Spring 2021

# Differential Gene Expression Analysis of Mice Thymi After Spaceflight

Roshani Codipilly
*San Jose State University*

# Differential Gene Expression Analysis of Mice Thymi After Spaceflight

A Project Report

Presented to

Department of Computer Science

San José State University

In Partial Fulfillment

Of the Requirements for the Degree

By

Roshani Codipilly

May, 2021

# ABSTRACT

With increases in space travel and a desire to inhabit the moon and Mars comes a pressing need to understand the impact of spaceflight on the body. Some effects are already known, such as reduced cardiac function and bone loss, but one area that needs to be further explored is the immune system. Differential gene expression analysis of mice thymi was performed to determine the impact of spaceflight on the immune system. The dataset that was analyzed, GLDS-289, was obtained from GeneLab, a space-omics database developed by NASA. Differential gene expression analysis was accomplished using a Nextflow implementation of GeneLab's RNA-Seq Consensus Pipeline. The results showed that microgravity has a significant effect on multiple cellular processes, such as regulation of the cell cycle and DNA organization. These changes in gene expression reduce the proliferation of new immune cells, hindering the immune response and resulting in a compromised immune system that is more prone to infection. Artificial gravity partially mitigates the impact of microgravity, but it does not completely rescue the body from the effects of spaceflight. Further research into the effect of microgravity on the immune system must be done before humans can safely inhabit the moon and beyond.



**Keywords:** differential gene expression, microgravity, immune system

# ACKNOWLEDGEMENTS

I would like the following people for their support:

My advisor, Dr. Heller, for his endless support and guidance throughout this master's program and final project.

The rest of my committee, Dr. Andreopoulos and Dr. Lee, for their valuable insight and feedback.

Jonathan Oribello, for his counsel and patience in my endless questions regarding his implementation.

Dr. Amanda Saravia-Butler, for her help and assistance with the RCP and DGE analysis.

The rest of the MSBI students, new and graduated, who have been such a supportive community, especially during these socially distant times.

My parents, for instilling in me the value of education and continuously supporting me in all my endeavors.

San José State University
Department of Computer Science

We hereby approve the project of

Roshani Codipilly

Candidate for the degree of Master of Science, Bioinformatics

Philip Heller, Ph.D.

Assistant Professor, Department of Computer Science, Advisor

Wendy Lee, Ph.D.

Assistant Professor, Department of Computer Science

William Andreopoulos, Ph.D.

Assistant Professor, Department of Computer Science

# Table of Contents

# Table of Figures

# Index of Tables

# I. Introduction

## *A. The Future of Space Travel and Effects on the Body*

In 1961, Soviet cosmonaut Yuri Gagarin became the first person to travel into outer space. Since then, around 480 individuals have followed him, and there has been continual human presence in space since October 31, 2000 [1]. Although impressive, this is just a fraction of what humankind hopes to achieve with space travel. As science and technology rapidly advance, so does the potential for space exploration. NASA is planning to send the first woman and a man to the Moon by 2024 and embark on the first voyage to Mars by the end of the decade [2]. The development of private spaceflight also expands space travel to civilians, with SpaceX planning to send four private citizens into space by the end of 2021 or early 2022 [3]. With this increase in space travel and desire to inhabit space comes a pressing need to learn the effects of spaceflight on the body. Some effects of microgravity on the human body are already known, such as bone loss, diminished cardiac function due to reduced cardiac muscle mass, and circadian rhythm issues resulting in sleep problems, but one area that needs to be further investigated is the immune system [4].

## *B. The Immune System and the Thymus*

The immune system defends the body against infection and consists of both an innate and adaptive response. The innate response is the first line of defense and its response is always the same, regardless of how many times the body is exposed to the same pathogen. It consists of physical barriers such as the skin and immune cells that attack foreign invaders. The adaptive response is more complex and specialized; it is slower but longer lasting than the

innate response. The adaptive response develops specific defenses catered to the pathogen. Additionally, the adaptive response has “memory”, becoming stronger with each exposure to a particular pathogen [5]. This response is carried out by two main classes of cells, B cells and T cells. B cells are important for the recognition of specific antigens, which are structures found on the surface of a pathogen. B cells then produce antibodies, which help remove the pathogens from the body. T cells are able to help B cells secrete antibodies or directly kill foreign cells [6]. The thymus is a critically important organ for the immune system since it is the site of T cell development. Both kinds of cell begin as pluripotent stem cells in the fetal liver and the bone marrow; then B cells reach maturation in the bone marrow while T cells reach maturation in the thymus [5]. Different types of T cells are distinguished by CD antigens, which are found on the surface of B and T cells. The two main types are CD4 and CD8 T cells. CD4 T cells, also known as helper T cells, send signals to B cells to make antibodies and help develop killer T cells. CD8 T cells, also known as killer T cells, can directly kill foreign cells [5].

### *C. Spaceflight Effects on the Immune System*

Spaceflight results in numerous detrimental effects to the immune system. Thymic mass significantly decreases as a result of spaceflight, reducing by approximately 50% [7]–[9]. Furthermore, the thymi masses did not recover in the 7 days after landing and continued to decrease, suggesting that the effects of spaceflight on the thymus are long-term [9]. Decreased thymus mass may result in reduced immune response capabilities, but more research into the effect of reduced thymic size on immune function should be performed to elucidate this impact [7].

Spaceflight has a multitude of effects on the development of the cell, from cell proliferation to differentiation, and all these disruptions have a deleterious impact on the immune system. One impact of spaceflight on the thymus is reduction in cell proliferation [7], [10], [11]. Cell proliferation is the process in which a cell grows and divides into two daughter cells, which is an important mechanism for tissue growth [12]. Reduction in cell proliferation hinders the immune system due to a decrease in T cells which depresses the adaptive immune response. An additional effect that spaceflight has on the cell cycle is the halting of cell cycle progression, which results in the dysregulation of T cell maturation [7], [13]. Spaceflight has also been found to prevent cell differentiation, which is the process in which a cell changes from one type to another [7], [8]. The process of T cell development is seen in Figure 1. Pluripotent stem cells first develop in the bone marrow and then migrate to the cortex, which is the outer portion of the thymus. Within the cortex, early T cell development occurs, with the pluripotent cells differentiating into CD4- CD8- double negative cells (DN). DN cells then develop into CD4+ CD8+ double positive (DP) cells before traveling into the medulla and developing into single positive (SP) mature T cells, CD4+ or CD8+. Lebsack et al. found that environment-induced arrest occurred after the DN stage but just before DP stage, meaning that differentiation of thymocytes into mature T cells was not occurring [11]. This inability to differentiate results in decreased T cell maturation, hindering the immune system's ability to respond to infection and foreign invaders.

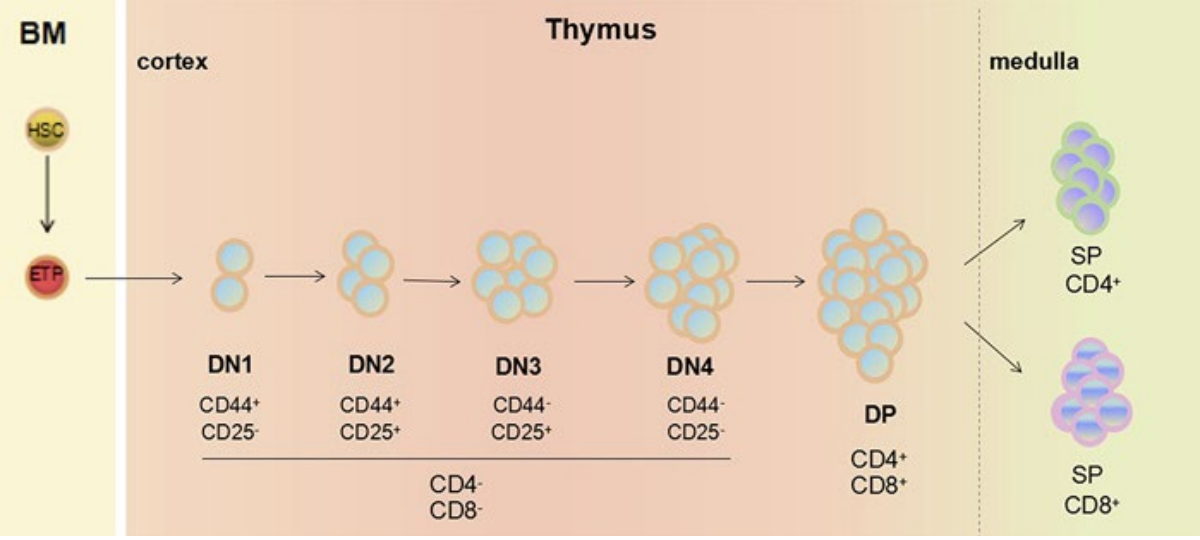


Figure 1. T cell development from pluripotent stem cell into mature T cell. Adapted from [14].

Prolonged exposure to microgravity also results in increased apoptosis, which is cell death [7], [11], [13]. Apoptosis is an essential component of the cell cycle, serving as a quality control mechanism for a cell to utilize if it does not properly develop. Uncontrolled regulation of apoptosis can result in developmental abnormalities or diseases [14]. Increased apoptotic thymic cell death due to the inability to regulate apoptosis causes serious damage to the immune system.

Alterations to the immune system lead to a compromised defense against infections, diseases and tumors [10]. Analysis of differential gene expression in astronauts who recently returned from space found that a decrease in the immune system's adaptive response led to higher susceptibility to disease. Out of 29 Apollo astronauts, 15 reported bacterial or viral infection during, immediately after, or within 1 week of landing back on Earth. Not only does a weakened immune system result in higher rates of disease, but it makes astronauts vulnerable

to uncommon pathogens as well. One astronaut on Apollo 13 contracted *Pseudomonas aeruginosa* and suffered from intense chills and fever. This pathogen rarely causes disease unless the person suffers from a break in epithelia or from immune suppression, which means that weakened immune responses due to spaceflight make astronauts vulnerable to previously unproblematic pathogens [15]. Another study found the reactivation of *Varcella zoster* virus (VZV) in astronauts during and after space flight. This virus is latent after primary infection, and typically reactivates in people with weakened immune systems, such as elderly individuals, cancer patients, organ transplant recipients, or patients with AIDS. The presence of VZV in astronauts suggests that space flight seriously damages immune responses [16]. Before colonizing the moon, it needs to be ensured that it is safe for humans to live there, and that will not be possible until the effects of spaceflight on the immune system are fully understood.

### *D. NASA GeneLab RNA-Seq Consensus Pipeline*

The objective of this project is to use an RNA Sequencing ("RNA-Seq") pipeline to analyze the impact of spaceflight on gene expression in the thymus. The dataset was obtained from GeneLab, a multi-omics space-related database curated by NASA [17]. The pipeline used was the NASA GeneLab RNA-Seq Consensus Pipeline (RCP), which analyzes short-read RNA-sequencing data to detect differential gene expression in space-related experiments. The major steps of the pipeline are quality control of reads, trimming, mapping to a reference genome, and gene quantification. In the end, detection of differentially expressed genes (DEGs) is reported [18]. A previous MS Bioinformatics student, Jonathan Oribello, implemented the workflow of this pipeline on Nextflow, a bioinformatics workflow manager [19]. Using Jonathan's implementation, a script was run that pulled the raw RNA reads from GeneLab.

Nextflow Tower, a web application for monitoring Nextflow pipelines, was used to observe the progression of the pipeline [20]. Upon completion of the pipeline, an unnormalized gene count matrix was returned and differential gene expression was performed in R using the DESeq2 package.

### *E. GLDS-289 Dataset*

The dataset chosen for this project was GeneLab Dataset 289 (GLDS-289), in which the gene expression of mice on board the International Space Station (ISS) was analyzed using transcription profiling of the thymus [21]. This experiment was carried out by the Japan Aerospace Exploration Agency (JAXA) and was conducted on two separate Space missions, MHU-1 and MHU-2, with the differences highlighted in Table I. For each mission, 6 C57BL/6 J male mice were housed on board the ISS. JAXA developed an experimental platform known as the Multiple Artificial-gravity Research System (MARS) which centrifuged the mouse cages at 1 x *g* as a countermeasure to the effects of weightlessness. Upon arrival to the ISS the 6 mice were housed in MARS, with 3 of them in centrifuged cages to study the impact of artificial gravity (AG) while the remaining 3 were exposed to microgravity (MG). An additional 3 mice were housed on Earth for the duration of the space mission as ground control (GC). After returning to Earth and following euthanasia, thymi were excised and cut for RNA preparation. Paired-end sequencing was performed using Illumina NextSeq500 [7], [21].

