

ACKNOWLEDGMENTS

This project is dedicated to the following people who helped me make this project possible:

- Mom and Dad: You both made a heart-wrenching decision to escape Vietnam during a turbulent time and came to the US as refugees with nothing but the hope for a better life. You guys did exactly that and are living the American dream. I am incredibly thankful to have parents who instill the importance of education and show me what it means to persevere no matter what life may throw. For everything I've achieved, I owe it all to you.

- Dr. Philip Heller: Thank you for letting me be a part of your lab, and thank you for letting me take on this project. The 8-year-old version of me, who read her astronomy book cover-to-cover, would be so proud of what I have achieved today.

- Dr. Leonard Wesley and Dr. William Andreopoulos: Thank you for being a part of my committee. Also, thank you for being amazing professors and for teaching the next generation of Bioinformaticians. I wouldn't have gotten my first internship without your classes.

- Dr. Elinor "Elly" Nelson: For being a prime example of an independent and educated woman. You are such an incredible neighbor and thank you for taking the time to proofread my papers throughout my educational career.

- MSBI Classmates: I have learned so much from all of you. Also, I couldn't have navigated my way through this program without you all.

- Jon: You are an incredible partner, and I appreciate everything that you do. Thank you for being my rock all these years.

- Skipper: Finally, thank you for being the best dog and little brother I could ever ask for. And most of all, thank you for sticking with me from elementary school until graduate school. I love you to the moon and back.

TABLE OF CONTENTS

TABLE OF FIGURES

I.   INTRODUCTION

Spaceflight has come a long way since the beginning of space exploration. The first successful mission in 1961 only lasted an hour and 48 minutes and completed one orbit around the Earth [1]. However, given that it was the first mission in which a human was successfully launched into space, it was a remarkable achievement. Today, astronauts are sent to the moon and robots are sent to Mars. In addition, researchers have made many discoveries from spaceflight with longer missions. For example, researchers found that spaceflight has various effects on the human body due to confinement, cosmic radiation, and microgravity [2]. Some of these effects are on the cardiovascular, and immune, pulmonary, and digestive systems. One well-known effect, since the first long-duration space mission, is the development of bone loss [3], [4]. However, no effective solution to this ongoing problem has been discovered.

A. Bone Remodeling

The skeleton is a metabolically active organ that undergoes a continuous cycle of regeneration throughout life [5]. The purpose of undergoing this cycle is to adjust the skeletal architecture to meet the body's mechanical needs and to repair any damages in the bone matrix. Bone remodeling involves two main multicellular units—osteoblasts and osteoclasts—found in the three stages of the remodeling cycle. The stages of the remodeling cycle involve resorption, reversal, and formation seen in Figure 1 from [6].
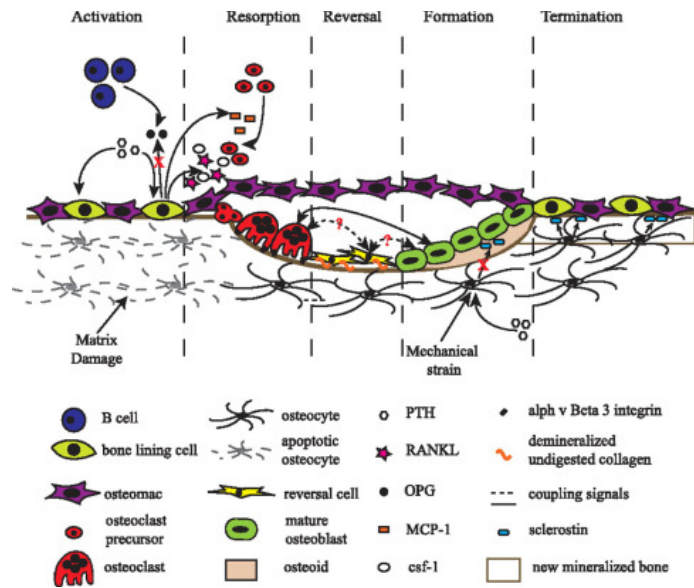
Figure 1. Representation of the bone remodeling cycle from [6].

In the resorption stage, osteoclasts digest old bone [6]. Then in the reversal stage, reversal cells cover the exposed bone surface and may send or receive signals that transition the bone from resorption to formation stage. Lastly, in the formation stage, osteoblasts produce new bone to replace the reabsorbed bone. As demonstrated in Figure 1, the bone remodeling cycle is a coupled process where bone reabsorption occurs at the same rate as bone formation. When the bone remodeling cycle becomes uncoupled, bone mass decreases as the rate of bone reabsorption becomes higher than the rate of bone formation. This ultimately leads to osteoporosis.

Osteoporosis is classified as a disorder of the skeleton where there is reduced bone mass and the increased risk of bone fractures, most likely related to metabolic changes in the bone and changes to the body's homeostasis [7]. Osteoporosis can occur in all ages and genders. However, it is more common in an aging population, particularly in postmenopausal women [8]. Those who are diagnosed with osteoporosis have a 40% lifetime chance of developing a bone fracture. The most common areas for bone fractures are in the wrist, spine, and hip. Other areas prone to

fracture are the humorous and ribs. Either way, a bone fracture can reduce a person's quality of life and even increase mortality up to 20%.

B. *Bone Loss in Space*

On Earth, bones can regenerate because gravity applies mechanical stress to the skeletal system [9]. The opposite is true in space, and one of the major obstacles that astronauts face during space missions is the risk of severe bone loss. In space, the human skeleton is not forced to regenerate when exposed to a microgravity environment. Therefore, in a weightless environment, astronauts may lose 1-2% of their bone density every month. According to Sirola and Iki et Al., bone loss in astronauts is ten times greater in a month than what postmenopausal women experience on average in a year on Earth [8], [10]. Luckily, bone loss from shorter missions is less severe because the body can reacclimate to a normal baseline. However, long-term missions may have more severe consequences for the human body. Such consequences are osteoporosis and a higher risk for bone fractures, which is why it is imperative that astronauts take certain countermeasures during spaceflight.

C. *Countermeasures in Space*

There are current countermeasures that astronauts can take to mitigate bone loss. The first is exercise. Astronauts are currently required to spend 3 hours a day exercising on weight bearing machines that replicate the resistance of Earth's gravity [3]. Astronauts can also take dietary supplements in addition to exercising since there is limited sun exposure during spaceflight. Some supplements that astronauts can take are vitamin D and calcium. Nonetheless, these countermeasures may not be adequate for longer space endeavors in the future [3], [11]. As

technologies advance, spaceflight will as well, and we must anticipate what the future entails. For example, because future space missions will send women to the Moon in 2024, astronauts to Mars, or even the public to space through private companies like Blue Origin, the current countermeasures are not enough to prevent osteoporosis.

D. *NASA GeneLab's RNA-Seq Consensus Pipeline*

The aim of this project is to utilize the Nextflow RNA-Seq Consensus Pipeline (Nextflow-RCP), an implementation of NASA Genelab's RNA-Seq Consensus Pipeline, to analyze the effects of bone loss in mice under the harsh conditions of microgravity [12],[13]. Nextflow-RCP is an RNA-sequencing pipeline that was created by a former San José State graduate student, Jonathan Oribello. A copy of the Nextflow-RCP pipeline can be found at https://github.com/J-81/Nextflow_RCP. Nextflow-RCP is implemented on Nextflow, which is open-source software that allows workflows using software containers [14]. This pipeline takes short read RNA-sequencing data to detect differentially expressed genes (DEGs) through quality control, trimming, and mapping the reads to a reference genome. By examining the DEGs, we may discover more about bone loss through the upregulated and downregulated genes. Ultimately, this research will give NASA insight into the mechanisms involved in bone loss and hopefully potential targets for bone loss treatments in astronauts.

E. *GLDS-241 Dataset*

The dataset used for this specific study is GLDS-241 version 8, which was originally implemented by UCLA in partnership with NASA. This dataset is found on the NASA Genelab

repository [15], [13]. The Genelab repository is a database that stores and analyzes omics data from spaceflight and other space-related experiments.

The GLDS-241 dataset involves RNA sequence data of femoral skin samples from a total of 19 mice. Of those 19, 10 were ground control mice while the other 9 were mice on the space mission. Table I. shows the schematic overview of this experiment.

TABLE I
GLDS-241 Treatment Protocol

Days from Launch	6	14	30	44
Ground Control	0.3 ml PBS	0.3 ml PBS 20 mg/kg Calcein green in 0.1 mL PBS	0.3 ml PBS and 0.1 mL of Alizarin Red	0.3 ml PBS and 0.1 ml Calcein green
Spaceflight Mice	0.3 ml PBS	0.3 ml PBS 20 mg/kg Calcein green in 0.1 mL PBS	0.3 ml PB and 0.1 mL of Alizarin Red	0.3 ml PBS and 0.1 ml Calcein green

A cohort of 40, 30-week-old female mice were flown to the International Space Station (ISS) for the Rodent Research-5 project (RR-5). All the mice were kept in the Rodent habitat for a total of 30 days. Six days after the launch, 20 mice received intraperitoneal injections of 0.3 ml phosphate-buffered saline (PBS) as a control. On day 14, this treatment was repeated. However, all the mice were also given intraperitoneal injections of 20 mg/kg Calcein green in 0.1 mL PBS as a bone marker.

After these treatments, 10 control mice were randomly selected for Live Animal Return (LAR) and sent back to Earth 30 days after the launch. After arriving back on Earth, the LAR mice spent another 30 days in a standard habitat to reacclimate to Earth's conditions. Within these 30 days, the LAR mice received another two treatments, the first consisting of 0.3 ml PBS

and 0.1 mL of Alizarin Red the day of arrival on Earth (or day 30 after launch) and the second with 0.3 ml PBS and 0.1 ml Calcein green (day 44 after launch).

Ground control mice treatments were conducted on a 3-day delay at Kennedy Space Center in ISS Environmental Simulators (ISSES) programmed with data from the Dragon capsule and the ISS cabin. The ground control animals received the same treatments with the same duration as the spaceflight mice.

After day 60, the mice ground control and spaceflight mice were euthanized with ketamine and xylazine. The femoral skin samples were preserved for RNA preparation, and the paired-end RNA was sequenced with Illumina NovaSeq600 [16], [17].