**TABLE I. DIFFERENCES BETWEEN SPACEFLIGHTS MHU-1 AND MHU-2**

| | Age of Mice at Launch | Flight Launch Date | Spaceflight Duration | Time of Sacrifice after return |
|---|---|---|---|---|
| **MHU-1** | 8 weeks old | 7/18/2016 | 35 days | 48 hours |
| **MHU-2** | 9 weeks old | 8/14/2017 | 30 days | 36.5 hours |

## II. Methods

### *A. High Level Overview of Workflow*

The RCP was performed on SJSU's COS-HPC using Jonathan's Nextflow implementation. In order to adapt his implementation for my own use, a config file referring to my dataset of interest (GLDS-289) was used. This config file directs to the GeneLab GLDS-289 repository page and pulls the raw paired end reads to be processed by the pipeline. An overview of the pipeline workflow is illustrated in Figure 2. The workflow is as follows: the raw paired end read data is first downloaded from GeneLab, then initial quality of the reads is checked using FastQC. Trim Galore is then used to remove adapters, low quality reads, and short reads. MultiQC then checks the quality of the trimmed reads across all samples at once to compare to the initial raw sequence data quality. After accessing the quality of the trimmed reads, the reads are aligned to the reference genome for their species. In order to build and annotate the reference genome, the mouse genome fasta and gene transfer format (GTF) files are downloaded from ENSEMBL and assembled using STAR. Trimmed reads are then aligned to the reference genome using STAR and RSEM is used to quantify gene counts. In the end, an unnormalized gene counts matrix for all the annotated genes found in sequencing is returned [18]. This gene count matrix

was then normalized and analyzed using an adapted DESeq2 script from Dr. Amanda Saravia-Butler, a scientist from GeneLab.

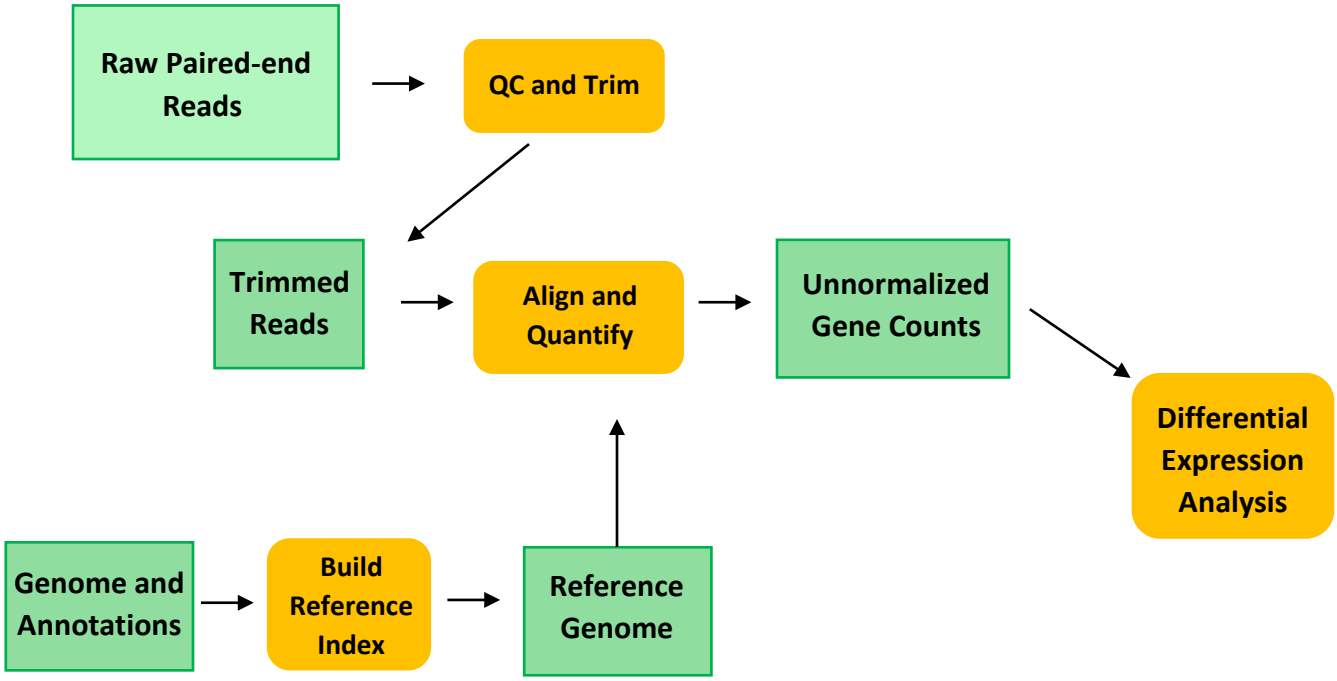


Figure 2. High Level Overview of NASA's RNA-Seq Consensus Pipeline. Green boxes represent files and yellow rounded rectangles represent processes. Adapted from [19].

*B. Normalization and Differential Expression Analysis using DESeq2*

Normalization of gene counts and gene dispersion was performed before proceeding to different expression analysis. Genes were filtered so that genes with count sums less than 10 across all 18 samples were removed. These genes have very low expression and would be uninformative in determining the probability of differential gene expression between the control and treatment groups. Size factor estimation is then done to correct for sample-wise differences in read depth. Read depth refers to the number of times a particular nucleotide is covered from all the short reads that were sequenced [22]. It is important to normalize for read

depth, otherwise it would not be possible to determine if differences in a gene's expression level between samples are due to biological reasons or variations in the sequencing depth. Size factor estimation was accomplished using the Median of Ratios method [23]. The first step of this method is to calculate a pseudo-reference sample of each gene, which is the geometric mean across all the samples. Then for each gene a ratio of the sample/pseudo-reference is calculated and the median value of all the ratios for a particular sample is used as a size factor for that sample. Once this normalization factor is found, each raw gene count for a sample is divided by the sample's size factor in order to normalize the gene counts [24].

The next step in normalization is estimating gene dispersions, which is the amount of variation in a gene's expression across all samples. It is especially important to normalize for gene expression in experiments that only have 2 to 3 replicates per group, since dispersion is typically more variable. Gene wise dispersions are estimated by sharing information across genes, making the assumption that genes with similar expression levels will have similar dispersions. By sharing information across genes rather than looking at each gene dispersion individually, unreliable dispersion estimates due to low mean counts are avoided. Maximum likelihood estimation is the method used to estimate the dispersion of each gene. This method is done by calculating the most likely estimate of the dispersion from the gene counts of all the replicates in a group [23]. After estimation is completed, the gene-wise dispersions are then adjusted to model the read depth normalized counts that were found for each gene.

Once normalization of read depth and dispersion are complete, hypothesis testing can begin. For RNA-Seq experiments, the null hypothesis is that gene expression is not significantly changed between the experimental groups and the control, and the alternative hypothesis

would be that gene expression is changed significantly between the groups. To perform hypothesis testing for this experiment, the Wald test was used. In order to perform a Wald test, a mathematical model must be selected that fits the distribution of the gene counts data. For RNA-Seq data, the mathematical model most often used is the negative binomial model, since the majority of genes have expression values equal or close to 0 [23]. Now that normalization is complete and hypothesis testing has been performed, differential gene expression analysis can begin.

When performing differential expression analysis on a large number of genes, adjustments must be made to account for the large data set before the normalized data can be interpreted. Each gene has its own hypothesis test, so for this experiment there are approximately 20,000 hypothesis tests occurring simultaneously. This large number of hypothesis tests increases the likelihood of a false discovery; this is known as the multiple testing problem. To correct for the multiple testing problem, the Benjamini-Hochberg procedure was used, which reduces false discovery rate for large numbers of independent hypothesis tests. This procedure works by controlling the false discovery rate below a given significance level through a simple Bonferroni-type procedure [25]. In the end each gene had a p-value calculated using the Wald Test and an adjusted p-value calculated using the Benjamini-Hochberg procedure before proceeding to differential gene expression analysis.

### *C. Gene Set Analysis*

Due to the high number of genes, it would be inefficient to look at the DEGs individually. It is much more informative to perform gene set analysis, highlighting affected pathways and grouping genes into functional groups in order to understand how biological processes are affected [26]. The tools used to perform gene set analysis are DAVID, GOrilla, and Gene Set Enrichment Analysis (GSEA), with visualization accomplished using Cytoscape.

#### *1. DAVID Gene Set Clustering*

DAVID Bioinformatics Resources, version 6.8, is a database that provides functional annotation tools for the analysis of large lists of genes in order to extract biological meaning [27]. Two lists of genes were separately uploaded, the genes differentially expressed in GC vs MG for MHU-1 and the genes differentially expressed in GC vs MG for MHU-2. Using the functional annotation tool, gene-term enrichment analysis was then performed in order to group overlapping genes into functional clusters, allowing for the identification of significantly enriched biological processes [28].

#### *2. GOrilla*

Gene Ontology (GO) terms are general gene descriptions that allow for the assigning of multiple genes of a similar function to a single GO term, linking similar genes together. The three types of ontologies covered by GO are cellular component (CP), molecular function (MF), and biological process (BP) [29]. The grouping of similar genes by ontology simplifies differential gene expression analysis by clearly highlighting affected biological pathways, processes, and cellular components. GOrilla is a gene ontology tool that identifies and visualizes enriched GO

terms given a target gene list and a background gene list [30]. Two lists were uploaded: a list of Ensembl gene IDs that were differentially expressed in both GC vs MG groups from MHU-1 and MHU-2 as a target gene list and a list of all the annotated genes found during RNA sequencing as the background list. A list of enriched GO BP terms was returned in descending order by p-value and FDR, with a p-value threshold of 0.01.

3. *Gene Set Enrichment Analysis*

Gene Set Enrichment Analysis (GSEA) is a tool that determines the enriched gene sets and expression difference between 2 biological states [31]. Unlike DAVID, it looks at the entire list of genes and the expression values in order to determine significantly enriched gene sets, not just the significantly differentially expressed genes. This allows for the identification of changes across a pathway or gene sets that may have not been otherwise identified. For example, if only one gene in a pathway is significantly differentially expressed, then that may not seem significant on its own, but if most genes in a pathway are differentially expressed, even if the change in expression is not significant, it could indicate an affected pathway. Since GSEA considers both the genes and the expression values, it is also able to calculate an enrichment score (ES) and a normalized enrichment score (NES) each gene set. GSEA version 4.1.0 was used to find the most significantly enriched upregulated and downregulated gene sets in MG vs GC. GSEA was run independently for the MG vs GC groups in MHU-1 and MHU-2 and then the gene sets that were significantly enriched in both analyses with nominal p-values less than 0.05 and an FDR less than 0.25 were identified. According to the GSEA documentation, an FDR cutoff of 0.25 was used instead of the traditional value of 0.05 due to the lack of overlap

in expression datasets and a relatively small number of gene sets. Therefore, a more lenient FDR avoids the possibility of overlooking potentially significant results.

*4. Cytoscape Enrichment Map of GO Biological Processes Terms*

Cytoscape is an open-source software platform for visualizing networks of gene sets, and is helpful for extracting meaningful biological insights from large datasets [32]. Cytoscape, version 3.8.2, has multiple different applications that can be used for the visualization of networks. The ClueGO plugin, version 2.5.7, was used to interpret the biological significance of a given gene list by choosing represented GO terms from selected ontologies and visualizing them as a functional network [33]. A list of the ENSEMBL gene IDs of differentially expressed genes present in both MHU-1 and MHU-2 was uploaded, and the two ontologies selected were biological processes and the immune system due to interest in how both overall biological function and function specific to the immune system were affected.

## III. Results

### *A. Quality Control of Untrimmed and Trimmed Reads*

MultiQC reports for the untrimmed and trimmed paired end reads were assessed to determine the quality of the reads before and after trimming [34]. Figure 3 presents the mean quality Phred scores per base for the untrimmed and trimmed reads. The untrimmed reads, left, had Phred scores of 30 and above for all bases, indicating high confidence in the base calls. A dip in the quality score is only seen towards the end of the reads at position 35, but the Phred scores are still above 30 so the quality of the base call is still acceptable. The trimmed reads, right, had a higher quality score at position 35, indicating an improvement in the mean quality scores after trimming.

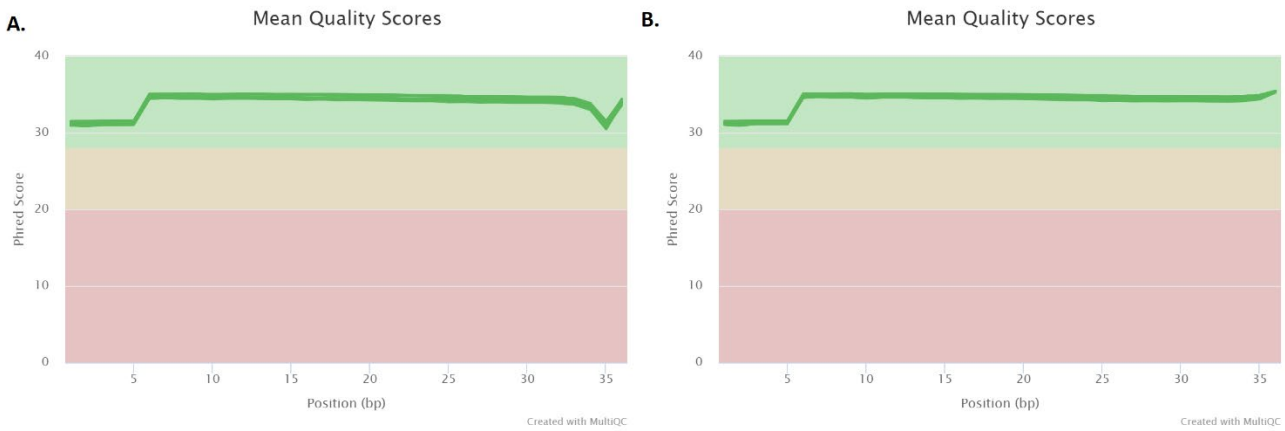


Figure 3. MultiQC Mean Quality Score per base. All the quality scores have a Phred score of 30 or above which indicates high confidence in the base calls.
(A)Untrimmed Reads (B) Trimmed Reads

MultiQC also reported that the per GC sequence content for both the untrimmed and trimmed reads had roughly normal distributions, with a majority of the reads having a per sequence GC content of around 50%. Another reported metric by MultiQC was the per base N content, seen in Figure 4. For the untrimmed reads, left, roughly 11% of bases at position 35 were substituted with N, indicating some uncertainty in the base call at this position. This uncertainty at position 35 also supports the drop in the mean quality score seen at the same position in Figure 3a. For the trimmed reads, right, there is close to 0% of bases called as N across the whole read length, indicating high confidence in the base calls. One other notable metric reported by MultiQC was the adapter content. For both the untrimmed and trimmed reads, no sample had any adapter contamination greater than 0.1%, which indicated successful removal of adapters. When looking at the per base sequence content, there was some biased sequence composition seen for the first 9 bases, but this not unusual at the start of a read [35].

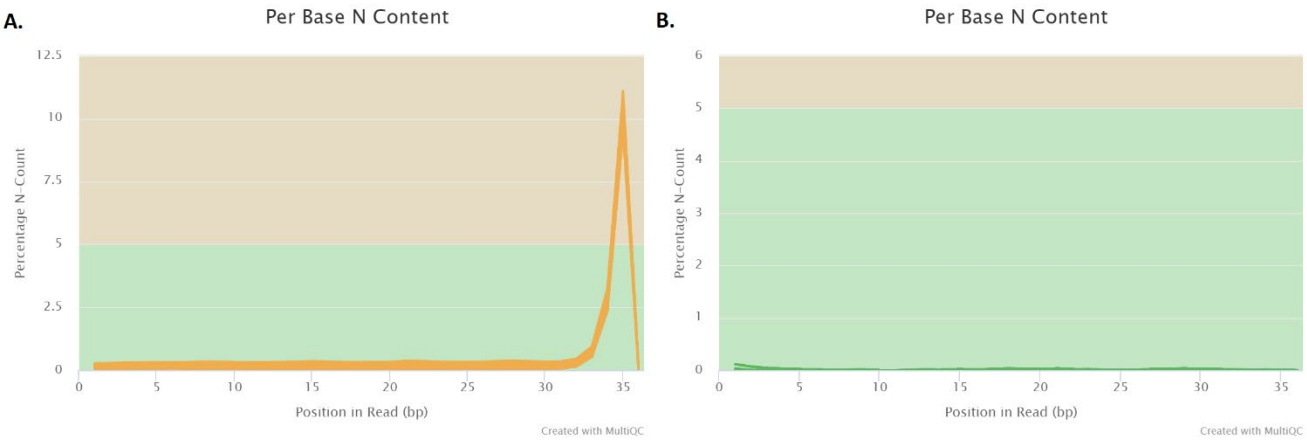


Figure 4. MultiQC Per Base N Content.
(A) Untrimmed Reads (B) Trimmed Reads

## *B. STAR Alignment Metrics*

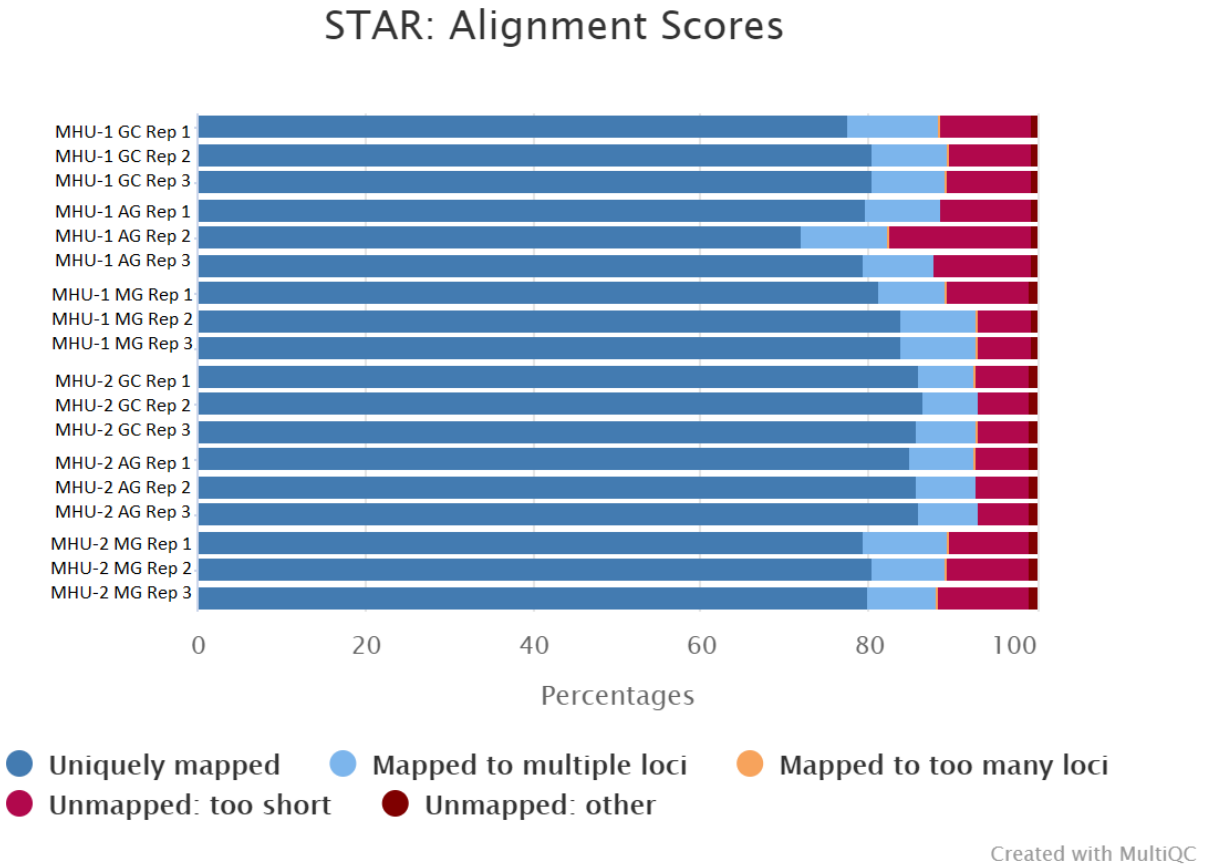


Figure 5. STAR Alignment metrics per sample, compiled with MultiQC

The alignment of the mapped reads for each sample is seen in Figure 5. With the exception of MHU-1 AG Rep 2, at least 77.6% of the reads for each sample were uniquely mapped. Approximately 9% of reads mapped to multiple loci, and approximately 8.3% were unmapped due to short length. A very small amount (1.1% or less) either mapped to too many loci or remained unmapped due to other reasons. The one outlier was MHU-1 AG Rep 2, which had 72.1% of reads that were uniquely mapped, 10.2% of reads that mapped to multiple loci, and 16.8% that were unmapped because the reads were too short.

*C. Preliminary Findings*

Differential expression was divided into a total of 6 groups: GC, MG and AG from MHU-1 and GC, MG, and AG from MHU-2. Each of these groups had 3 mice each, resulting in 18 mice in total. From this experimental design there were 6 comparative groups in total, 3 from each space mission: GC vs MG, GC vs AG and AG vs MG for MHU-1 and the same groups for MHU-2. The same group from both space flights were then compared to find the differentially expressed genes in common. A gene was classified as differentially expressed for a group if the FDR < 0.05 and the Log 2-fold change was less than -1 or greater than 1. The number of differentially expressed genes is seen in Table II. Principal Component Analysis (PCA) of the gene counts after normalization was also performed to reduce dimensionality and visualize the difference in gene expression profiles between the 6 groups. The PCA plot, seen in Figure 6, suggests that variation within each of the groups was minimal. The biological triplicates for each group had consistent gene expression, with the exception of one AG sample from MHU-1. For the MG group from MHU-1, there appears to be only 2 samples instead of 3, but this is due to a close similarity between 2 of the samples, which are overlapping. Additionally, the PCA plot

shows that the GC groups were relatively similar, but the MG and AG groups were notably different. This may be due to the differences in flight schedules, which are detailed in Table I.

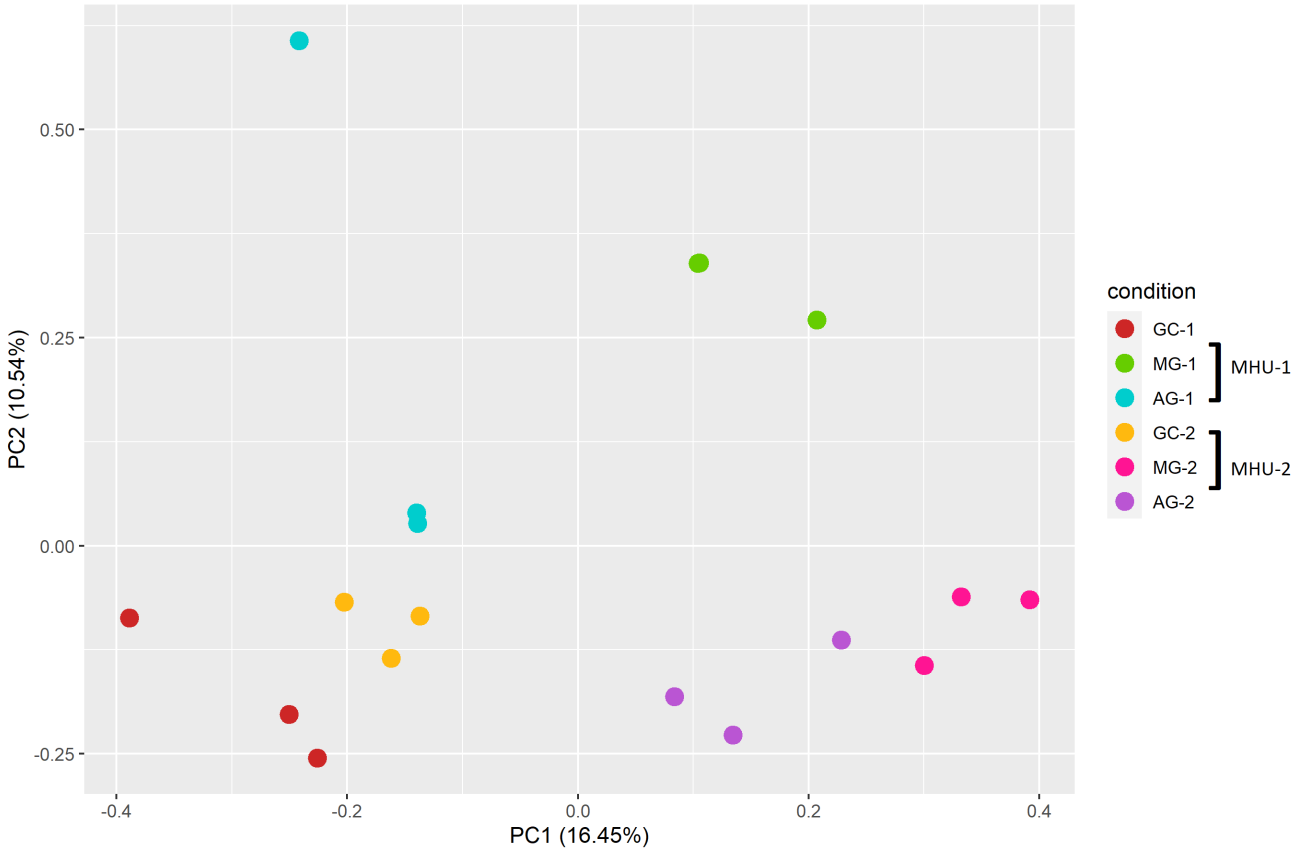


Figure 6. Principal Component Analysis (PCA) plots for normalized gene counts

**TABLE II. TOTAL NUMBER OF DIFFERENTIALLY EXPRESSED GENES PER GROUP**

| | GC vs MG | GC vs AG | AG vs MG |
|---|---|---|---|
| **MHU-1** | 665 | 11 | 84 |
| **MHU-2** | 1039 | 153 | 313 |


Figure 7. Venn diagram depicting the number of shared differentially expressed genes between the spaceflights MHU-1 and MHU-2

### *D. Comparison to GeneLab Results*

Figure 8 depicts the shared and unique genes identified in the Nextflow implementation of the RCP compared to GeneLab. This comparison was accomplished by obtaining the differential expression analysis data from the GLDS-289 repository. The GeneLab data was then processed in the same way as the Nextflow implementation. Differentially expressed genes were first split according to the 6 comparative groups, GC vs MG, GC vs AG and AG vs MG for flights MHU-1 and MHU-2, then groups between spaceflights were compared to identify the DEGs in common between MHU-1 and MHU-2, resulting in 3 groups. The GeneLab results were then compared to the Nextflow implementation to identify shared and unique DEGs per group. The Nextflow implementation and the GeneLab results had significant overlap for GC vs MG, but there was still a number of unique genes found from both processes. The results for GC vs AG and AG vs MG were less promising, with little overlap between the Nextflow implementation and the GeneLab results. Gene set analysis was then performed on the DEGs identified by GeneLab but not detected by the Nextflow implementation to determine if the differences between the Nextflow implementation and GeneLab lead to different conclusions.