II. METHODS

A. High Performance Computing Cluster

The San Jose State University College of Science High-Performance Computing Cluster (SJSU CoS HPC) was used to run Nextflow-RCP since the programs needed to analyze GLDS-241 are computationally extensive and require a lot of memory [18]. SJSU CoS HPC is a Linux-based computing system that consists of 36 nodes and 128 to 258 gigabytes (GB) of RAM. To submit jobs on SJSU CoS HPC, a SLURM script is required.

B. Nextflow-RCP
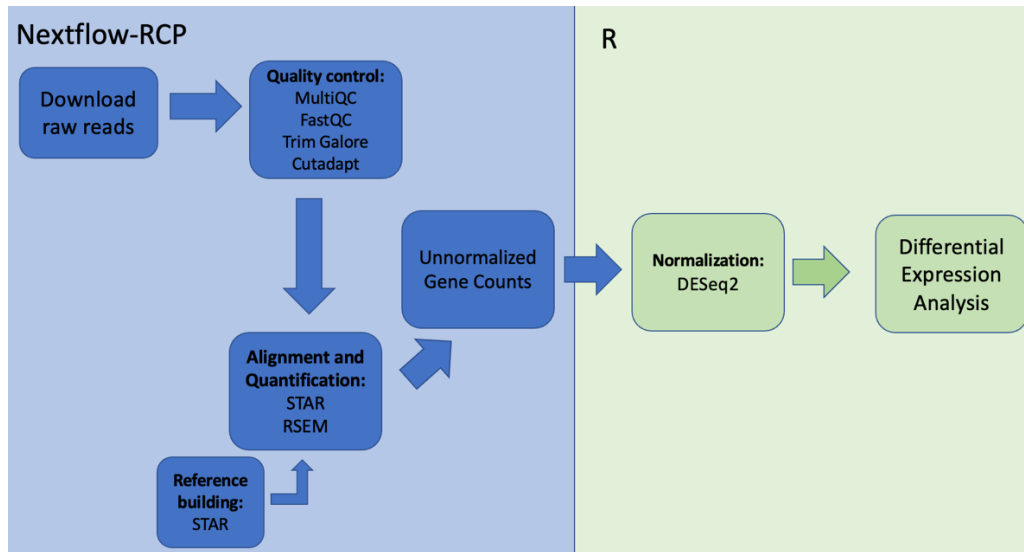
Figure 2. Nextflow-RCP workflow. The blue represents the processes that occurs in the pipeline. The green indicates differential gene expression analysis in R after the Nextflow-RCP.

Figure 2. Demonstrates the overall workflow for Nextflow-RCP. The steps to the pipeline involve downloading raw reads, quality control with FastQC and MultiQC, trimming raw reads with Trim Galore, aligning with STAR, and quantifying gene counts with RSEM.

1) Obtaining Raw Data

Nextflow-RCP first downloads the GLDS-241 FastQ files from the Genelab repository using a Slurm script containing GLDS-241 as an argument. Unlike the previous implementation of the Nextflow-RCP, the new update does not require a configuration file that directs Nextflow-RCP to download the GLDS-241 raw reads from the online repository.

2) Quality Control and Trimming

FastQC, MultiQC, Trim Galore and Cutadapt software were used to perform quality control of the FASTQ files [19]–[21]. FastQC is a software program that is used to assess the

- 14 -

quality of the individual FASTQ files for each sample. However, it's time-consuming to view the quality of each file. Therefore, MultiQC was used to summarize the quality from FastQC across all the raw reads. The FastQC and MultiQC versions used were 0.11.9 and 1.9, respectively.

The paired-end reads and the adapters were trimmed with Trim Galore and Cutadapt, versions 0.6.6 and 3.2, which remove low-quality bases and the adapter from the reads. After the raw reads are trimmed, FastQC and MultiQC generate the quality metrics for the trimmed reads.

3) Aligning and Quantifying Reads

The trimmed reads were aligned to the *Mus musculus* reference genome using Spliced Transcripts Alignment to a Reference (STAR), version 2.7.7a [22]. STAR is an open-source splice-aware alignment software that involves a seed search and a clustering, stitching, and scoring step in its algorithm in order to map reads to a reference genome. STAR first creates the indexed genome files that help with read mapping using the reference FASTA file and the gene transfer format (GTF) annotation file. The *M. musculus* reference FASTA and GTF annotation files in this project are from the Ensembl release version 104 [23]. Then, STAR uses the trimmed reads and indexed genome files to map the reads to the reference genome.

To quantify the mapped reads to each gene and transcript, Nextflow-RCP uses RSEM version 1.3.3 [24]. RSEM computes the maximum likelihood estimate to reads across multiple genes. Like the first step in STAR, RSEM creates an indexed gene file using the GTF and reference genome files. With the indexed gene file and the mapped reads, the genes are assigned a count.

The output files from this step include two files. One file contains the counts assigned to each gene and another that contains counts to each isoform. Since this project was only

concerned with the differential gene expression between ground control and space mice, only the gene counts file was required for the differential gene expression analysis.

4) *Differential Gene Expression Analysis*

After downloading the counts, an R script was created to obtain the DEGs. The version of R that was used was 4.1.1 on macOS 11.6. The DESeq2 package on R was used for differential gene expression [25].

a. *Normalization*

The first step in the R script was to normalize the reads from the downloaded unnormalized gene counts file. The data was formatted as a DEG object with the DESeq2 package on R. The genes were pre-filtered so that genes with a count less than 10 across all samples were removed because these would have a low expression in both ground control and space flight conditions [26]. In addition, keeping the genes with no or low counts can increase the number of final DEGs reported. Therefore, we can ignore these genes to improve the reliability between the mean-variance.

b. *DESeq()*

With the DEG object, the DESeq() function in R can perform the size factor estimation, estimate the dispersion, and hypothesis testing.

(a) *Size factor Estimation*

$$\widehat{s}_j = \operatorname*{median}_i \frac{k_{ij}}{\left(\prod_{v=1}^m k_{iv}\right)^{1/m}}.$$

The size factor is estimated using the method described by Equation 5 in [27]. This method takes the median of ratio of the observed counts where the numerator is the jth sample's

- 16 -

counts and denominator is a pseudo-reference sample created by the geometric mean of all samples. By estimating the size factors, we can compare the counts between different samples if there are discrepancies between sequencing depth among the samples.

(b) Estimating dispersions

The next step to normalization involves estimating the dispersion, or the variance of the data. In the DESeq() function, the dispersion is estimated for each gene, using an empirical Bayes shrinkage method [25]. Each gene is treated separately, and the gene-wise dispersion is estimated. For the estimates, the location parameter of the distribution is determined. The gene-wise dispersion estimates are shrunken based on the distribution using empirical Bayes.

(c) Hypothesis testing

After normalizing the counts data, hypothesis testing was performed to determine which genes were differentially expressed between space flight mice and ground control. The Wald significance test was used for hypothesis testing. For differential gene expression analysis, Wald's test is the default method for hypothesis testing when two groups are being compared. However, by only using the p-values from the Wald's test to obtain DEGs, this would result in a multiple testing problem that increases the number of false positives as more genes are tested. To correct the multiple testing problem, the Benjamini-Hochberg (BH) procedure was used [28]. The BH procedure controls the false discovery rate (FDR) through a Bonferroni correction for multiple hypothesis testing. In the BH method, the genes assigned ranks to its p-value, then the BH critical value for each p-value is calculated, and finally, the original p-values are compared to the BH critical value.

Traditionally, genes that have an FDR, or adjusted p-value ≤ 0.05 and a log2 fold change ($\log_2$FC) of 1, are considered to be differentially expressed. However, since the sample sizes of

the space flight and ground control samples are relatively small, using the traditional method for filtering DEGs may result in no or very few genes. Therefore, this study considered genes with a p-value threshold lower than 0.01 and a log_2FC of 1 as differentially expressed.

5) Gene Enrichment Analysis

A gene enrichment analysis was performed to gain a further understanding of biological mechanisms that the DEGs play by looking for the classes of genes that represent functions. The software used for gene enrichment analysis were DAVID, gprofiler2, and GSEA [29]–[32].

a. DAVID

Database for Functional Annotation, Visualization, and Integrated Discovery (DAVID), version 6.8, is an open-source web server that provides functional annotation tools for large gene lists in genomics studies [29]. DAVID avoids duplicated genes by calculating the Fisher Exact test based on the DAVID gene IDs and removing repeated IDs. The functional annotation tool accepts a ranked gene list as the input and groups them into annotation groups by measuring the degree of association with a Kappa statistic and a fuzzy clustering algorithm. As the gene list, the Ensembl IDs for the GC vs. FLT DEGs were uploaded onto DAVID to determine which biological processes were enriched. On DAVID, the thresholds can be selected. For this project, the default thresholds were used where the Kappa similarity threshold was 0.5 and the maximum EASE score/p-value was set at 1.

b. gprofiler2 (0.2.1)

Gprofiler2, version 0.2.1, is a functional enrichment analysis and visualization package on R available through CRAN or conda-forge [32]. This R package accompanies g:profiler, the web-based server, and facilitates access to g:profiler's computations and database [33].

Gprofiler2 also provides access to many annotation sources including KEGG, GO, Reactome, Wikipathways, miRTarBase, CORUM, Human Protein Atlas, etc. [33]–[40]. One benefit to using gprofiler2 is that these annotation sources are readily available without having to download the database.

Gprofiler has four main wrapper functions: gost, gconvert, gorth and gsnpense. This study only uses the gost function for gene enrichment. gost accepts either a list of DEG Ensembl IDs or separate lists of ordered upregulated or downregulated DEGs. To visualize the difference between the upregulated and downregulated genes, separate lists of ordered upregulated or downregulated DEGs were used in this study. As for visualization of the enrichment analysis results, gprofiler uses ggplot2 and plotly similar to g:profiler [41], [42].

c. GSEA

Gene Set Enrichment Analysis (GSEA) is a software package that interprets gene expression data by grouping genes based on common biological function, chromosomal location, or regulation [30], [31]. Most methods for gene expression data interpretation, like DAVID and gprofiler2, accept the gene list of the DEGs. But, by using a list of DEGs, some of the genes with biological relevance may be filtered out due to the correction of multiple hypothesis testing, or a long list of DEGs may not have a meaningful relationship as a whole. GSEA can overcome these challenges and extract more meaning from the data by using the normalized RNA-Seq. count data as the input. With the count data, GSEA calculates its own metrics: the estimation of significance level (ES) and the normalized enrichment score (NES). These two metrics take the degree of overrepresentation and its statistical significance into account. In addition, the metrics solve the problem with multiple hypothesis testing.