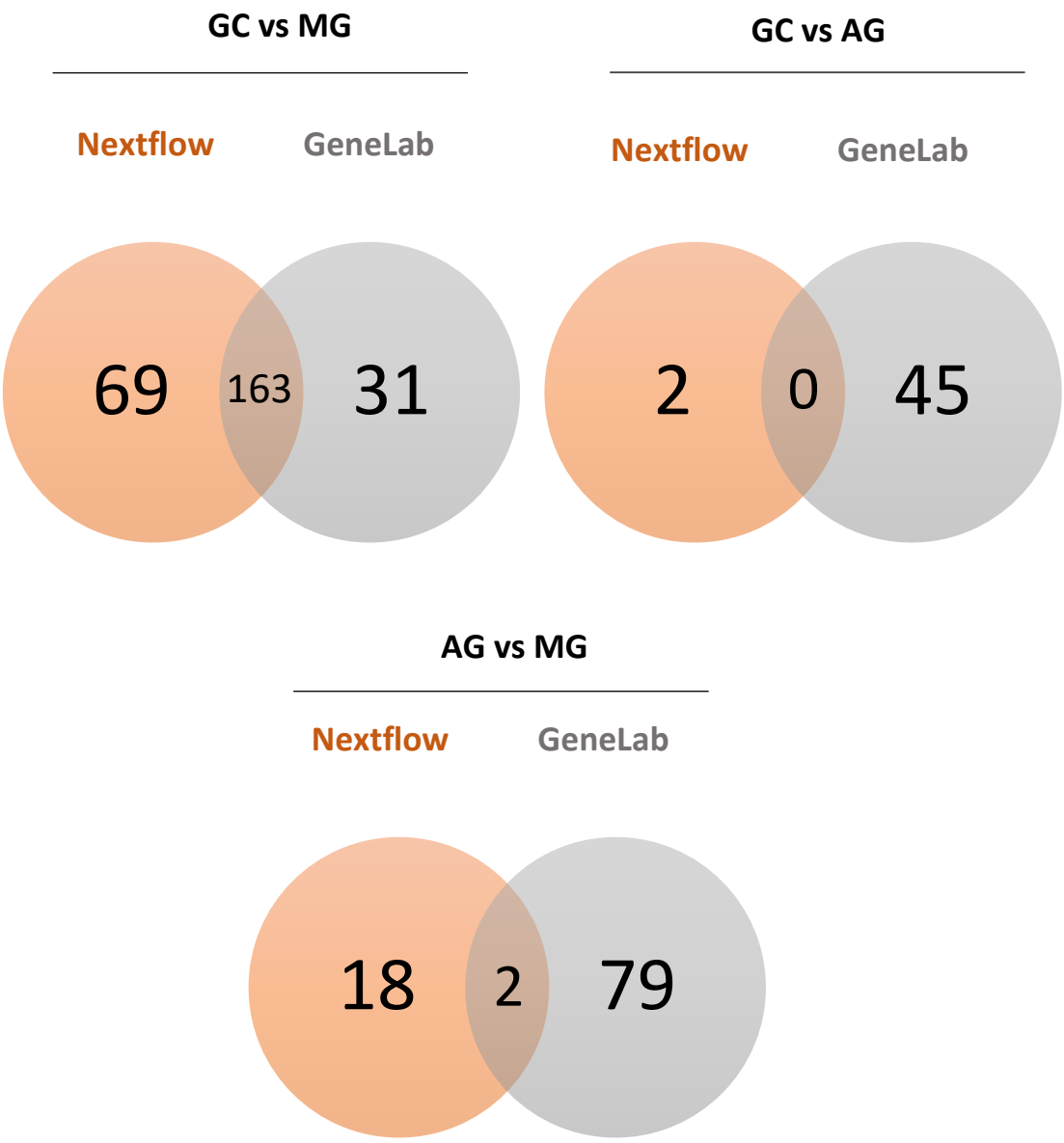


Figure 8. Venn Diagram depicting the shared and unique DEGs between the Nextflow implementation versus the results posted on GeneLab.

## *E. Gene Set Analysis*

### *1. DAVID*

Table III summarizes the top 3 DAVID clusters found from the DEGs in the GC vs MG group from MHU-1. All three clusters suggest disruptions in the nucleosome and histone function, which are essential for DNA organization. Histones are small, positively charged proteins that negatively charged DNA tightly coils around to form complexes known as nucleosomes. This complex allows from the compaction of large volumes of DNA to be stored in the nucleus [36]. This disruption in histone and nucleosome function is also seen in third annotation cluster from Table IV, which depicts the top 3 DAVID clusters found from the DEGs in the GC vs MG group from MHU-2. The first cluster suggests disruptions in the cell cycle, which is presented in Figure 9. The cell cycle is a series of stages occurring within a cell that results in cell division. It is comprised of 2 main parts: interphase and mitosis. Interphase is divided into 3 stages: G1, S, and G2. During G1, the cells grow in preparation of cell division. In the S phase, all the DNA in the cells are copied. Finally, in G2, genetic material is organized and begins to condense in preparation of cell division, which occurs in mitosis [37]. The disruption of the cell cycle at any of these stages would lead to a reduction in cell division and a decrease in cell proliferation, suppressing the immune response due to the decrease in newly generated immune cells.

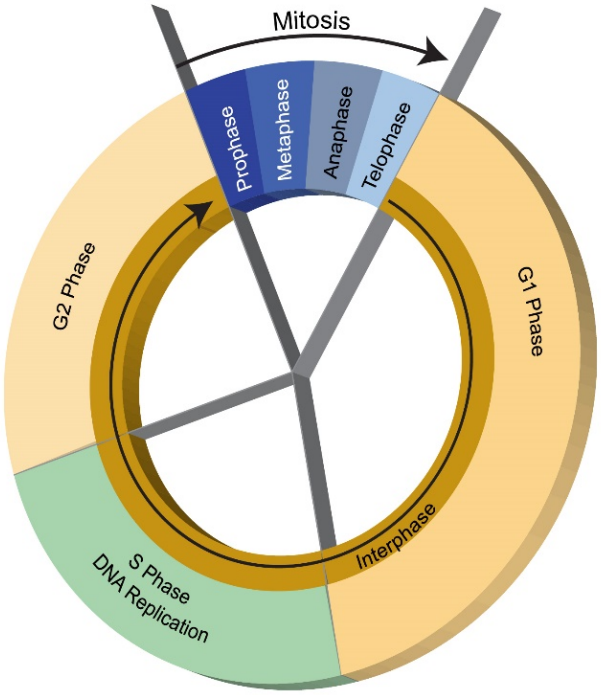


Figure 9. Depiction of the Cell Cycle through Interphase and Mitosis. Adapted from [37].

**TABLE III. TOP 3 DAVID CLUSTERS FOR GC VS MG MHU-1**

| Category | Term | FDR |
|---|---|---|
| **Cluster 1** | | |
| UP_KEYWORDS | Nucleosome core | 6.08E-35 |
| INTERPRO | IPR007125:Histone core | 4.52E-32 |
| INTERPRO | IPR009072:Histone-fold | 4.11E-31 |
| **Cluster 2** | | |
| INTERPRO | IPR007125:Histone core | 4.52E-32 |
| UP_KEYWORDS | Citrullination | 1.61E-26 |
| INTERPRO | IPR002119:Histone H2A | 2.18E-07 |
| UP_SEQ_FEATURE | chain:Histone H2A type 1 | 1.81E-05 |
| UP_SEQ_FEATURE | chain:Histone H2A type 1-H | 1.81E-05 |
| **Cluster 3** | | |
| GOTERM_BP_DIRECT | GO:0006334~nucleosome assembly | 2.29E-21 |
| INTERPRO | IPR000558:Histone H2B | 1.35E-10 |
| SMART | SM00427:H2B | 1.65E-09 |

**TABLE IV. TOP 3 DAVID CLUSTERS FOR GC VS MG MHU-2**

| Category | Term | FDR |
|---|---|---|
| **Cluster 1** | | |
| UP_KEYWORDS | Cell cycle | 3.05E-61 |
| GOTERM_BP_DIRECT | GO:0007049~cell cycle | 2.14E-54 |
| UP_KEYWORDS | Cell division | 2.00E-53 |
| **Cluster 2** | | |
| GOTERM_CC_DIRECT | GO:0005694~chromosome | 6.03E-48 |
| GOTERM_CC_DIRECT | GO:0000775~chromosome, centromeric region | 1.81E-34 |
| UP_KEYWORDS | Centromere | 6.48E-32 |
| **Cluster 3** | | |
| UP_KEYWORDS | Chromosome | 4.62E-94 |
| UP_KEYWORDS | Nucleosome core | 1.59E-56 |
| INTERPRO | IPR009072:Histone-fold | 1.12E-50 |

*2. GOrilla*

Table V presents the top 20 enriched GO Biological Process (BP) terms for GC vs MG DEGs found in both flights. Two lists were uploaded to GOrilla: the list of GC vs MG DEGs that were common in MHU-1 and MHU-2 and the background list of genes, which were all the annotated genes that were found in sequencing. The list returned from GOrilla is organized in ascending order based on the P-value and FDR. The top 3 enriched GO BP terms are related to the cell cycle, which supports the results seen in the first DAVID cluster in Table IV. There are also multiple enriched GO terms regarding chromosomes, nucleosomes, and chromatin, all which are important for the packaging and organization of DNA. Another interesting set of GO terms to note are the ones regarding the regulation of gene expression. The packaging of DNA through chromatin and chromosome organization has a large impact on the regulation of gene expression [38]. Therefore, it makes sense that the impact of microgravity on DNA organization would also affect gene expression regulation.

**TABLE V. GORILLA TOP 20 ENRICHED GO BP TERMS FOR COMBINED GC VS MG**

| GO term | Description | P-value | FDR | Enrichment |
|---|---|---|---|---|
| **GO:1903047** | mitotic cell cycle process | 4.37E-11 | 6.62E-7 | 5.13 |
| **GO:0007049** | cell cycle | 3.05E-10 | 2.31E-6 | 3.87 |
| **GO:0022402** | cell cycle process | 7.3E-10 | 3.69E-6 | 3.73 |
| **GO:0006334** | nucleosome assembly | 1.26E-9 | 4.76E-6 | 14.74 |
| **GO:0051301** | cell division | 1.53E-9 | 4.63E-6 | 4.95 |
| **GO:0034728** | nucleosome organization | 6.13E-9 | 1.55E-5 | 10.70 |
| **GO:0065004** | protein-DNA complex assembly | 6.85E-9 | 1.48E-5 | 10.59 |
| **GO:0035458** | cellular response to interferon-beta | 7.74E-9 | 1.47E-5 | 18.86 |
| **GO:0071824** | protein-DNA complex subunit organization | 1.85E-8 | 3.12E-5 | 8.45 |
| **GO:0051276** | chromosome organization | 3.93E-8 | 5.95E-5 | 5.47 |
| **GO:0035456** | response to interferon-beta | 4.19E-8 | 5.77E-5 | 15.40 |
| **GO:0030261** | chromosome condensation | 5.59E-8 | 7.06E-5 | 19.42 |
| **GO:0006323** | DNA packaging | 1.53E-7 | 1.78E-4 | 16.93 |
| **GO:0040029** | regulation of gene expression, epigenetic | 2.63E-7 | 2.85E-4 | 7.46 |
| **GO:0006325** | chromatin organization | 5.53E-7 | 5.59E-4 | 3.51 |
| **GO:0006342** | chromatin silencing | 1.03E-6 | 9.8E-4 | 12.95 |
| **GO:0071103** | DNA conformation change | 2.04E-6 | 1.82E-3 | 9.43 |
| **GO:0045814** | negative regulation of gene expression, epigenetic | 2.84E-6 | 2.39E-3 | 11.19 |
| **GO:0016584** | nucleosome positioning | 1.12E-5 | 8.95E-3 | 26.95 |
| **GO:0000226** | microtubule cytoskeleton organization | 1.41E-5 | 1.07E-2 | 3.69 |

*3. GSEA*

The gene sets enriched in downregulated gene sets with respect to MG are presented in Table VI. For MG vs GC from MHU-1, there are 12 significantly enriched gene sets with a nominal p-value less than 0.05 and an FDR less than 0.25 and 16 significantly enriched gene sets for MG vs GC from MHU-2 under the same conditions. From both these spaceflights, all 12 gene sets that were significant for MHU-1 were significant for MHU-2 as well. Both the

downregulated and upregulated gene set tables are organized in descending order based on the combined NES. Hallmark gene sets were utilized for both the downregulated and upregulated gene sets since they effectively summarize well defined biological states and processes [39]. The top 2 enriched downregulated gene sets, HALLMARK_E2F_TARGETS and HALLMARK_G2M_CHECKPOINT, are related to the regulation of the cell cycle, which supports the results found by multiple other studies [7], [13], [40].

Table VII presents the upregulated hallmark gene sets with respect to MG. From MHU-1, there are 20 significantly enriched gene sets based on the previously mentioned conditions and for MHU-2 there are 19. The overlap of significant gene sets for MG vs GC from both MHU-1 and MHU-2 is 18. Two of the top three upregulated gene sets relate to interferon response, which are key in activating the immune response to infection through cell signaling [41]. Two other upregulated gene sets, HALLMARK_IL2_STAT5_SIGNALING and HALLMARK_IL6_JAK_STAT3_SIGNALING, are related to interleukins, which are another type of cell signaling protein that facilitates communication between immune and inflammation cells [42]. The presence of both interferons and interleukins in the upregulated hallmark gene sets indicates that microgravity has a profound effect on communication between cells.