This study uses GSEA version 4.1.0. Per the documentation, there are multiple levels of thresholds for the gene set permutation, depending on how conservative or permissive the gene set permutation should be. The permissive threshold involves a p-value < 0.05 and an FDR < 0.25 while a conservative threshold involves a p-value < 0.001 and an FDR < 0.05. For this study, the most permissive threshold, p-value > 0.05 and FDR < 0.25, was used to prevent from filtering out significant results.

III. RESULTS

A. *Nextflow-RCP*

  1) Quality Control and Trimming

FastQC: Mean Quality Scores

(b)
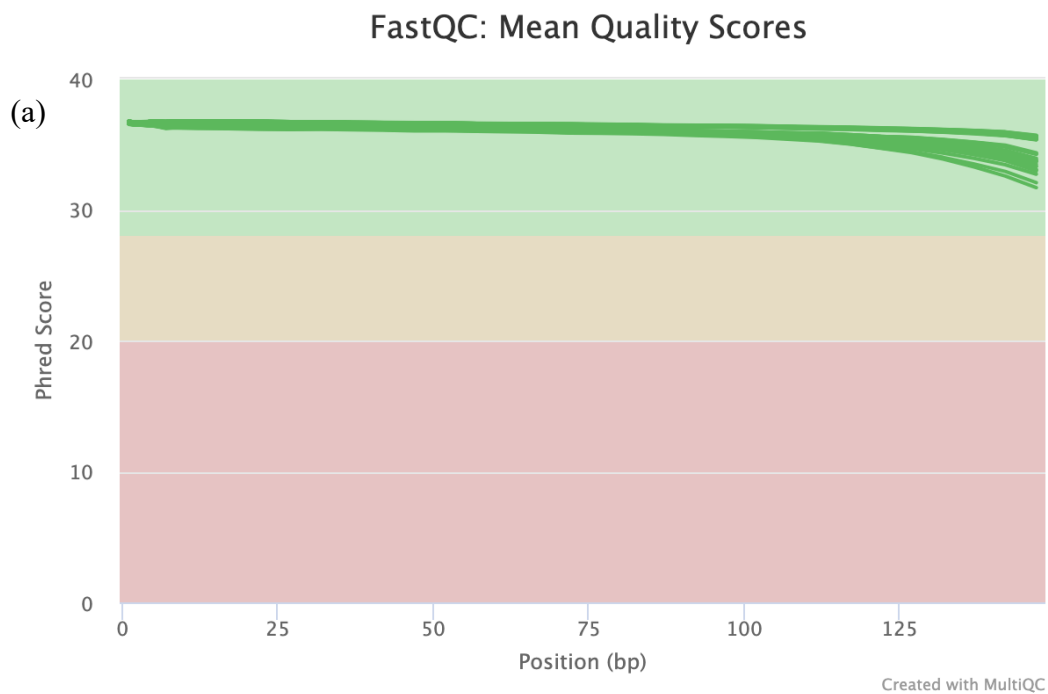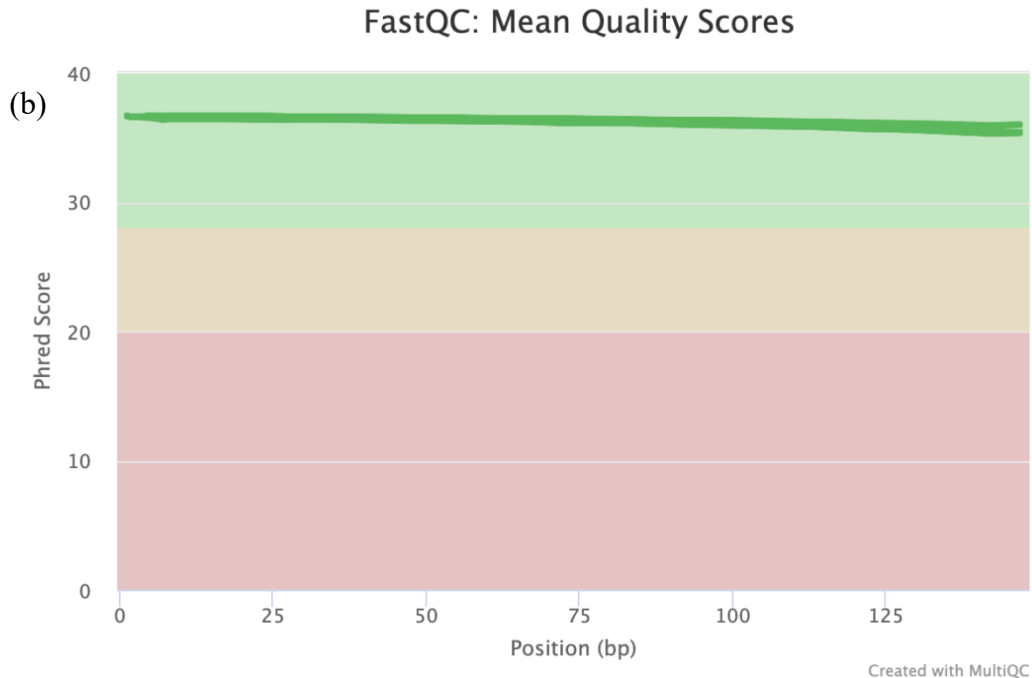
Figure 3.(a) FastQC Mean Quality Scores Per Base. A (b) Fast QC Mean Quality Score

The quality of the raw and trimmed reads from the MultiQC reports is seen in Figure 3. Each line, in green, represents the average per base Phred score for each read. A Phred score over 30 indicates that the base call has an accuracy of 99.9%, in other words, the probability of an incorrect base is 1 out of 1000 bases [43]. Figure 3A. indicates that the overall quality for all the raw reads was high since they were around 38. Although, towards 115 base pairs and onwards, the average quality per base decreases because of the adapter content placed at the end of the reads.

After trimming the reads with Trim Galore and Cutadapt, the quality of the reads slightly improved, demonstrated in Figure 3B. In addition, MultiQC reported that there were no samples with an adapter content over 0.1%, indicating that Cutadapt successfully removed the adaptor content at the end of each sample. Figure 4 shows the adapter content in each sample before trimming the reads.

FastQC: Adapter Content
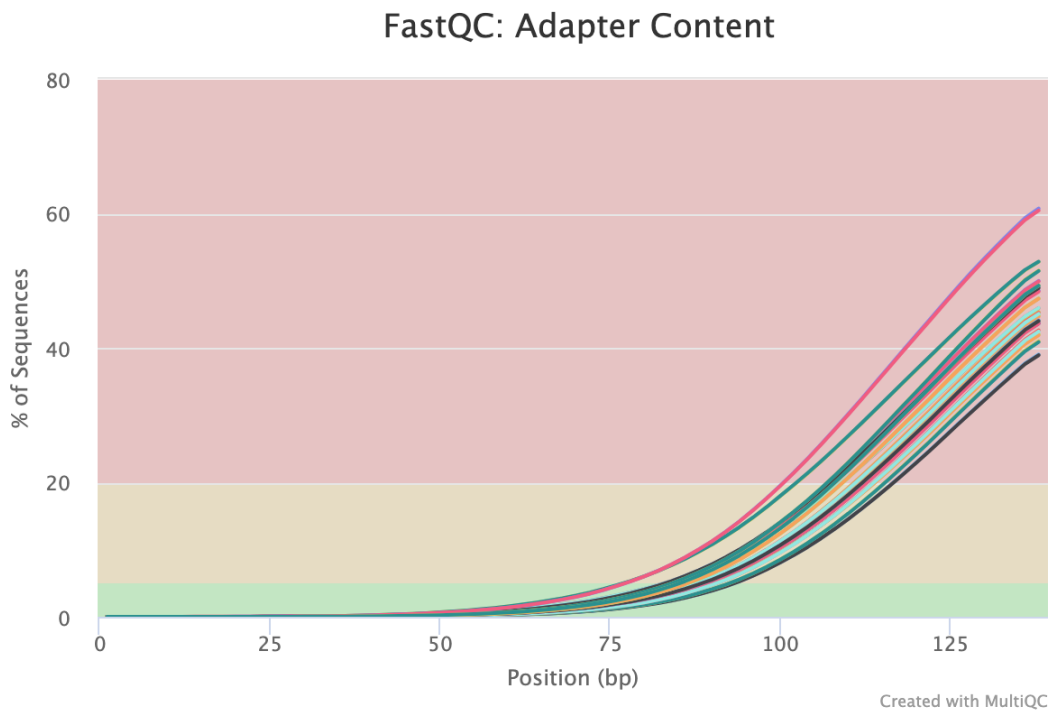
Figure 4. FastQC: Adapter Content

2) STAR Alignment
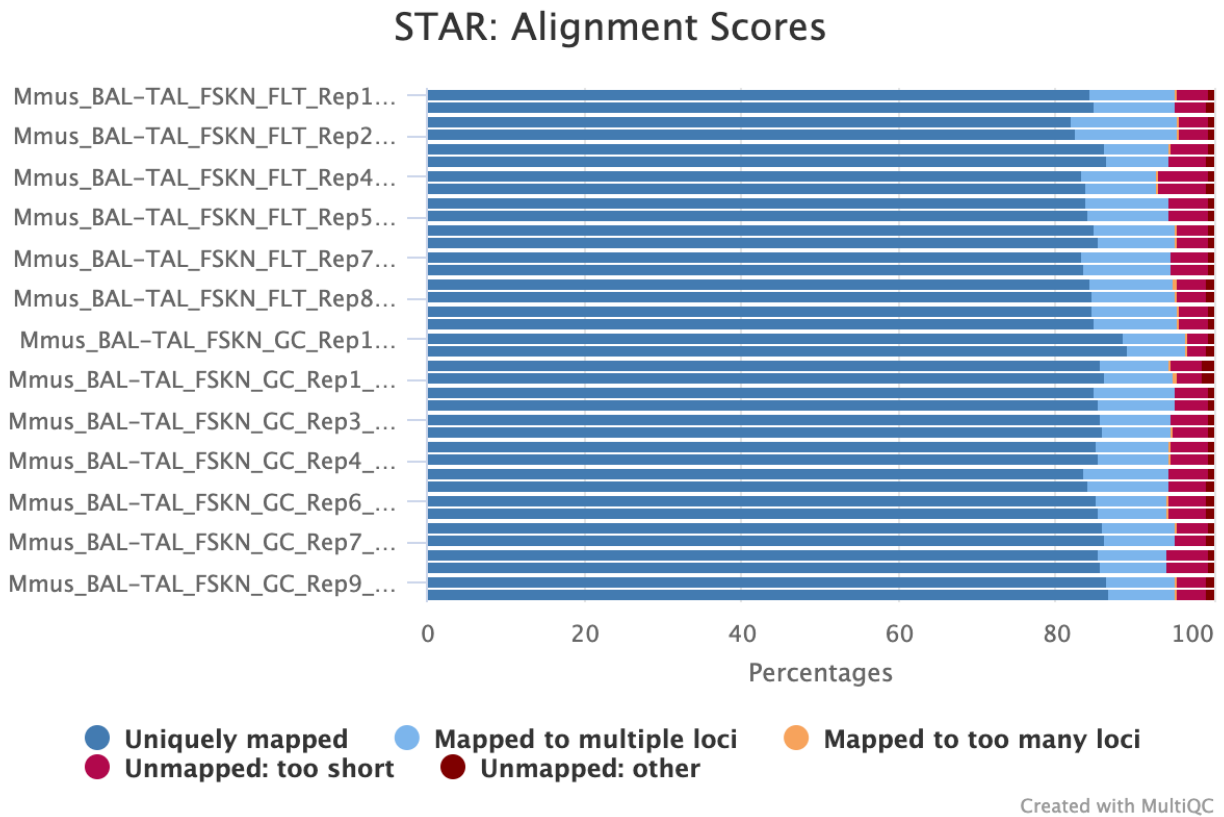
Figure 5. STAR Alignment Scores

The STAR alignment results across all the samples are seen in the MultiQC report in Figure 5. The color bands are correlated to the percentage of genes that were or were not mapped to a locus or loci. 81-89% were uniquely mapped with GC sample 1 having the highest percentage and FLT sample 2 with the lowest. 8-11% were mapped to multiple or too many loci in all the samples. As for the unmapped genes about 2-6% were unmapped across all the samples.

B. Differential Gene Expression

1) Normalization

The Principal Analysis Component (PCA) plots in Figure 6 use dimensionality reduction to visualize the variance among multidimensional samples. There was some variance among the GC and FLT mice before normalizing the gene counts with DESeq2 in Figure 6A. Some of the GC samples, in pink, are clustered with each other in the lower-left corner while the FLT samples, in teal, tend to cluster towards the upper right diagonal of the plot. But some of the GC and FLT samples were grouped near each other. After normalization, in Fig. 6B, one of the GC samples clustered closer to the FLT samples. Overall, most of the samples in the GC and FLT groups had a consistent expression before and after normalization.

(a)

(b)
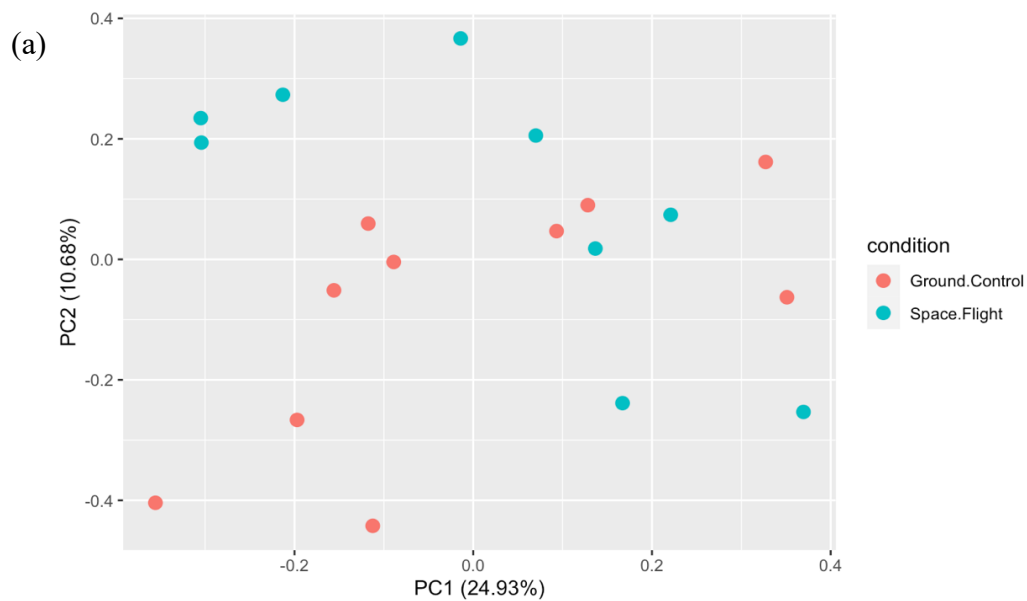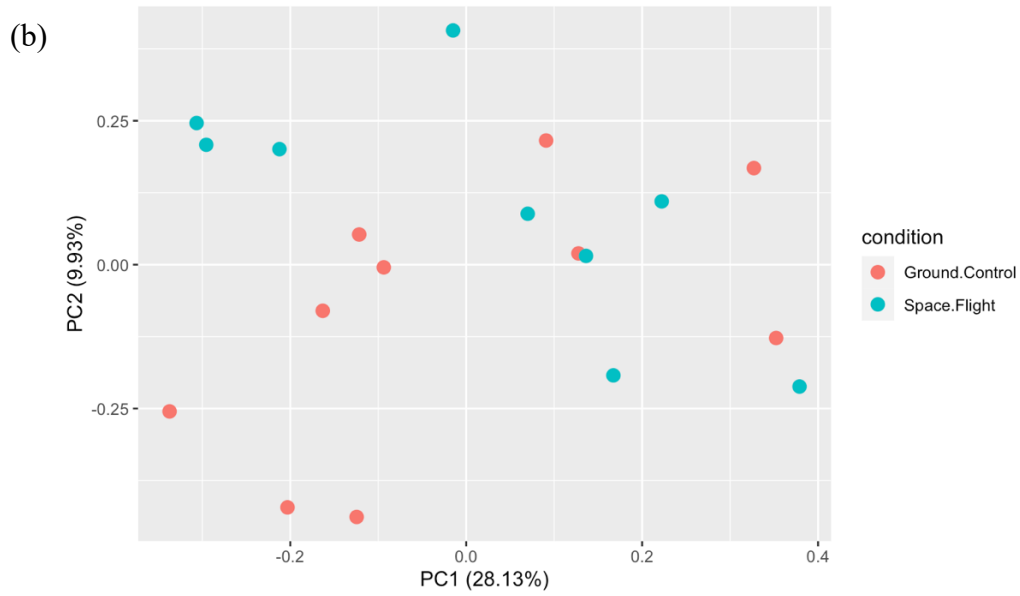
Figure 6. (a) top PCA plot (b) bottom PCA plot

2) DEGs
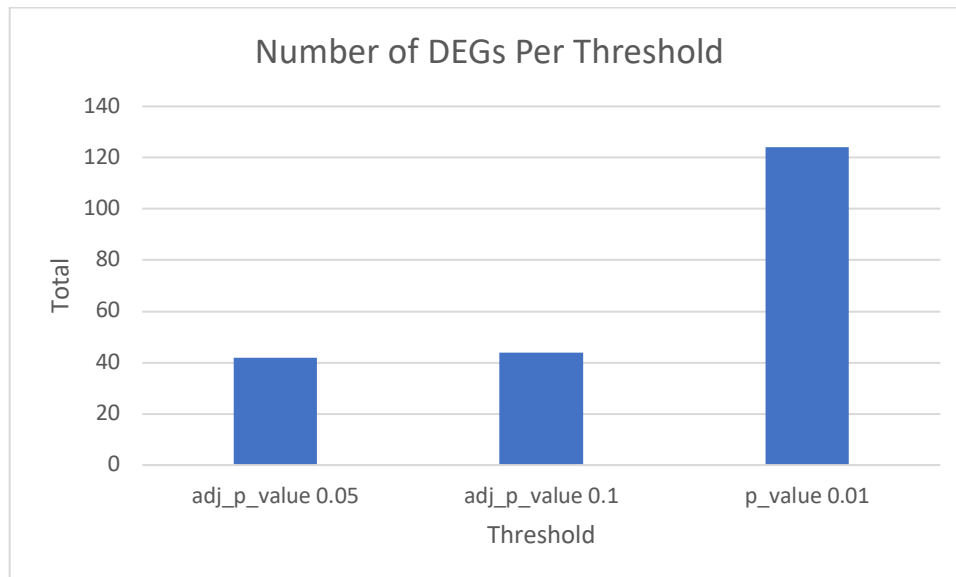
Figure 7. The Number of DEGs found per threshold

Figure 7 shows the number of DEGs using the different levels of the adjusted p-value and

p-value for GC vs. FLT samples. For both the adjusted p-value and p-value, the threshold for the

log_2FC was 1. At an adjusted p-value of 0.05 and 0.1, the number of DEGs found were 42 and

44, respectively. At a p-value of 0.05, the number of DEGs increases to 124. The number of DEGs using the adjusted value of 0.05 and 0.01 are low. This could potentially filter out important findings. Therefore, a p-value of 0.01 rather than the adjusted p-value of 0.05 was used for the differential expression analysis in this project.
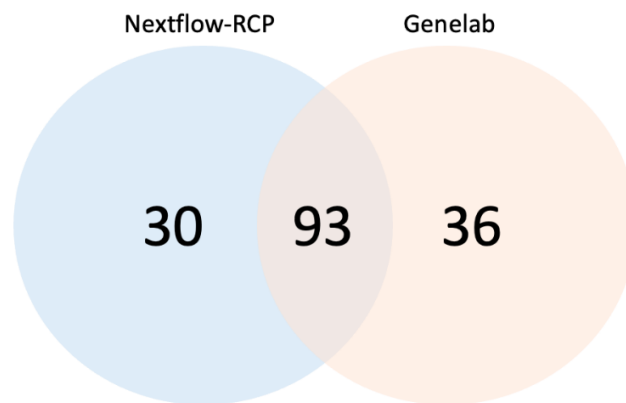
Figure 8. Venn Diagram of the DEGs from Nextflow-RCP and Genelab.