## TABLE VI. GSEA DOWNREGULATED MG HALLMARK GENE SETS

| | MHU-1 | | | MHU-2 | | |
|---|---|---|---|---|---|---|
| **Name** | **NES** | **NOM p-value** | **FDR** | **NES** | **NOM p-value** | **FDR** |
| **HALLMARK_E2F_TARGETS** | -3.76 | 0.000 | 0.000 | -3.71 | 0.000 | 0.000 |
| **HALLMARK_G2M_CHECKPOINT** | -3.34 | 0.000 | 0.000 | -3.53 | 0.000 | 0.000 |
| **HALLMARK_MYC_TARGETS_V1** | -2.73 | 0.000 | 0.000 | -3.18 | 0.000 | 0.000 |
| **HALLMARK_DNA_REPAIR** | -2.15 | 0.000 | 0.000 | -2.60 | 0.000 | 0.000 |
| **HALLMARK_OXIDATIVE_PHOSPHORYLATION** | -2.08 | 0.000 | 0.000 | -2.45 | 0.000 | 0.000 |
| **HALLMARK_MITOTIC_SPINDLE** | -1.71 | 0.000 | 0.002 | -2.46 | 0.000 | 0.000 |
| **HALLMARK_MTORC1_SIGNALING** | -1.95 | 0.000 | 0.000 | -2.06 | 0.000 | 0.000 |
| **HALLMARK_SPERMATOGENESIS** | -1.80 | 0.000 | 0.000 | -1.96 | 0.000 | 0.000 |
| **HALLMARK_PEROXISOME** | -1.82 | 0.000 | 0.000 | -1.36 | 0.039 | 0.059 |
| **HALLMARK_CHOLESTEROL_HOMEOSTASIS** | -1.41 | 0.009 | 0.032 | -1.67 | 0.002 | 0.004 |
| **HALLMARK_UNFOLDED_PROTEIN_RESPONSE** | -1.24 | 0.077 | 0.113 | -1.61 | 0.003 | 0.006 |
| **HALLMARK_PI3K_AKT_MTOR_SIGNALING** | -1.46 | 0.000 | 0.028 | -1.19 | 0.159 | 0.173 |

## TABLE VII. GSEA UPREGULATED MG HALLMARK GENE SETS

| | MHU-1 | | | MHU-2 | | |
|---|---|---|---|---|---|---|
| **Name** | **NES** | **NOM p-value** | **FDR** | **NES** | **NOM p-value** | **FDR** |
| **HALLMARK_INTERFERON_GAMMA_RESPONSE** | 2.16 | 0.000 | 0.000 | 2.61 | 0.000 | 0.000 |
| **HALLMARK_INFLAMMATORY_RESPONSE** | 1.95 | 0.000 | 0.000 | 2.67 | 0.000 | 0.000 |
| **HALLMARK_INTERFERON_ALPHA_RESPONSE** | 2.08 | 0.000 | 0.000 | 2.51 | 0.000 | 0.000 |
| **HALLMARK_KRAS_SIGNALING_UP** | 1.77 | 0.000 | 0.001 | 2.36 | 0.000 | 0.000 |
| **HALLMARK_EPITHELIAL_MESENCHYMAL_TRANSITION** | 1.98 | 0.000 | 0.000 | 1.98 | 0.000 | 0.000 |
| **HALLMARK_TNFA_SIGNALING_VIA_NFKB** | 1.80 | 0.000 | 0.001 | 2.11 | 0.000 | 0.000 |
| **HALLMARK_COMPLEMENT** | 1.64 | 0.000 | 0.002 | 2.18 | 0.000 | 0.000 |
| **HALLMARK_APICAL_SURFACE** | 2.19 | 0.000 | 0.000 | 1.63 | 0.008 | 0.005 |
| **HALLMARK_ALLOGRAFT_REJECTION** | 1.66 | 0.000 | 0.002 | 2.12 | 0.000 | 0.000 |
| **HALLMARK_COAGULATION** | 1.57 | 0.003 | 0.005 | 2.20 | 0.000 | 0.000 |
| **HALLMARK_IL2_STAT5_SIGNALING** | 1.84 | 0.000 | 0.000 | 1.92 | 0.000 | 0.000 |
| **HALLMARK_IL6_JAK_STAT3_SIGNALING** | 1.60 | 0.000 | 0.004 | 2.05 | 0.000 | 0.000 |
| **HALLMARK_ANGIOGENESIS** | 1.75 | 0.001 | 0.001 | 1.64 | 0.004 | 0.004 |
| **HALLMARK_UV_RESPONSE_DN** | 1.58 | 0.001 | 0.005 | 1.72 | 0.000 | 0.001 |
| **HALLMARK_APICAL_JUNCTION** | 1.70 | 0.000 | 0.001 | 1.59 | 0.000 | 0.007 |
| **HALLMARK_ESTROGEN_RESPONSE_EARLY** | 1.68 | 0.000 | 0.002 | 1.58 | 0.000 | 0.007 |
| **HALLMARK_KRAS_SIGNALING_DN** | 1.71 | 0.000 | 0.002 | 1.51 | 0.000 | 0.013 |
| **HALLMARK_TGF_BETA_SIGNALING** | 1.72 | 0.001 | 0.001 | 1.43 | 0.032 | 0.026 |

*4. Cytoscape ClueGO Application*

Figure 10 displays the biological network of significantly enriched GO terms from the list of MG vs GC DEGs found in both MHU-1 and MHU-2. The two ontologies selected were biological processes and immune system process. Out of the 232 DEGs, 181 were annotated to at least one of the selected ontologies. From these 181 genes, after filtering based on a general selection criterion, then fusing similar GO terms and filtering out terms with a p-value greater than 0.05, there were 45 GO terms representing the list of DEGs left. The color of a node is based on functional group, and clusters are formed for groups of nodes greater than 3 that belong to exactly 1 functional group. The total number of groups and percentage of GO terms represented per group is presented in Figure 11. Nodes that are multicolored belong to multiple functional groups. Nodes are connected to other nodes within their group and to nodes belonging to other groups by edges, which are calculated based on the number of shared genes between the nodes [33].

The three most significantly enriched groups are chromatin organization, chromosome organization, and cell cycle process. There are multiple nodes connecting chromatin and chromosome organization, indicating that the two groups are highly connected. The presence of a large functional group for the cell cycle supports the DAVID cluster results seen in Table IV and the GOrilla results seen in Table V, further emphasizing the impact of microgravity on cell cycle regulation. Another interesting functional group to note is the response to interferon-beta. The upregulation of gene sets related to interferons was seen in the GSEA results presented in Table VII, but those were gene sets for two different interferons, interferon alpha

and interferon gamma. However, the impact of microgravity on interferon beta gene expression is seen in other studies, although the results were conflicting [43], [44].

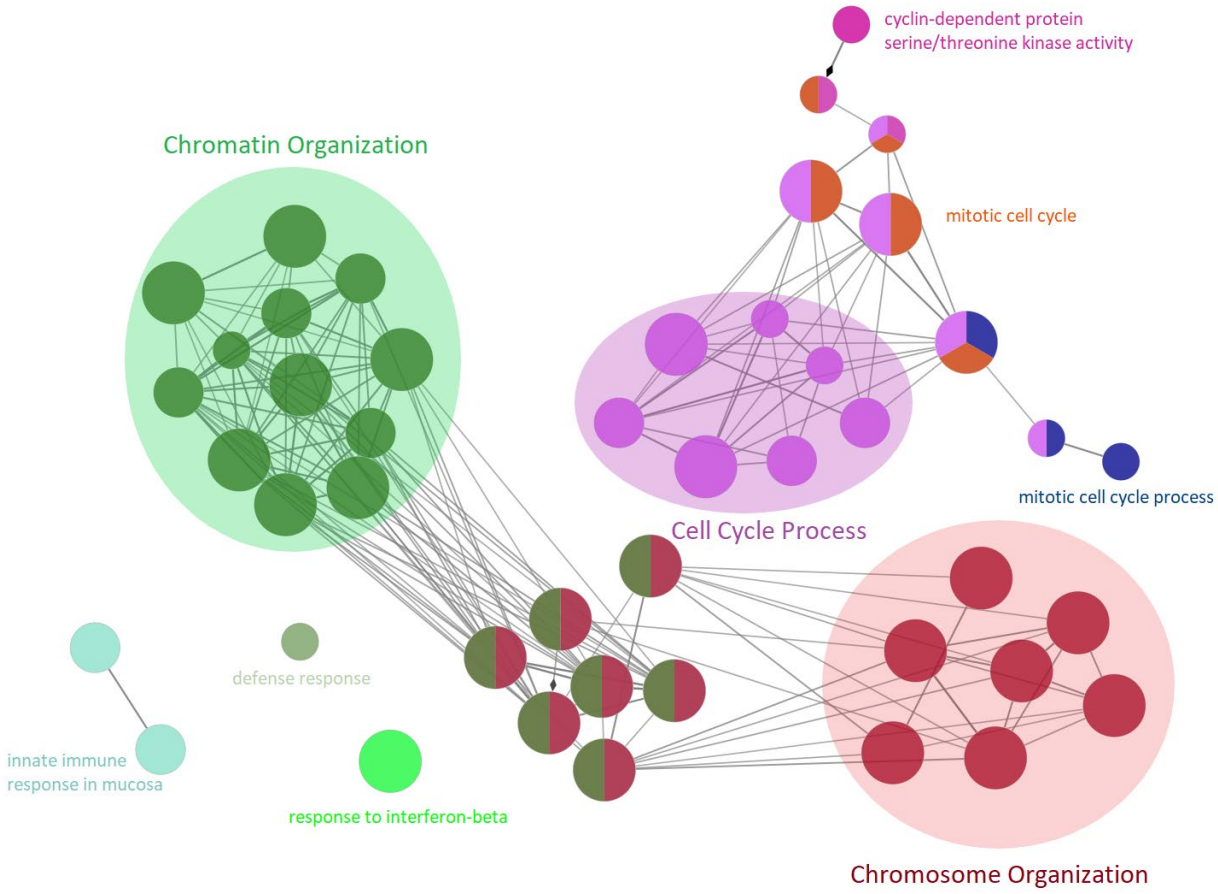


Figure 10. ClueGO generated biological network of enriched GO terms from the biological processes and immune system process ontologies. Clusters contain at least 3 nodes belonging to only one group and were manually annotated. Nodes are connected by edges, which are calculated based on the number of shared genes between the nodes.

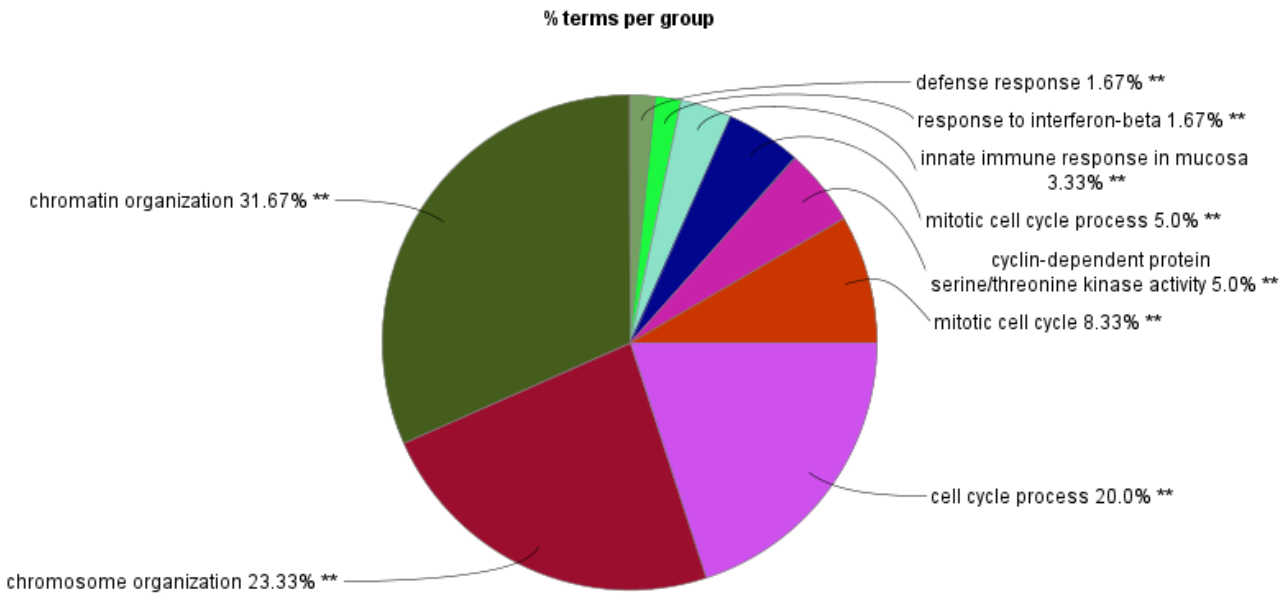


Figure 11. Pie chart summarizing the number of functional groups represented in the network generated by ClueGO and the percentage of the total GO terms represented per group.

### *F. Use of Artificial Gravity for Lessening Impact of Microgravity*

Figure 12 displays a clustered heatmap of the scaled gene expression of histone and histone related genes from MHU-1 and MHU-2, organized by sample group. The cladogram was removed to simplify the figure. The MHU-1 heatmap, left, does not show a significant difference in gene expression between MG, AG and GC. However, the MHU-2 heatmap, right, does present a notable difference the gene expression in MG mice compared to the GC and AG mice. Horie et al. also found variations in histone gene expression in MG mice compared to GC and AG mice [7]. Additionally, the number of DEGs, presented in Table II, were notably lower in the GC vs AG groups than the GC vs MG groups, which suggests that exposure to artificial gravity partially alleviates the impact of microgravity.