The results from the Nextflow-RCP and Genelab-RCP were compared by downloading the differential gene analysis results from the Genelab repository for the GLDS-241 dataset and running a comparison on R. The Genelab-RCP obtained a total of 129 DEGs between the GC vs. FLT groups, using a p-value threshold of 0.01 and a $\log_2$FC of 1. The number of DEGs from the Genelab-RCP was not far off from the 124 DEGs from the Nextflow-RCP at the same thresholds. Between the two pipelines, there were 93 DEGs in common, seen in the center of the Venn-diagram in Figure 8. The left indicates the number of unique DEGS found from the Nextflow-RCP while the right is the number of unique DEGs from the Genelab-RCP.
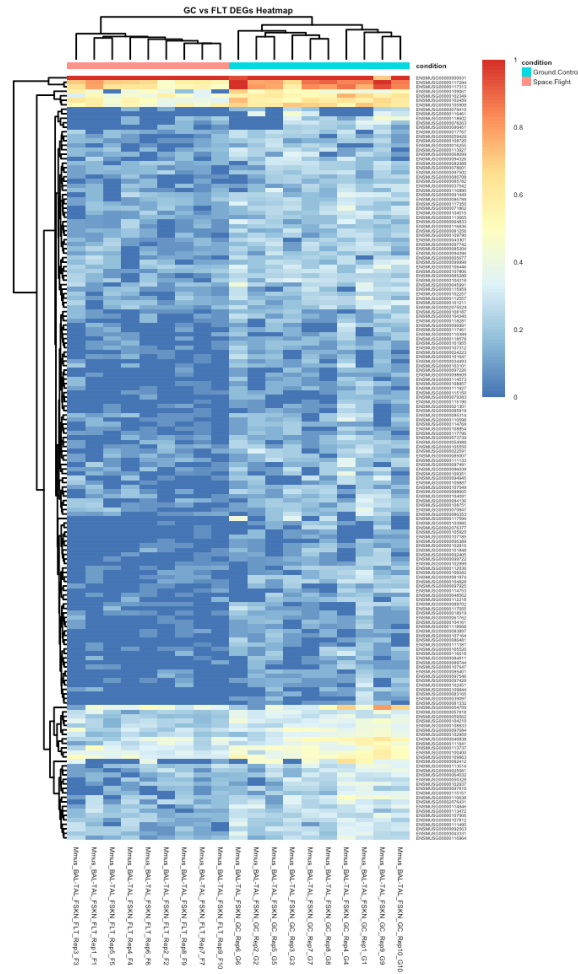
Figure 9. GC vs FLT DEG Heatmap. The GC samples are located on the right of the heatmap in teal and the FLT samples are on the left in pink.

A clustered heatmap was created to visualize the difference between the DEGs in the GC and FLT samples in Figure 9. The x-axis represents the GC and FLT samples, and the y-axis consists of the 124 DEGs. The top of the heatmap shows a dendrogram where the GC and FLT samples clustered in their respective groups. The 9 FLT samples are on the left of the heatmap in pink, and the 10 GC samples are on the right in teal.

The colors in the heatmap indicate the levels of expression in the genes. The red denotes that the gene is highly expressed, and the blue denotes low expression. The expression levels of the genes in the GC samples are all similar as well as in the FLT samples. When comparing the

gene expression levels between the GC and FLT groups, the FLT samples overall have a lower expression across all genes.

C. Gene Enrichment Analysis

  1) DAVID

Table II consists of the top 4 annotation clusters from the GC vs. FLT groups ordered by the overall enrichment scores. The higher the enrichment score, the more enriched the group is. The first cluster has the highest enrichment score of 6.07 and suggests that many protein transporters of hydrophobic molecules are impacted in space. The second cluster has an enrichment score of 5.06 and shows that urinary proteins are also affected in space. Interestingly, the third cluster suggests that there is an impact on insulin secretion and metabolism. As for the fourth cluster, immunoglobulins are shown to be impacted. However, the fourth cluster has the lowest enrichment score of 0.59 and an FDR of 1. These values suggest that the annotations in the cluster may not be strongly enriched as the FDR value ranges from 0 to 1. An FDR value of 0 suggests perfect enrichment and conversely, an FDR equal to 1 is poorly enriched.

Table II
Top 4 DAVID Annotation Clusters for GC vs. FLT

Category	Term	FDR
Cluster 1, Enrichment Score 6.07		
INTERPRO	Lipocalin fatty-acid binding protein	3.42E-06
INTERPRO	Major Urinary protein	3.42E-06
INTERPRO	Calycin	3.42E-06
INTERPRO	Calycin	3.42E-06
INTERPRO	Lipocalin	4.00E-06
INTERPRO	Lipocalin conserved site	5.52E-06
GOTERM_MF_DIRECT	Small molecule binding	3.53E-05
GOTERM_MF_DIRECT	Pheromone binding	2.54E-03
GOTERM_MF_DIRECT	Transporter activity	1.71E-02
Cluster 2, Enrichment Score 5.56		
UP_SEQ_FEATURE	Major urinary protein 1	1.17E-04
UP_SEQ_FEATURE	Major urinary protein 2	1.17E-04
UP_SEQ_FEATURE	Major urinary protein 6	1.17E-04
Cluster 3, Enrichment Score 2.98		
GOTERM_MF_DIRECT	Insulin activated receptor activity	8.00E-03
GOTERM_BP_DIRECT	Heat generation	2.21E-02
GOTERM_BP_DIRECT	Positive regulation of lipid metabolism	2.21E-02
GOTERM_BP_DIRECT	Cellular response to lipid	2.21E-02
GOTERM_BP_DIRECT	Negative regulation of lipid biosynthetic process	2.21E-02
GOTERM_BP_DIRECT	Negative regulation of lipid storage	2.21E-02
GOTERM_BP_DIRECT	Positive regulation of glucose metabolic process	2.21E-02
GOTERM_BP_DIRECT	Energy reserve metabolic process	2.21E-02
GOTERM_BP_DIRECT	Negative regulation of insulin secretion	2.21E-02
GOTERM_BP_DIRECT	Negative regulation of gluconeogenesis	2.35E-02
GOTERM_BP_DIRECT	Locomotor rhythm	2.48E-02
GOTERM_BP_DIRECT	Mitochondrion morphogenesis	3.42E-02
GOTERM_BP_DIRECT	Aerobic respiration	4.65-E02
GOTERM_BP_DIRECT	Positive regulation of protein kinase B signaling	0.25
GOTERM_BP_DIRECT	Glucose homeostasis	0.42
Cluster 4, Enrichment Score 0.59		
SMART	IG	1
UP_KEYWORDS	Immunoglobulin domain	1
INTERPRO	Immunoglobulin subtype	1
INTERPRO	Immunoglobulin-like domain	1
INTERPRO	Immunoglobulin-like fold	1

2) Gprofiler2
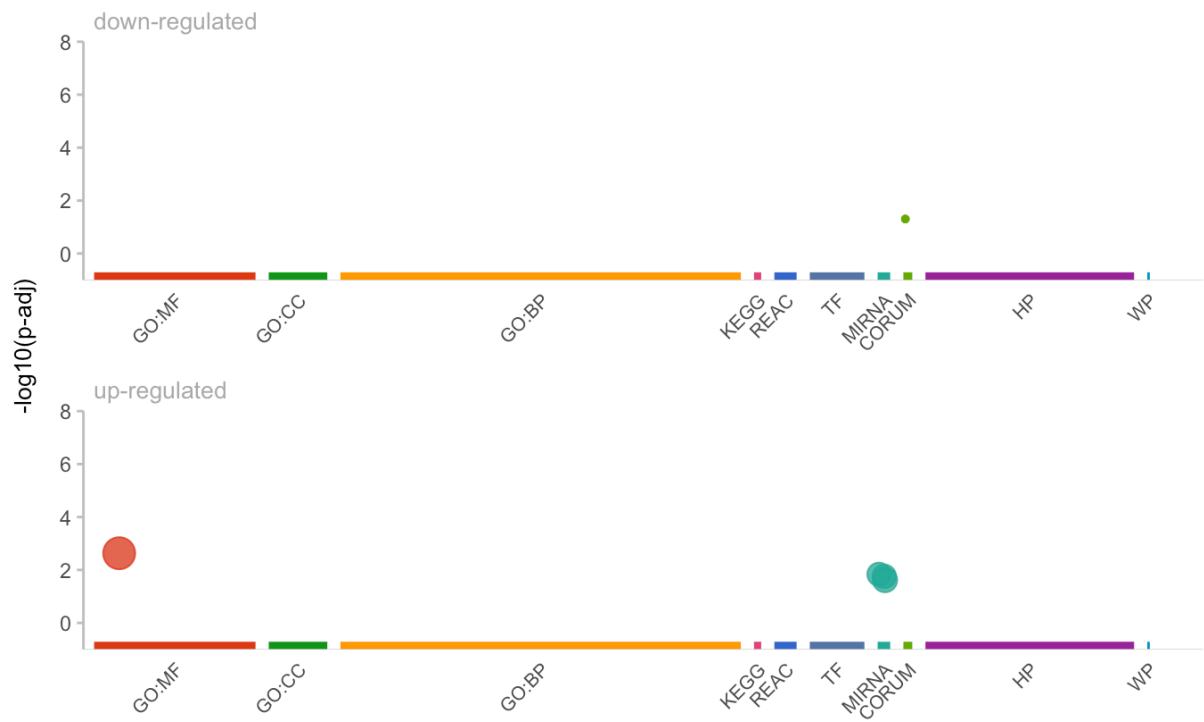
Figure 10. Manhattan Plot from gprofiler2.