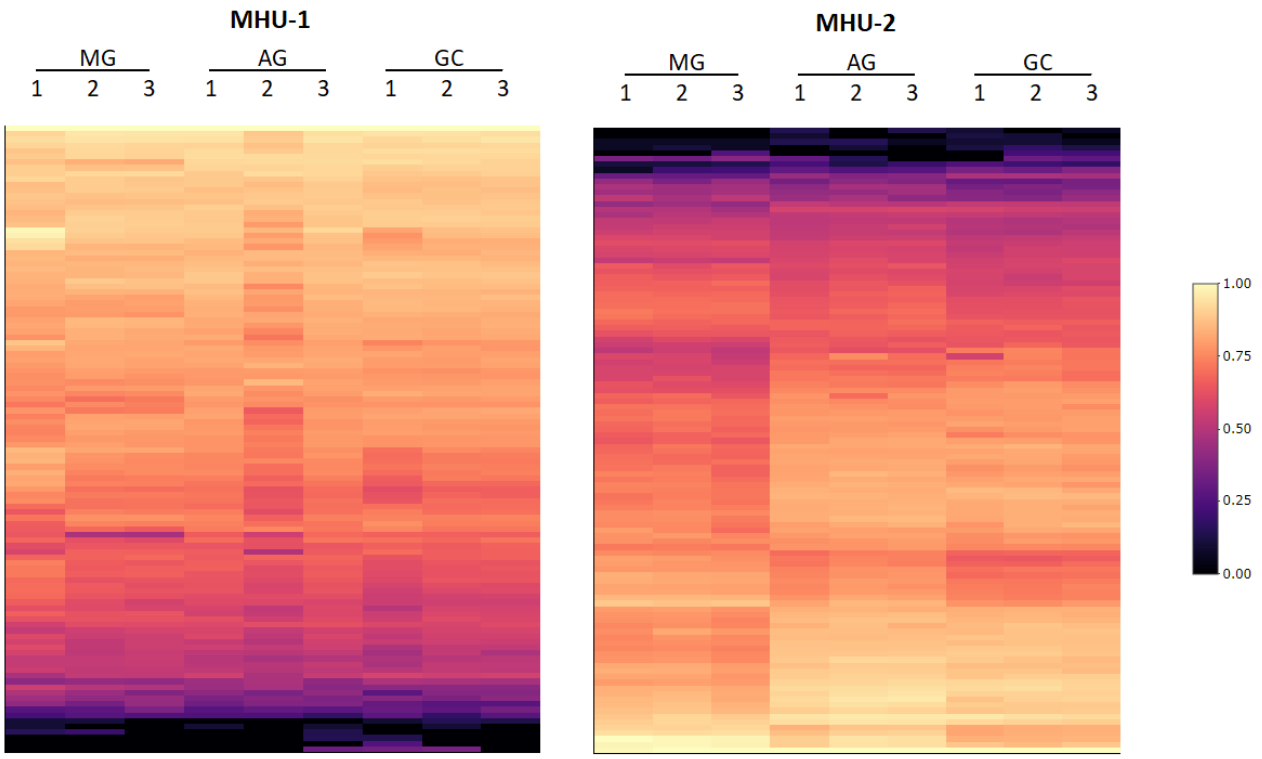


Figure 12. Clustered heatmap of histone and histone related scaled gene expression in the thymus, organized by sample group. The cladogram was removed to simplify the figure. Yellow indicates high gene expression and black indicates low expression.

## IV. Discussion

### *A. Preliminary Findings*

Table II presents the total number of differentially expressed genes (DEGs) per group. The gene expression profiles were markedly different in the GC vs MG groups than GC vs AG and AG vs MG, which had considerably fewer DEGs. The PCA plot presented in Figure 6 demonstrates that the triplicate samples for each group had relatively similar gene expression, with the exception of 1 AG-1 sample. The PCA plot also indicates that gene expression was similar for the GC groups but notably differed for the MG and AG groups. This large difference in gene expression in the spaceflight groups is also seen in the small overlap of DEGs, as seen in

Figure 7. According to Table II, MHU-1 had 665 DEGs for GC vs MG and MHU-2 had 1039 DEGs for the same group, but only 232 DEGs overlapped between the 2 groups in Figure 7. This small overlap of DEGs can possibly be attributed to the difference in spaceflight missions, as explained in Table I. Mice were 8 weeks old at the time of launch for MHU-1 and 9 weeks old at time of launch for MHU-2. This difference in age should not result in differences in development, since female mice reach maturity at 6 weeks old while male mice reach maturity by 8 weeks old [45]. Therefore, both spaceflights should have mice who have already reached full maturity. There were also differences in date of launch, duration of spaceflight and time of sacrifice after returning to Earth. Mice from MHU-1 spent 5 days longer in space than MHU-2, and were sacrificed after 48 hours as opposed to 36.5 hours. Overall, differential gene expression was greater for MHU-2 mice than MHU-1, which may be due to shorter waiting time before sacrifice. A longer duration of time before sacrifice once returning to Earth may allow for the gene expression to adjust a bit more to normal levels.

### *B. Comparison to GeneLab Results*

Figure 8 presents the shared and uniquely identified genes found per group in the Nextflow implementation of RCP compared to the posted GeneLab results. Although a majority of the GC vs MG genes were found in both the Nextflow implementation and the GeneLab results, there were a significant number that were unique to each, which is disappointing. The results for GC vs AG and AG vs MG were worse, with no genes overlapping for GC vs AG and only 2 genes overlapping for AG vs MG. One possible explanation for the discrepancies could be differences in the operating system or variations in the tool versions used within the RCP. It was also noted that a Docker based version of the Nextflow implementation may produce results

that more closely align with the results obtained by GeneLab, so the use the Docker based version in future iterations of the Nextflow implementation may be worth considering [19]. Future work using the pipeline should further explore the differences between the Nextflow implementation and GeneLab in order to improve reproducibility.

Gene set analysis was performed on the GeneLab GC vs MG DEGs identified by GeneLab but not by the Nextflow implementation to see if the results significantly differed. The GeneLab analysis returned similar results found with the Nextflow implementation from DAVID, GOrilla, and GSEA. Although the number of uniquely identified genes between the Nextflow implementation and the GeneLab results are concerning, it is promising to see that the same alterations to biological processes and structures can be seen from both applications.

## *C. Gene Set Analysis*

### *1. Histone Function and DNA Organization*

All three DAVID clusters for GC vs MG from MHU-1 seen in Table III suggest disruptions in histone function due to space flight, which is supported by other studies [7], [46]. Histones are important for the regulation of gene expression due to their role in DNA packaging, which is essential for storing large amounts of DNA into a small cell. The organization of DNA is illustrated in Figure 13. The most basic organizational unit is a nucleosome, which is a DNA strand wrapped around 8 histones. The next level of organization is chromatin, which are repeating units of nucleosome packaged together. Chromatin are then further folded to make up the building blocks of chromosomes, allowing for large amounts of DNA to be compacted into a small area. Chromatin varies in its level of compaction. In its most compact form, DNA is inaccessible for transcription, and when the compaction of chromatin is more relaxed, DNA is more accessible for RNA polymerase to access, making chromatin important in the regulation of gene expression [47]. The effects of microgravity on DNA organization and packaging are also seen in the GOrilla results presented in Table V, where there were numerous enriched GO terms regarding DNA packaging, nucleosomes, and chromatins. [48]

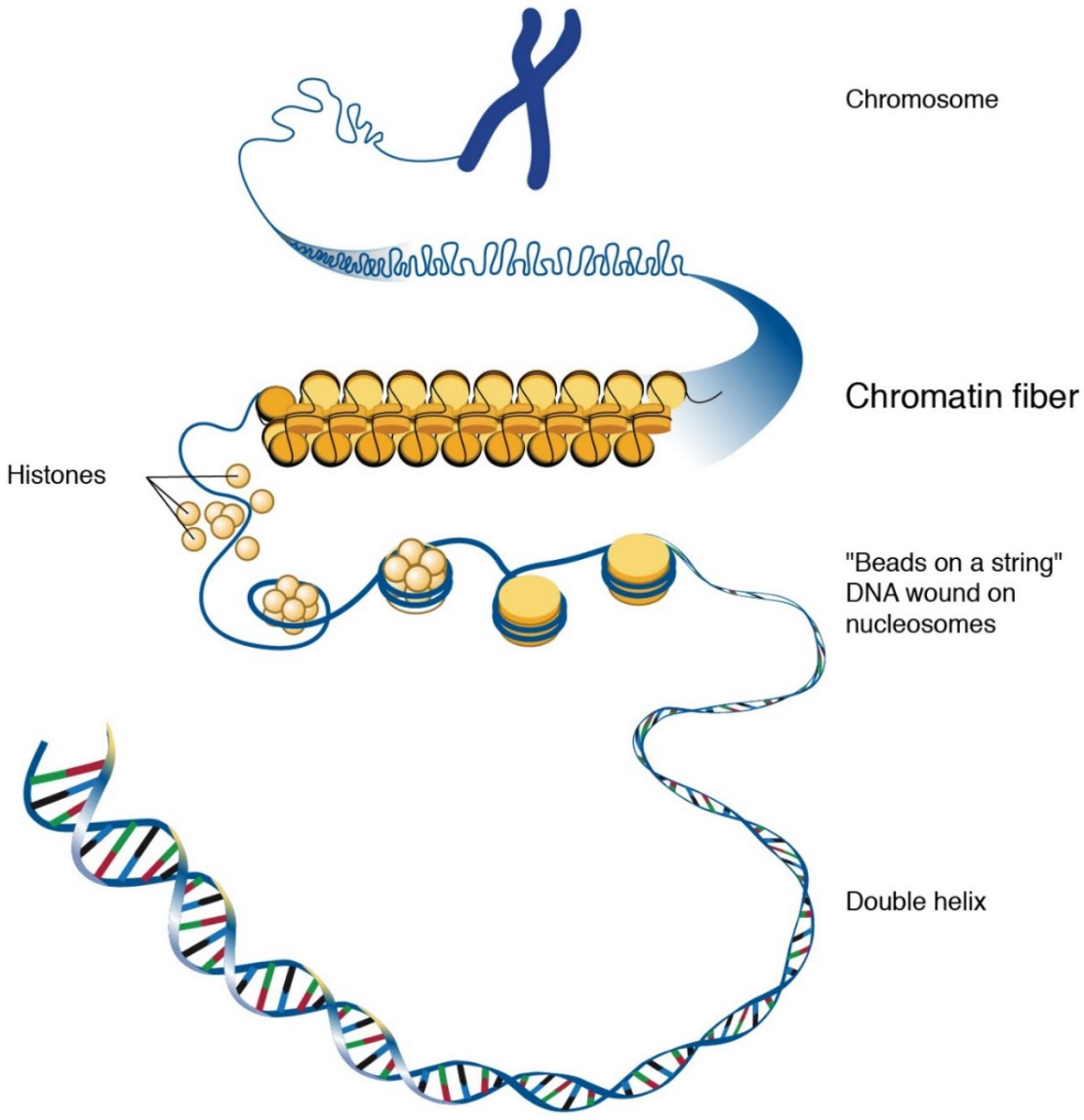


Figure 13. The hierarchy of DNA organization, from DNA to chromosome. Starting from bottom to top, the first level is DNA. The following level is a nucleosome, which is DNA wrapped around 8 histones to form “beads on a string”. The next level is chromatin, which are repeating units of nucleosomes packed together. Chromatin are then folded to tightly pack large amounts of DNA onto a chromosome. Illustration from [48].

The changes in DNA organization due to reduced histone expression also impacts other aspects of the cell, such as the cell cycle. Saunders et al. found that histone depletion led to changes to chromatin structure, which resulted in cell cycle arrest [49]. Chromatins and cell cycle progression are closely linked, since the packaging of DNA affects DNA replication, chromosome segregation and other processes that occur during the early phases of the cell cycle [50]. The cell cycle controls cell division, so cell cycle arrest results in a depression of cell proliferation. Horie et al. found a downregulation of many histone genes in their MG mice, noting that this resulted in a decrease in cell proliferation [7]. Decreased cell proliferation was also found in other studies, and is important for the generation of new cells, so a decrease in cell proliferation depresses the immune response through the decrease in T cell generation [10], [11]. Microgravity significantly impacts histones through changes in DNA packaging and the cell cycle, which severely impacts the regulation of gene expression, cell proliferation, and the immune system.

2. *Disruptions in the Cell Cycle*

The first DAVID cluster for MHU-2 in Table IV and multiple enriched GO terms found by GOrilla in Table V suggest disruptions to the cell cycle due to microgravity, which was also found in other studies [7], [13], [40]. Gridley et al. performed pathway analysis on gene expression data from the thymus and spleen and found that cell progression was halted in both the spleen and thymus. This cell cycle arrest was due to the downregulation of early interphase checkpoint genes in addition to the upregulation of cell cycle progression inhibitors [13]. Dysregulation of cell cycle was also found due to the activation of regulatory transcription factors. Novoselova et al. found an increase in activated p53 in thymic lymphocytes of mice

who had recently returned from space [9]. P53 is a transcription factor that is induced by stress signals such as nutrient deprivation and DNA damage. Once activated, it results in cell cycle arrest and apoptosis [51].