Table III
Down and Up-regulation

id	source	term_id	term_name	term_size	intersection_size_up-regulated	intersection_size_down-regulated	p_value_up-regulated	p_value_down-regulated
1	CORUM	CORUM:1018	Pcdha7-Pcdhgb4 complex	2	NA	1	NA	5.0e-02
2	CORUM	CORUM:1019	Pcdha7-Pcdhga3 complex	2	NA	1	NA	5.0e-02
3	CORUM	CORUM:1020	Pcdha7-Pcdhgb2 complex	2	NA	1	NA	5.0e-02
4	CORUM	CORUM:1021	Pcdha7-Pcdhga1 complex	2	NA	1	NA	5.0e-02
5	GO:MF	GO:0005550	pheromone binding	107	4	NA	2.3e-03	NA
6	MIRNA	MIRNA:mmu-miR-151-3p	mmu-miR-151-3p	11	2	NA	1.5e-02	NA
7	MIRNA	MIRNA:mmu-miR-378b	mmu-miR-378b	12	2	NA	1.7e-02	NA
8	MIRNA	MIRNA:mmu-miR-463-3p	mmu-miR-463-3p	14	2	NA	2.4e-02	NA
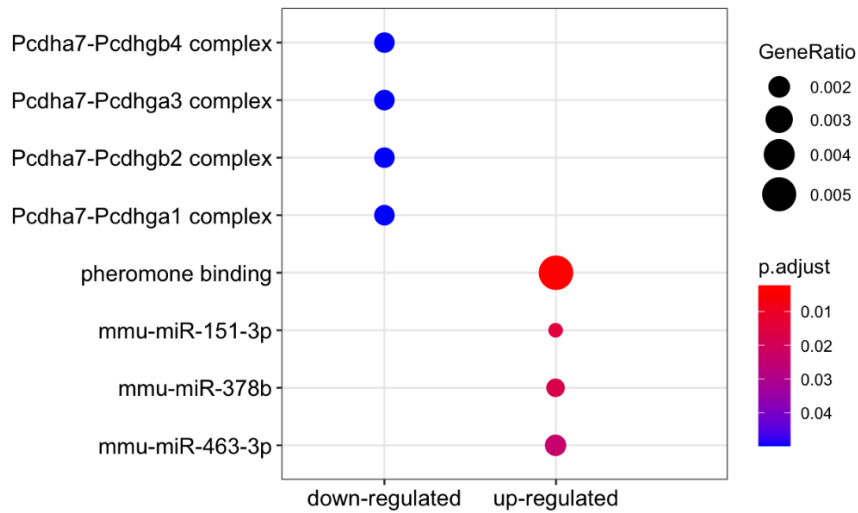
Figure 11. Side-By-Side Vertical Dot plot representation of term sizes between upregulated and downregulated gene sets.

Figure 11. illustrates the enrichment results visualized on a Manhattan plot using the two upregulated and downregulated lists. The top plot shows the enrichment results for downregulated genes while the bottom shows the results for the upregulated genes. The x-axis indicates the grouped functional terms ordered according to the different data sources and positioned according to the source order [32]. The source order is defined in such a way that the closer the dots are to each other, the closer they are in the source hierarchy. The y-axis in the Manhattan plot shows the adjusted p-values using the -$\log_{10}$ as a scale. The dots represent a single term and the size of the dot is correlated to the term size, so the larger the dot, the larger the term size.

Table III shows the corresponding statistics to Figure 8, while figure 11 shows a side-by-side dot plot of the term sizes between up and downregulated gene sets. Overall, pheromone binding had the lowest adjusted p-value at 2.3E-03 and the largest term size of 107. This corresponds to the results seen in DAVID, which also suggests that pheromone binding is a significantly enriched biological pathway as seen in Cluster 1 in Table II.

The other enriched pathways are the three Pcdha7-Pcdhgb* complexes found in the enrichment results for the downregulated genes. Pcdha7 is a member of the protocadherin gene cluster linked on chromosome 5 [44]. The *Pcdha7* gene encodes protein complexes that are organized similarly to the structure of B and T cell receptors.

Another fascinating find is the mmu-miR enrichment groups in the upregulated genes from the miRNA database. Mmu-miR is a mouse miRNA, a single-stranded non-encoding RNA that plays a role in gene expression by regulating transcription [45].

3) GSEA

Table IV
GSEA DOWNREGULATED HALLMARK GENES

Gene Set	NES	NOM p.value	FDR	FWER p.value
HALLMARK_ADIPOGENESIS	-1.66	0.035	0.254	0.142
HALLMARK_DNA_REPAIR	-1.63	0.034	0.226	0.217
HALLMARK_OXIDATIVE_PHOSPHORYLATION	-1.62	0.038	0.152	0.218
HALLMARK_FATTY_ACID_METABOLISM	-1.59	0.057	0.173	0.296
HALLMARK_REACTIVE_OXYGEN_SPECIES_PATHWAY	-1.53	0.043	0.256	0.421
HALLMARK_PEROXISOME	-1.50	0.072	0.236	0.486
HALLMARK_PROTEIN_SECRETION	-1.48	0.110	0.243	0.531

11 out of 50 genes sets were upregulated in the GC phenotype. But none of these genes were significantly enriched. In the FLT phenotype, 39 out of 50 gene sets were found to be downregulated. Of the 59, 5 gene sets were significantly enriched with an FDR of less than 25%. In addition, 6 gene sets were significantly enriched at a nominal p-value less than 5%. Table IV. lists the downregulated gene sets from the Hallmarks database that either met the FDR or nominal p-value cutoff [46]. From Table III, DNA repair and oxidative phosphorylation both meet the thresholds that GSEA considers as significant findings.

IV. DISCUSSION

A. Genelab-RCP

The Nextflow-RCP is a good tool to use for differential gene expression analysis. To a certain extent, the Nextflow-RCP is a reproducible implementation of the Genelab-RCP. As seen in Figure 3, there is a huge number of shared DEGs for GC vs. FLT comparison that was reported. A possible reason for the minor discrepancy between the two pipelines could be the tool versions used during each step. In this study, Ensembl version 104 was used during the alignment step, which is the most updated version released as of May 2021. On the other hand, Genelab used version 96, released in April 2019, for the alignment. Therefore, there could have been some updates to the *M. musculus* genome in between the two versions, which could affect overall STAR alignment and RSEM counts. Another reason for the discrepancy in the results between the two pipelines is the difference in operating systems. Nevertheless, it is important to address the differences in these findings so that there is improved reproducibility in the future.

B. DEG Analysis

DEGs are commonly filtered by the adjusted p-values and $\log_2$FC of 0.05 and 1, respectively. Using the typical filter resulted in 42 DEGs as demonstrated in Figure 8, but this number of DEGs is extremely low. Increasing the adjusted p-value to 0.1 slightly increased the number of DEGs to 44, but this number is still low. A possible explanation for this finding is that the mice were given a month to reacclimate after arriving back on Earth from the ISS. After a month or so, the skeletal system may have had time to restore itself [47]. If this study obtained the mice samples upon immediate return to Earth, it's possible that there could be more DEGs.

Due to the lack of DEGs found using the adjusted p-value and log$_2$FC of 0.05 and 1, a p-value of 0.01 and log$_2$FC of 1 were used to filter for the DEGs. A total number of 124 DEGs were found and these DEGs were used throughout this study. But by using the p-value rather than the adjusted p-value, false positives are a possibility. Nonetheless, the DEGs had differential expression levels on the clustered heatmap in Figure 9. In addition, the GC and FLT samples clustered within their respective groups, proving that the samples under different conditions varied with another.

C. Gene Enrichment Analysis

1) DAVID

a. Protein Transport and Binding

The first cluster in DAVID suggests that there are changes to protein binding and transporters. Some protein transporters that are mentioned in the first cluster are calycin and lipocalin. Calycin forms a large protein superfamily and can be divided into families including lipocalins and fatty-acid binding proteins [48]. Of the calycin superfamily, the lipocalins are the most diverse. Lipocalins are extracellular proteins that transport small hydrophobic molecules like steroids, for example. Some studies have associated lipocalins with osteoporosis. One mice study confirmed the association of lipocalins and bone loss. Lipocalin deficient mice of all ages had osteopenia and showed fewer osteoblasts compared to the wild-type mice [49]. The lipocalin deficient mice also had an increase in body weight and showed tolerance to glucose and insulin. This is surprising because Cluster 2 from the DAVID results also suggests that regulation of blood glucose is affected. As a whole, these findings show that that lipocalin plays a role in bone regeneration. Not only that but also the mechanisms involved in bone loss may be

- 34 -

more complex. However, the link between lipocalins and how it's involved in bone regeneration remains unclear.

Another interesting find, within the first DAVID cluster, is pheromone binding. This result coincides with the results found from gprofiler2. Figure 10 and Table II from the gprofiler2 results denote that the pheromone binding gene set is upregulated and has the highest gene ratio. Both findings strongly suggest that pheromone binding is impacted in mice with bone loss. A single study by Daev et. Al indicates this as well. In the study, a single-cell gel electrophoresis of mice exposed to pheromone 2,5-dimethylpyrazine for 4 to 24 hours has shown increases in the DNA damage during interphase in bone marrow cells [50]. This finding from [50] also shows that an increase in pheromone binding causes disturbances in mitosis which in turn play a role in bone loss. However, there are not enough studies to prove a relationship between pheromones and bone loss.

2) Urinary Proteins

Urinary proteins were found in the first and second cluster in Table II. This finding is noteworthy since urinary proteins have a known association with issues in blood sugar metabolism, which also appears in the first cluster of the DAVID results [51]–[54]. Two correlation studies among hospitalized patients have shown that the urine protein level and the incidence rate of osteoporosis in patients with diabetic nephropathy are significantly higher than those in patients only with diabetes. In one of the studies, about 35 % of the patients also had osteoporosis [51]. The patients with a higher urine albumin-to-creatinine ratio had a significantly higher incidence of osteoporosis compared to those with normal ratios of urine albumin-to-creatinine. The other reported a correlation between high levels of 24h urine protein (24hndb)

and inflammatory cytokines, which are proteins associated with diabetes and osteoporosis [52]. Therefore, these findings suggest that renal function could also affect osteoporosis.

3) Immunoglobins

The results from DAVID indicate that the immunoglobins are affected in outer space. But DAVID indicates that these gene sets are not significantly enriched because the FDR is 1 as indicated in Table II. Although, the previous studies found would suggest otherwise. The skeletal and immune systems work as a close unit as B-cells are created in the bone marrow [55]. In some cases, immune cells can induce changes in key factors of bone regenerators, thereby affecting bone health. However, the interaction between bone and the immune system is not straightforward and is largely unexplored.