The effects of microgravity on the cell cycle are also seen in Table VI, which contains the GSEA enriched downregulated MG hallmark gene sets. The most enriched hallmark gene set from this table is HALLMARK_E2F_TARGETS, which contains genes that encode for cell cycle related targets of E2F transcription factors [52]. E2F is a group of genes that regulates cell proliferation and is particularly involved in the progression from the G1 stage into the S stage during interphase [53]. Another enriched downregulated gene set is HALLMARK_G2M_CHECKPOINT, which includes the genes involved in the G2/M checkpoint [54]. Checkpoints are important regulatory pit stops for the cell to be examined before proceeding to the next step of the cell cycle. The G2/M checkpoint is the last step before entering mitosis, ensuring that DNA is not damaged and chromosomes have been properly replicated before beginning cell division [55]. The downregulation of both these gene sets results in the halting of cell cycle progression, and the appearance of cell cycle related gene sets in three different gene set analysis tools indicates the monumental impact that microgravity has on the regulation of the cell cycle.

*3. Chromosomal Abnormalities*

Two DAVID clusters for GC vs MG in MHU-2 seen in Table IV and a number of GOrilla enriched GO BP terms in Table V indicate that microgravity impacts the chromosome. This result is supported by other studies, who have found that exposure to microgravity and radiation led to an increase in chromosome aberrations [56]–[58]. Disruptions to the

chromosome can affect the regulation of gene expression, disrupt exons, and create fusion genes [59]. Furthermore, chromosomal abnormalities are inheritable, and can result in diseases such as Turner Syndrome or Down Syndrome [60]. One major step in the future of space travel is inhabiting space, and the cumulation of chromosomal aberrations being passed down from generation to generation could result in an increased number of developmental disabilities.

4. *DNA Repair*

Another interesting GSEA downregulated hallmark gene set seen in Table VI is HALLMARK_DNA_REPAIR. The reduction in DNA repair genes is supported by multiple other studies [40], [61], [62]. Kumari et al. tested whether simulated microgravity conditions affected the expression of DNA repair and apoptosis genes and found that the expression of DNA repair genes decreased [40]. DNA repair is crucial for the correction of DNA damage that could result from external sources such as the environment or errors in replication. Without DNA repair mechanisms, unchecked modifications to the DNA may result in cell cycle arrest, apoptosis, or cancer [63].

5. *GSEA Upregulated Hallmark Gene Sets*

The top 3 enriched upregulated gene sets presented in Table VII, HALLMARK_INTERFERON_GAMMA_RESPONSE, HALLMARK_INTERFERON_ALPHA_RESPONSE and HALLMARK_INFLAMMATORY_RESPONSE, are related to interferon and inflammatory response. Interferons are a type of regulatory signaling protein known as a cytokine. They are secreted by cells in response to virus or foreign invader, and they stimulate the surrounding cells to secrete proteins that will prevent the virus from replicating. Additionally, interferons are involved in other cellular processes such as cell growth, cell division, and cell proliferation.

There are two types of interferons: Type I, which includes multiple subtypes such as interferon alpha (IFN-α) and interferon beta (IFN-β), and Type II, which just has one subtype, interferon gamma (IFN-γ). The two types of interferons differ in their structure and respond to different cell surface receptors [41]. Members from both types of interferons were upregulated in the gene sets, IFN-α and IFN-γ. Another type of cytokine, interleukins, were also enriched in the upregulated hallmark gene set. Interleukins are responsible for mediating communication between immune and inflammatory cells. The two upregulated interleukins seen in the GSEA gene sets, IL-2 and IL-6, are produced by T cells [42].

In addition to cytokines, inflammation is important for the immune response to a foreign invader. Inflammation occurs in response to infection, resulting in the release of signals that produce cytokines and chemokines, which are small cytokines. Inflammation is important for protecting tissues and defense against infection but can also result in decline in tissue function [64]. The upregulation of interferons and inflammation is consistent with the results found by other studies [8], [65]. Chang et al. found an increase in the production of proinflammatory cytokines in flight mice when compared to ground mice, observing a 5-fold increase in IFN-γ [65]. Gridley et al. reported an increase in IFN-γ and another interleukin IL-10 [8]. However, these results conflict with numerous other studies who found that interferon response was inhibited by microgravity [9], [66]. Gridley et al. also noted these conflicting results, suggesting that the differences in interferon response may be due to the cell phenotype, time of assessment after landing, or differences in the mice, such as age, gender, or genetic background [8]. However, even after considering all these possible variables in experimental design, the wide variations in interferon response are perplexing. Further

research into cytokine and interferon response to space flight needs to performed in order to elucidate the impact of microgravity on cell communication.

One other interesting upregulated gene set seen in Table VII is HALLMARK_COAGULATION. Coagulation is the transformation of blood from liquid to solid, and is important for clot formation. Kim et al. also found data supporting the upregulation of coagulation, suggesting that astronauts may be exposed to an enhanced coagulation state due to alterations in venous flow, pressure, distension, and damage to the inner linings of the small intestine [67]. Increased coagulation as a result of spaceflight can be fatal since it may result in blood clots that can become life threatening. Marshall-Goebel et al. assessed the internal jugular vein (IJV) blood flow and morphology during spaceflight in 11 astronauts and found stagnant blood flow in 6 astronauts and a concerning blood clot in 1 astronaut [68]. Coagulation also plays a role in the innate immune response, with infection eliciting a coagulation response that can limit the invasiveness of a foreign invader. However, excessive coagulation has to be regulated in order to prevent the damage that could be done on the host immune system [69]. Therefore, an increase in coagulation can cause damage not just to the vascular system but to innate immunity as well. Nevertheless, coagulation is a poorly understood area in space biology and must be further researched to understand the ramifications of space flight on blood clotting.

*6. ClueGO*

Figure 10 presents the network of functional groups differentially expressed in MG vs GC that supports the results seen with the other gene set analysis tools. One benefit of the network generated by ClueGO is that it displays the connections between multiple groups, demonstrating how interconnected the expression of genes affecting biological pathways and structures are. The large number of nodes connecting the chromatin organization and chromosome organization functional groups is unsurprising considering that both chromatins and chromosome packaging are a part of the same hierarchy of DNA organization, so changes in gene expression for histones and nucleosomes would affect all the structural elements of DNA packaging.

Another notable functional group seen in Figure 10 is the cell cycle process group. The impact of microgravity on the cell cycle is already noted, but one interesting functional group connected to the cell cycle is the cyclin-dependent protein serine/threonine kinase activity group seen in dark purple. Cyclins are important proteins in the regulation of cell cycle progression, and increased cyclin gene expression leads to disruptions in the cell cycle and can even result in cancers such as sarcoma or carcinoma [70]. Microgravity affects the cell cycle in numerous ways through changes in the expression of regulatory proteins or the progression through checkpoints, and this disruption to normal cell cycle progression can have grave consequences.

### *D. Artificial Gravity Partially Alleviates Effects of Microgravity*

The heatmap for histone and histone related gene expression from MHU-2 in Figure 12 suggests that artificial gravity partially alleviates the negative impact of microgravity. The positive impact of artificial gravity is also seen in Table II, since the number of DEGs for the GC vs AG groups were much lower than the number of DEGs in the GC vs MG groups. Horie et al. found that artificial gravity reduced the amount of thymic atrophy that occurred during space flight. Other cellular components were partially rescued by artificial gravity, such as cyclins, which are important for the regulation of the cell cycle. Although artificial gravity lessened the damage caused by microgravity, it did not completely prevent it. Thymic size was significantly smaller in AG mice when compared to GC mice, which suggests that factors other than microgravity result in changes to the body, such as $CO_2$ levels or stress [7]. Therefore, the use of artificial gravity is not a complete solution, but is still a worthwhile inclusion to a space flight in order to partially alleviate the effect of microgravity.

## V. Conclusion

The usage of Jonathan Oribello's Nextflow implementation of the GeneLab RCP for my dataset was a success. I was able to easily adapt his implementation to analyze my dataset of interest by providing a configuration file that navigated to GLDS-289, downloaded the raw paired end files for each sample, and proceeded through the pipeline. There were some discrepancies when comparing the results obtained from the Nextflow implementation to the GeneLab results, but the overall impact of microgravity on multiple functional groups and biological processes still aligned. Hopefully the results from this RNA analysis demonstrates that

Jonathan's Nextflow implementation of the RCP is easily adaptable to multiple GeneLab datasets and other SJSU students will utilize it to perform their own RNA-Seq analysis.

Microgravity has a profound impact on the immune system, particularly on the regulation of the cell cycle and histone gene expression. The cell cycle is essential for the production of new cells, and the progression through the cell cycle is highly regulated. Microgravity impacts this regulation in numerous ways, from cyclins to checkpoints, and these disruptions put the cell cycle in disarray. Changes in the cell cycle halts production of new immune cells, leaving the immune system highly compromised. Another significant impact of microgravity on the immune system is the change is histone gene expression. Histones are essential in DNA organization and packaging, which in turn affects the regulation of gene expression. Disruptions to histone function has a widespread impact, affecting biological processes and structures such as cell proliferation and organization, which negatively impacts the immune system.

The effects of microgravity are partially mitigated by artificial gravity. Histone gene expression for AG mice in MHU-2 was relatively similar to GC mice than compared to MG mice, suggesting that artificial gravity partially rescues histones from the impact of microgravity. Additionally, the number of DEGs in the GC vs AG groups compared to the GC vs MG groups were markedly less, indicating that gene expression profiles are less altered by artificial gravity compared to microgravity. The use of technology that simulates artificial gravity on space crafts could protect astronauts from some of the dangerous effects space has on the human body. However, the commitment to use artificial gravity technology to combat the deleterious effects of microgravity is a very expensive undertaking. Further work into understanding the use of

artificial gravity for alleviating the impact of microgravity must be done before any further actions can be taken.

## References

[1] M. Heppener, "Spaceward ho!," *EMBO Rep.*, vol. 9, no. S1, pp. S4–S12, Jul. 2008, doi: 10.1038/embor.2008.98.

[2] C. Warner, "NASA Perseveres Through Pandemic, Looks Ahead in 2021." *NASA*, Jan. 05, 2021. https://www.nasa.gov/feature/nasa-perseveres-through-pandemic-looks-ahead-in-2021 (accessed Jan. 14, 2021).

[3] S. O'Kane, "SpaceX will launch private citizens into orbit," *The Verge*, Feb. 18, 2020. https://www.theverge.com/2020/2/18/21142137/spacex-tourism-orbit-earth-private-citizens-dragon-space-flight (accessed Mar. 25, 2021).

[4] R. J. White and M. Averner, "Humans in space," *Nature*, vol. 409, no. 6823, pp. 1115–1118, Feb. 2001, doi: 10.1038/35059243.

[5] P. J. Delves and I. M. Roitt, "The immune system. First of two parts," *N. Engl. J. Med.*, vol. 343, no. 1, pp. 37–49, Jul. 2000, doi: 10.1056/NEJM200007063430107.

[6] B. Alberts, A. Johnson, J. Lewis, M. Raff, K. Roberts, and P. Walter, "Helper T Cells and Lymphocyte Activation," *Mol. Biol. Cell 4th Ed.*, 2002, Accessed: Mar. 19, 2021. [Online]. Available: https://www.ncbi.nlm.nih.gov/books/NBK26827/.

[7] K. Horie *et al.*, "Impact of spaceflight on the murine thymus and mitigation by exposure to artificial gravity during spaceflight," *Sci. Rep.*, vol. 9, no. 1, Art. no. 1, Dec. 2019, doi: 10.1038/s41598-019-56432-9.

[8] D. S. Gridley *et al.*, "Spaceflight effects on T lymphocyte distribution, function and gene expression," *J. Appl. Physiol.*, vol. 106, no. 1, pp. 194–202, Jan. 2009, doi: 10.1152/japplphysiol.91126.2008.

[9] E. G. Novoselova *et al.*, "Changes in immune cell signalling, apoptosis and stress response functions in mice returned from the BION-M1 mission in space," *Immunobiology*, vol. 220, no. 4, pp. 500–509, Apr. 2015, doi: 10.1016/j.imbio.2014.10.021.

[10] B. Morukov, M. Rykova, E. Antropova, T. Berendeeva, S. Ponomaryov, and I. Larina, "T-cell immunity and cytokine production in cosmonauts after long-duration space flights," *Acta Astronaut.*, vol. 68, no. 7, pp. 739–746, Apr. 2011, doi: 10.1016/j.actaastro.2010.08.036.

[11] T. W. Lebsack *et al.*, "Microarray analysis of spaceflown murine thymus tissue reveals changes in gene expression regulating stress and glucocorticoid receptors," *J. Cell. Biochem.*, vol. 110, no. 2, pp. 372–381, 2010, doi: https://doi.org/10.1002/jcb.22547.

[12] C. Molnar and J. Gair, "6.2 The Cell Cycle," in *Concepts of Biology - 1st Canadian Edition*, BCcampus, 2015.

[13] D. S. Gridley *et al.*, "Changes in Mouse Thymus and Spleen after Return from the STS-135 Mission in Space," *PLOS ONE*, vol. 8, no. 9, p. e75097, Sep. 2013, doi: 10.1371/journal.pone.0075097.

[14] P. Meier, A. Finch, and G. Evan, "Apoptosis in Development," *Nature*, vol. 407, pp. 796–801, Nov. 2000, doi: 10.1038/35037734.

[15] J. B. Boonyaratanakornkit *et al.*, "Key gravity-sensitive signaling pathways drive T cell activation," *FASEB J. Off. Publ. Fed. Am. Soc. Exp. Biol.*, vol. 19, no. 14, pp. 2020–2022, Dec. 2005, doi: 10.1096/fj.05-3778fje.

[16] S. K. Mehta, R. J. Cohrs, B. Forghani, G. Zerbe, D. H. Gilden, and D. L. Pierson, "Stress-induced subclinical reactivation of varicella zoster virus in astronauts," *J. Med. Virol.*, vol. 72, no. 1, pp. 174–179, 2004, doi: https://doi.org/10.1002/jmv.10555.

[17] "GeneLab Project Overview," *NASA GeneLab*. https://genelab.nasa.gov/overview (accessed Mar. 20, 2021).

[18] E. G. Overbey *et al.*, "NASA GeneLab RNA-Seq Consensus Pipeline: Standardized Processing of Short-Read RNA-Seq Data," *bioRxiv*, p. 2020.11.06.371724, Nov. 2020, doi: 10.1101/2020.11.06.371724.

[19] 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.

[20] "Why Nextflow Tower?," Oct. 20, 2020. https://help.tower.nf/docs/welcome/introduction/#why-nextflow-tower (accessed Mar. 25, 2021).

[21] Ohno H, Akiyama T. "Impact of spaceflight on gene expression in the thymus", GeneLab, Version 5, doi: http://doi.org/10.26030/s0ag-yx79

[22] D. Sims, I. Sudbery, N. E. Ilott, A. Heger, and C. P. Ponting, "Sequencing depth and coverage: key considerations in genomic analyses," *Nat. Rev. Genet.*, vol. 15, no. 2, Art. no. 2, Feb. 2014, doi: 10.1038/nrg3642.

[23] M. I. Love, W. Huber, and S. Anders, "Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2," *Genome Biol.*, vol. 15, no. 12, p. 550, Dec. 2014, doi: 10.1186/s13059-014-0550-8.

[24] 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.

[25] Y. Benjamini and Y. Hochberg, "Controlling The False Discovery Rate - A Practical And Powerful Approach To Multiple Testing," *J R. Stat. Soc Ser. B*, vol. 57, pp. 289–300, Nov. 1995, doi: 10.2307/2346101.

[26] D. W. Huang, B. T. Sherman, and R. A. Lempicki, “Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists,” *Nucleic Acids Res.*, vol. 37, no. 1, pp. 1–13, Jan. 2009, doi: 10.1093/nar/gkn923.

[27] 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.

[28] X. Jiao *et al.*, “DAVID-WS: a stateful web service to facilitate gene/protein list analysis,” *Bioinformatics*, vol. 28, no. 13, pp. 1805–1806, Jul. 2012, doi: 10.1093/bioinformatics/bts251.

[29] M. Ashburner *et al.*, “Gene Ontology: tool for the unification of biology,” *Nat. Genet.*, vol. 25, no. 1, Art. no. 1, May 2000, doi: 10.1038/75556.

[30] E. Eden, R. Navon, I. Steinfeld, D. Lipson, and Z. Yakhini, “GOrilla: a tool for discovery and visualization of enriched GO terms in ranked gene lists,” *BMC Bioinformatics*, vol. 10, no. 1, p. 48, Feb. 2009, doi: 10.1186/1471-2105-10-48.

[31] 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.

[32] P. Shannon *et al.*, “Cytoscape: a software environment for integrated models of biomolecular interaction networks,” *Genome Res.*, vol. 13, no. 11, pp. 2498–2504, Nov. 2003, doi: 10.1101/gr.1239303.

[33] G. Bindea *et al.*, “ClueGO: a Cytoscape plug-in to decipher functionally grouped gene ontology and pathway annotation networks,” *Bioinformatics*, vol. 25, no. 8, p. 1091, Apr. 2009, doi: 10.1093/bioinformatics/btp101.

[34] 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.

[35] “Per Base Sequence Content.” https://www.bioinformatics.babraham.ac.uk/projects/fastqc/Help/3%20Analysis%20Modules/4%20Per%20Base%20Sequence%20Content.html (accessed May 05, 2021).

[36] “DNA Packaging: Nucleosomes and Chromatin | Learn Science at Scitable.” https://www.nature.com/scitable/topicpage/dna-packaging-nucleosomes-and-chromatin-310/ (accessed Mar. 26, 2021).

[37] “Cell Cycle,” *Genome.gov*. https://www.genome.gov/genetics-glossary/Cell-Cycle (accessed Mar. 27, 2021).

[38] P. L. Jones and A. P. Wolffe, “Relationships between chromatin organization and DNA methylation in determining gene expression,” *Semin. Cancer Biol.*, vol. 9, no. 5, pp. 339–347, Oct. 1999, doi: 10.1006/scbi.1999.0134.

[39] A. Liberzon, C. Birger, H. Thorvaldsdóttir, M. Ghandi, J. P. Mesirov, and P. Tamayo, "The Molecular Signatures Database (MSigDB) hallmark gene set collection," *Cell Syst.*, vol. 1, no. 6, pp. 417–425, Dec. 2015, doi: 10.1016/j.cels.2015.12.004.

[40] R. Kumari, K. P. Singh, and J. W. DuMond, "Simulated microgravity decreases DNA repair capacity and induces DNA damage in human lymphocytes," *J. Cell. Biochem.*, vol. 107, no. 4, pp. 723–731, 2009, doi: https://doi.org/10.1002/jcb.22171.

[41] G. C. Sen, "Viruses and Interferons," *Annu. Rev. Microbiol.*, vol. 55, no. 1, pp. 255–281, 2001, doi: 10.1146/annurev.micro.55.1.255.

[42] S. B. Mizel, "The interleukins," *FASEB J.*, vol. 3, no. 12, pp. 2379–2388, 1989, doi: https://doi.org/10.1096/fasebj.3.12.2676681.

[43] A. Rose, J. M. Steffen, X. J. Musacchia, A. D. Mandel, and G. Sonnenfeld, "Effect of Antiorthostatic Suspension on Interferon-α/β Production by the Mouse," *Proc. Soc. Exp. Biol. Med.*, vol. 177, no. 2, pp. 253–256, Nov. 1984, doi: 10.3181/00379727-177-41939.

[44] M. Tálas *et al.*, "Results of space experiment program 'interferon,'" *Acta Astronaut.*, vol. 11, no. 7, pp. 379–386, Jul. 1984, doi: 10.1016/0094-5765(84)90078-X.

[45] "TMF - Mouse Husbandry, Breeding and Development," Jul. 04, 2007. https://web.archive.org/web/20070704035747/http://www.research.uci.edu/tmf/husbandry.htm#guidelines (accessed Mar. 25, 2021).

[46] A. Hekmat, M. S. Manesh, Z. Hajebrahimi, and S. Hatamie, "Microgravity-Induced Alterations in the H3.3B (H3F3B) Gene Expression and the Histone H3 Structure," *Adv. Sci. Eng. Med.*, vol. 12, no. 8, pp. 1084–1094, Aug. 2020, doi: 10.1166/asem.2020.2672.

[47] G. Felsenfeld and M. Groudine, "Controlling the double helix," *Nature*, vol. 421, no. 6921, Art. no. 6921, Jan. 2003, doi: 10.1038/nature01411.

[48] "Nucleosome," *Genome.gov*. https://www.genome.gov/genetics-glossary/Nucleosome (accessed Apr. 16, 2021).

[49] M. J. Saunders, E. Yeh, M. Grunstein, and K. Bloom, "Nucleosome depletion alters the chromatin structure of Saccharomyces cerevisiae centromeres.," *Mol. Cell. Biol.*, vol. 10, no. 11, pp. 5721–5727, Nov. 1990, doi: 10.1128/MCB.10.11.5721.

[50] F. Prado, S. Jimeno-González, and J. C. Reyes, "Histone availability as a strategy to control gene expression," *RNA Biol.*, vol. 14, no. 3, pp. 281–286, May 2016, doi: 10.1080/15476286.2016.1189071.

[51] J. Chen, "The Cell-Cycle Arrest and Apoptotic Functions of p53 in Tumor Initiation and Progression," *Cold Spring Harb. Perspect. Med.*, vol. 6, no. 3, Mar. 2016, doi: 10.1101/cshperspect.a026104.

[52] "HALLMARK_E2F_TARGETS." https://www.gsea-msigdb.org/gsea/msigdb/cards/HALLMARK_E2F_TARGETS (accessed Apr. 03, 2021).

[53] D. G. Johnson and R. Schneider-Broussard, "Role of E2F in cell cycle control and cancer," *Front. Biosci. J. Virtual Libr.*, vol. 3, pp. d447-448, Apr. 1998, doi: 10.2741/a291.

[54] "HALLMARK_G2M_CHECKPOINT." https://www.gsea-msigdb.org/gsea/msigdb/cards/HALLMARK_G2M_CHECKPOINT (accessed Apr. 03, 2021).

[55] "Cell Cycle Checkpoints | Biology for Majors I." https://courses.lumenlearning.com/wmopen-biology1/chapter/cell-cycle-checkpoints/ (accessed Apr. 03, 2021).

[56] M. R. I. TESTARD, "Radiation-induced chromosome damage in astronauts' lymphocytes," *Int. J. Radiat. Biol.*, Jul. 2009, Accessed: Apr. 05, 2021. [Online]. Available: https://www.tandfonline.com/doi/abs/10.1080/095530096144879.

[57] G. Obe, I. Johannes, C. Johannes, K. Hallman, G. Reitz, and R. Facius, "Chromosomal aberrations in blood lymphocytes of astronauts after long-term space flights," *Int. J. Radiat. Biol.*, vol. 72, no. 6, pp. 727–734, Dec. 1997, doi: 10.1080/095530097142889.

[58] T. C. Yang, K. George, A. S. Johnson, M. Durante, and B. S. Fedorenko, "Biodosimetry Results from Space Flight Mir-18," *Radiat. Res.*, vol. 148, no. 5, pp. S17–S23, 1997, doi: 10.2307/3579712.

[59] W. P. Kloosterman and R. Hochstenbach, "Deciphering the pathogenic consequences of chromosomal aberrations in human genetic disease," *Mol. Cytogenet.*, vol. 7, Dec. 2014, doi: 10.1186/s13039-014-0100-9.

[60] A. Theisen and L. G. Shaffer, "Disorders caused by chromosome abnormalities," *Appl. Clin. Genet.*, vol. 3, pp. 159–174, Dec. 2010, doi: 10.2147/TACG.S8884.

[61] G. Horneck, "Impact of microgravity on radiobiological processes and efficiency of DNA repair," *Mutat. Res. Mol. Mech. Mutagen.*, vol. 430, no. 2, pp. 221–228, Dec. 1999, doi: 10.1016/S0027-5107(99)00133-5.

[62] M. Moreno-Villanueva, M. Wong, T. Lu, Y. Zhang, and H. Wu, "Interplay of space radiation and microgravity in DNA damage and DNA damage response," *Npj Microgravity*, vol. 3, no. 1, Art. no. 1, May 2017, doi: 10.1038/s41526-017-0019-7.

[63] O. D. Schärer, "Chemistry and Biology of DNA Repair," *Angew. Chem. Int. Ed.*, vol. 42, no. 26, pp. 2946–2974, 2003, doi: https://doi.org/10.1002/anie.200200523.

[64] R. Medzhitov, "Inflammation 2010: New Adventures of an Old Flame," *Cell*, vol. 140, no. 6, pp. 771–776, Mar. 2010, doi: 10.1016/j.cell.2010.03.006.

[65] T. T. Chang, S. M. Spurlock, T. L. T. Candelario, S. M. Grenon, and M. Hughes-Fulford, "Spaceflight impairs antigen-specific tolerance induction in vivo and increases inflammatory cytokines," *FASEB J.*, vol. 29, no. 10, pp. 4122–4132, 2015, doi: https://doi.org/10.1096/fj.15-275073.

[66] B. Crucian, R. P. Stowe, S. Mehta, H. Quiriarte, D. Pierson, and C. Sams, “Alterations in adaptive immunity persist during long-duration spaceflight,” *Npj Microgravity*, vol. 1, no. 1, Art. no. 1, Sep. 2015, doi: 10.1038/npjmgrav.2015.13.

[67] D. S. Kim *et al.*, “The effect of microgravity on the human venous system and blood coagulation: a systematic review,” *Exp. Physiol.*, Mar. 2021, doi: 10.1113/EP089409.

[68] K. Marshall-Goebel *et al.*, “Assessment of Jugular Venous Blood Flow Stasis and Thrombosis During Spaceflight,” *JAMA Netw. Open*, vol. 2, no. 11, p. e1915011, Nov. 2019, doi: 10.1001/jamanetworkopen.2019.15011.

[69] C. T. Esmon, J. Xu, and F. Lupu, “Innate immunity and coagulation,” *J. Thromb. Haemost.*, vol. 9, no. s1, pp. 182–188, 2011, doi: https://doi.org/10.1111/j.1538-7836.2011.04323.x.

[70] R. Roskoski, “Cyclin-dependent protein serine/threonine kinase inhibitors as anticancer drugs,” *Pharmacol. Res.*, vol. 139, pp. 471–488, Jan. 2019, doi: 10.1016/j.phrs.2018.11.035.