D. Gprofiler2

Two unique findings from gprofiler2 were the upregulation and downregulation of mmu-miRNA and pchda7, respectively. Neither DAVID nor GSEA have reported these results. However, it's because DAVID and GSEA do not use the CORUM and miRNA databases as a source for gene annotations [29], [30].

1) Mmu-miRNA

Mmu-miRNA is a non-encoding mRNA found in M. *musculus*. These non-encoding mRNAs play a role in many diseases. In Gao et Al's review, there is a vast amount of studies on how different mmu-miRNAs upregulate and downregulated osteoclasts and osteoblasts shown in Tables 3, 4, and 5 in [56]. However, none of the tables in [56] mention the roles of mmu-miRNA 151-3P, 378B, and 463-3P found in grprofiler2. The only miRNA that has a possible connection to bone loss is Mmu-miRNA-151-3P. In a comprehensive miRNA sequencing analysis, Mmu-miRNA-151-3P was differentially expressed in the tibialis muscle between young and old mice

[57]. Therefore, it's possible that there is a connection between bone loss and skeletal muscle, but more studies are needed.

2) Pcdha7

The *pdch7* gene encodes the protein Pcdha7, which is a protocadherin in the cadherin superfamily [58]. Previously, the function of protocadherins was elusive, but recent studies have uncovered their roles. A study by Kim et Al. demonstrated the function Pcdha7 in bone regulation through the gene deletion of *Pcdh7* in mice [44]. The mice had increased bone mass and a reduced number of osteoclasts but not of osteoblasts. With in-vitro cell cultures, Kim et Al. also showed that the defect could be restored by transducing the *Pcdh7* gene. Given Kim et Al's findings, the downregulation of the *pcdha7* gene in space mice is plausible. Because the *pcdha7* gene is downregulated in a microgravity environment, the number of osteoclasts, which aids in bone removal, increases.

E. GSEA

The only significant downregulated gene sets were oxidative phosphorylation and DNA repair. While DNA repair did not coincide with the other results found from gprofiler2 and DAVID, oxidative phosphorylation was related to the results in DAVID. Oxidative phosphorylation is one of the common pathways for ATP generation and plays a role in glucose metabolism. During the Krebs cycle, pyruvate oxidation is coupled with oxidative phosphorylation to obtain ATP from glucose [59]. Oxidative phosphorylation also has a role in bone development. The generated ATP from oxidative phosphorylation provides the osteoblast with energy for bone formation [60]. At the same time, oxidative phosphorylation also stimulates

osteoclastogenesis. Osteoclasts differentiate through a macrophage stimulating factor (M-CSF) or a receptor activator of nuclear factor kappa-B ligand (RANKL)[61]. In RANKL osteoclast differentiation, glycolysis and oxidative phosphorylation increases. Glycolysis is the process where glucose is broken down to pyruvate to generate energy for the cell. Osteoclastogenesis is optimal at certain glucose levels, but at higher glucose levels osteoclastogenesis is reduced [62]. In addition, inhibiting oxidative phosphorylation impedes with osteoclastogenesis, indicating that osteoclast differentiation is a coupled process that involves both glycolysis and oxidative phosphorylation. Since oxidative phosphorylation is downregulated under microgravity conditions, it's possible that a metabolic dysregulation impedes the energy expenditure and formation of osteoblasts and osteoclasts. Therefore, one promising way to combat bone loss is to target the metabolic pathways. But the mechanisms involving the metabolic pathway are complex and would require further investigation.

V.   CONCLUSION

This project aimed to examine the factors involved in bone loss with the GLDS-241 dataset, which contains RNA sequencing data of femoral skin from mice. The dataset was analyzed using Nextflow-RCP, which is an implementation of NASA Genelab-RCP. Overall, both pipelines resulted in a similar number of DEGs and many of the DEGs were shared between the two pipelines. But, there are some discrepancies between the two pipelines that need to be addressed for future reproducibility.

Even though the skin samples were prepared a month after arrival on Earth from the ISS, the samples still showed differential expression in the genes. Nonetheless, the differential expression between the GC and FLT groups may show profound insights if the samples were immediately prepared.

Gene enrichment analysis was performed using three types of tools: DAVID, gprofiler2, and GSEA. Overall, the findings from the gene enrichment analysis suggest that there are local and systemic-wide factors that play a role in bone loss. Many previous studies examined these factors individually. But some studies have shown a correlation between osteoporosis and glucose metabolism. However, there has yet been a study that unifies all these factors to find a single mechanism that causes bone loss. Hopefully, the findings from this study shed some light to future studies. Most of all, this will hopefully help make space travel safer for everyone.

REFERENCES

[1] M. G. Smith, M. Kelley, and M. Basner, "A brief history of spaceflight from 1961 to 2020: An analysis of missions and astronaut demographics," *Acta Astronaut.*, vol. 175, pp. 290–299, Oct. 2020, doi: 10.1016/j.actaastro.2020.06.004.

[2] A. Hawkey, "The physical price of a ticket into space," *J. Br. Interplanet. Soc.*, vol. 56, no. 5–6, pp. 152–159, Jun. 2003.

[3] R. Cappellesso, L. Nicole, A. Guido, and D. Pizzol, "Spaceflight osteoporosis: current state and future perspective," *Endocr. Regul.*, vol. 49, no. 04, pp. 231–239, 2015, doi: 10.4149/endo_2015_04_231.

[4] M. F. Holick, "Perspective on the Impact of Weightlessness on Calcium and Bone Metabolism," *Bone*, vol. 22, no. 5, pp. 105S-111S, May 1998, doi: 10.1016/S8756-3282(98)00014-3.

[5] D. J. Hadjidakis and I. I. Androulakis, "Bone remodeling," *Ann. N. Y. Acad. Sci.*, vol. 1092, pp. 385–396, Dec. 2006, doi: 10.1196/annals.1365.035.

[6] L. J. Raggatt and N. C. Partridge, "Cellular and Molecular Mechanisms of Bone Remodeling *," *J. Biol. Chem.*, vol. 285, no. 33, pp. 25103–25108, Aug. 2010, doi: 10.1074/jbc.R109.041087.

[7] I. Akkawi and H. Zmerly, "Osteoporosis: Current Concepts," *Joints*, vol. 6, no. 2, pp. 122–127, Jun. 2018, doi: 10.1055/s-0038-1660790.

[8] J. Sirola *et al.*, "Factors affecting bone loss around menopause in women without HRT: a prospective study," *Maturitas*, vol. 45, no. 3, pp. 159–167, Jul. 2003, doi: 10.1016/S0378-5122(03)00150-6.

[9] T. S. Keller and A. M. Strauss, "Predicting skeletal adaptation in altered gravity environments," *J. Br. Interplanet. Soc.*, vol. 46, no. 3, pp. 87–96, Mar. 1993.

[10] M. Iki *et al.*, "Age, menopause, bone turnover markers and lumbar bone loss in healthy Japanese women," *Maturitas*, vol. 25, no. 1, pp. 59–67, Aug. 1996, doi: 10.1016/0378-5122(96)01042-0.

[11] A. LeBlanc *et al.*, "Bone mineral and lean tissue loss after long duration space flight," *J. Musculoskelet. Neuronal Interact.*, vol. 1, no. 2, pp. 157–160, Dec. 2000.

[12] J. Oribello, "Differential Gene Expression Analysis of Rodents Exposed to Long-Term Space Flight and Insights into Physiological Effects," *Masters Proj.*, Jan. 2021, doi: https://doi.org/10.31979/etd.d4e7-x2g3.

[13] "NASA GeneLab Data Systems : /genelab/accession/GLDS-241/." https://genelab-data.ndc.nasa.gov/genelab/accession/GLDS-241/ (accessed Nov. 03, 2021).

[14] "A DSL for parallel and scalable computational pipelines | Nextflow." https://www.nextflow.io/index.html#Features (accessed Nov. 12, 2021).

[15] "NASA GeneLab: Open Science for Life in Space," *NASA GeneLab*. https://131.110.128.77/ (accessed Nov. 12, 2021).

[16] "NovaSeq 6000 Sequencing System," p. 4.

[17] "TruSeq Stranded Total RNA Reference (1000000040499)," p. 45.

[18] "HPC Cluster." http://coe-hpc-web.sjsu.edu/ (accessed Nov. 20, 2021).

[19] "Babraham Bioinformatics - FastQC A Quality Control tool for High Throughput Sequence Data." https://www.bioinformatics.babraham.ac.uk/projects/fastqc/ (accessed Nov. 13, 2021).

[20] P. Ewels, M. Magnusson, S. Lundin, and M. Käller, "MultiQC: summarize analysis results for multiple tools and samples in a single report," *Bioinformatics*, vol. 32, no. 19, pp. 3047–3048, Oct. 2016, doi: 10.1093/bioinformatics/btw354.

[21] "Babraham Bioinformatics - Trim Galore!" https://www.bioinformatics.babraham.ac.uk/projects/trim_galore/ (accessed Nov. 13, 2021).

[22] A. Dobin *et al.*, "STAR: ultrafast universal RNA-seq aligner," *Bioinformatics*, vol. 29, no. 1, pp. 15–21, Jan. 2013, doi: 10.1093/bioinformatics/bts635.

[23] K. L. Howe *et al.*, "Ensembl 2021," *Nucleic Acids Res.*, vol. 49, no. D1, pp. D884–D891, Jan. 2021, doi: 10.1093/nar/gkaa942.

[24] B. Li and C. N. Dewey, "RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome," *BMC Bioinformatics*, vol. 12, no. 1, p. 323, Aug. 2011, doi: 10.1186/1471-2105-12-323.

[25] M. Love *et al.*, *DESeq2: Differential gene expression analysis based on the negative binomial distribution*. Bioconductor version: Release (3.14), 2021. doi: 10.18129/B9.bioc.DESeq2.

[26] R. Bourgon, R. Gentleman, and W. Huber, "Independent filtering increases detection power for high-throughput experiments," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 107, no. 21, pp. 9546–9551, May 2010, doi: 10.1073/pnas.0914005107.

[27] S. Anders and W. Huber, "Differential expression analysis for sequence count data," *Genome Biol.*, vol. 11, no. 10, p. R106, Oct. 2010, doi: 10.1186/gb-2010-11-10-r106.

[28] Y. Benjamini and Y. Hochberg, "Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing," *J. R. Stat. Soc. Ser. B Methodol.*, vol. 57, no. 1, pp. 289–300, 1995.

[29] D. W. Huang, B. T. Sherman, and R. A. Lempicki, "Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources," *Nat. Protoc.*, vol. 4, no. 1, pp. 44–57, 2009, doi: 10.1038/nprot.2008.211.

[30] A. Subramanian *et al.*, "Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles," *Proc. Natl. Acad. Sci.*, vol. 102, no. 43, pp. 15545–15550, Oct. 2005, doi: 10.1073/pnas.0506580102.

[31] V. K. Mootha *et al.*, "PGC-1α-responsive genes involved in oxidative phosphorylation are coordinately downregulated in human diabetes," *Nat. Genet.*, vol. 34, no. 3, pp. 267–273, Jul. 2003, doi: 10.1038/ng1180.

[32] L. Kolberg, U. Raudvere, I. Kuzmin, J. Vilo, and H. Peterson, "gprofiler2 -- an R package for gene list functional enrichment analysis and namespace conversion toolset g:Profiler," *F1000Research*, vol. 9, p. ELIXIR-709, Nov. 2020, doi: 10.12688/f1000research.24956.2.

[33] U. Raudvere *et al.*, "g:Profiler: a web server for functional enrichment analysis and conversions of gene lists (2019 update)," *Nucleic Acids Res.*, vol. 47, no. W1, pp. W191–W198, Jul. 2019, doi: 10.1093/nar/gkz369.

[34] M. Kanehisa, Y. Sato, M. Furumichi, K. Morishima, and M. Tanabe, "New approach for understanding genome variations in KEGG," *Nucleic Acids Res.*, vol. 47, no. D1, pp. D590–D595, Jan. 2019, doi: 10.1093/nar/gky962.

[35] A. Fabregat *et al.*, "The Reactome Pathway Knowledgebase," *Nucleic Acids Res.*, vol. 46, no. D1, pp. D649–D655, Jan. 2018, doi: 10.1093/nar/gkx1132.

[36] D. N. Slenter *et al.*, "WikiPathways: a multifaceted pathway database bridging metabolomics to other omics research," *Nucleic Acids Res.*, vol. 46, no. D1, pp. D661–D667, Jan. 2018, doi: 10.1093/nar/gkx1064.

[37] A. Ruepp *et al.*, "CORUM: the comprehensive resource of mammalian protein complexes--2009," *Nucleic Acids Res.*, vol. 38, no. Database issue, pp. D497-501, Jan. 2010, doi: 10.1093/nar/gkp914.

[38] M. Ashburner *et al.*, "Gene ontology: tool for the unification of biology. The Gene Ontology Consortium," *Nat. Genet.*, vol. 25, no. 1, pp. 25–29, May 2000, doi: 10.1038/75556.

[39] C.-H. Chou *et al.*, "miRTarBase update 2018: a resource for experimentally validated microRNA-target interactions," *Nucleic Acids Res.*, vol. 46, no. D1, pp. D296–D302, Jan. 2018, doi: 10.1093/nar/gkx1067.

[40] M. Uhlén *et al.*, "Proteomics. Tissue-based map of the human proteome," *Science*, vol. 347, no. 6220, p. 1260419, Jan. 2015, doi: 10.1126/science.1260419.

[41] H. Wickham *et al.*, *ggplot2: Create Elegant Data Visualisations Using the Grammar of Graphics*. 2021. Accessed: Nov. 16, 2021. [Online]. Available: https://CRAN.R-project.org/package=ggplot2

[42] C. Sievert *et al.*, *plotly: Create Interactive Web Graphics via "plotly.js."* 2021. Accessed: Nov. 16, 2021. [Online]. Available: https://CRAN.R-project.org/package=plotly

[43] S. Zhang, B. Wang, L. Wan, and L. M. Li, "Estimating Phred scores of Illumina base calls by logistic regression and sparse modeling," *BMC Bioinformatics*, vol. 18, p. 335, Jul. 2017, doi: 10.1186/s12859-017-1743-4.

[44] "PCDHA7 protocadherin alpha 7 [Homo sapiens (human)] - Gene - NCBI." https://www.ncbi.nlm.nih.gov/gene/56141 (accessed Nov. 17, 2021).

[45] J. O'Brien, H. Hayder, Y. Zayed, and C. Peng, "Overview of MicroRNA Biogenesis, Mechanisms of Actions, and Circulation," *Front. Endocrinol.*, vol. 9, p. 402, 2018, doi: 10.3389/fendo.2018.00402.

[46] A. Liberzon, C. Birger, H. Thorvaldsdóttir, M. Ghandi, J. P. Mesirov, and P. Tamayo, "The Molecular Signatures Database Hallmark Gene Set Collection," *Cell Syst.*, vol. 1, no. 6, pp. 417–425, Dec. 2015, doi: 10.1016/j.cels.2015.12.004.

[47] D. Williams, A. Kuipers, C. Mukai, and R. Thirsk, "Acclimation during space flight: effects on human physiology," *CMAJ Can. Med. Assoc. J.*, vol. 180, no. 13, pp. 1317–1323, Jun. 2009, doi: 10.1503/cmaj.090628.

[48] D. R. Flower, A. C. T. North, and C. E. Sansom, "The lipocalin protein family: structural and sequence overview," *Biochim. Biophys. Acta BBA - Protein Struct. Mol. Enzymol.*, vol. 1482, no. 1, pp. 9–24, Oct. 2000, doi: 10.1016/S0167-4838(00)00148-5.

[49] M. Capulli *et al.*, "A Complex Role for Lipocalin 2 in Bone Metabolism: Global Ablation in Mice Induces Osteopenia Caused by an Altered Energy Metabolism," *J. Bone Miner. Res. Off. J. Am. Soc. Bone Miner. Res.*, vol. 33, no. 6, pp. 1141–1153, Jun. 2018, doi: 10.1002/jbmr.3406.

[50] E. V. Daev, M. V. Petrova, L. S. Onopa, V. A. Shubina, and T. S. Glinin, "DNA damage in bone marrow cells of mouse males in vivo after exposure to the pheromone: Comet assay," *Russ. J. Genet.*, vol. 53, no. 10, pp. 1105–1112, Oct. 2017, doi: 10.1134/S1022795417100027.

[51] Y. Y. Lee *et al.*, "The Association between Urine Albumin to Creatinine Ratio and Osteoporosis in Postmenopausal Women with Type 2 Diabetes," *J. Bone Metab.*, vol. 23, no. 1, pp. 1–7, Feb. 2016, doi: 10.11005/jbm.2016.23.1.1.

[52] Z. Zhao, "Correlation analysis of urine proteins and inflammatory cytokines with osteoporosis in patients with diabetic nephropathy," *J. Musculoskelet. Neuronal Interact.*, vol. 18, no. 3, pp. 348–353, Sep. 2018.

[53] G. N. Thomas, J. a. J. H. Critchley, B. Tomlinson, C. S. Cockram, and J. C. N. Chan, "Peripheral vascular disease in Type 2 diabetic Chinese patients: associations with metabolic indices, concomitant vascular disease and genetic factors," *Diabet. Med. J. Br. Diabet. Assoc.*, vol. 20, no. 12, pp. 988–995, Dec. 2003, doi: 10.1046/j.1464-5491.2003.01046.x.

[54] I. Kanazawa, T. Yamaguchi, K. Hayashi, H. Takase, T. Shimizu, and T. Sugimoto, "Effects of treatment with risedronate and alfacalcidol on progression of atherosclerosis in postmenopausal women with type 2 diabetes mellitus accompanied with osteoporosis," *Am. J. Med. Sci.*, vol. 339, no. 6, pp. 519–524, Jun. 2010, doi: 10.1097/MAJ.0b013e3181db6dfe.

[55] R. K. Srivastava, H. Y. Dar, and P. K. Mishra, "Immunoporosis: Immunology of Osteoporosis—Role of T Cells," *Front. Immunol.*, vol. 9, p. 657, 2018, doi: 10.3389/fimmu.2018.00657.

[56] Y. Gao, S. Patil, and A. Qian, "The Role of MicroRNAs in Bone Metabolism and Disease," *Int. J. Mol. Sci.*, vol. 21, no. 17, p. 6081, Aug. 2020, doi: 10.3390/ijms21176081.

[57] H. J. Jung *et al.*, "Comprehensive miRNA Profiling of Skeletal Muscle and Serum in Induced and Normal Mouse Muscle Atrophy During Aging," *J. Gerontol. Ser. A*, vol. 72, no. 11, pp. 1483–1491, Oct. 2017, doi: 10.1093/gerona/glx025.

[58] S. Hayashi and M. Takeichi, "Emerging roles of protocadherins: from self-avoidance to enhancement of motility," *J. Cell Sci.*, vol. 128, no. 8, pp. 1455–1464, Apr. 2015, doi: 10.1242/jcs.166306.

[59] T. O. Alabduladhem and B. Bordoni, "Physiology, Krebs Cycle," in *StatPearls*, Treasure Island (FL): StatPearls Publishing, 2021. Accessed: Nov. 18, 2021. [Online]. Available: http://www.ncbi.nlm.nih.gov/books/NBK556032/

[60] W.-C. Lee, A. R. Guntur, F. Long, and C. J. Rosen, "Energy Metabolism of the Osteoblast: Implications for Osteoporosis," *Endocr. Rev.*, vol. 38, no. 3, pp. 255–266, May 2017, doi: 10.1210/er.2017-00064.

[61] S. Lemma *et al.*, "Energy metabolism in osteoclast formation and activity," *Int. J. Biochem. Cell Biol.*, vol. 79, pp. 168–180, Oct. 2016, doi: 10.1016/j.biocel.2016.08.034.

[62] J.-M. Kim, D. Jeong, H. K. Kang, S. Y. Jung, S. S. Kang, and B.-M. Min, "Osteoclast precursors display dynamic metabolic shifts toward accelerated glucose metabolism at an early stage of RANKL-stimulated osteoclast differentiation," *Cell. Physiol. Biochem. Int. J. Exp. Cell. Physiol. Biochem. Pharmacol.*, vol. 20, no. 6, pp. 935–946, 2007, doi: 10.1159/000110454.