A Characterization of Cephalad Transcriptome Changes in the Spaceflight Environment and Mapping the Epitranscriptomic Landscape of Pluripotent Stem Cells

Eliah G. Overbey

A dissertation
submitted in partial fulfillment of the
requirements for the degree of

Doctor of Philosophy

University of Washington
2021

Reading Committee:
R David Hawkins, Chair
Carol Ware
Phil Green

Program Authorized to Offer Degree:
Genome Sciences

University of Washington

Abstract

A Characterization of Cephalad Transcriptome Changes in the Spaceflight Environment and Mapping the Epitranscriptomic Landscape of Pluripotent Stem Cells

Eliah G. Overbey

Chair of the Supervisory Committee:
Associate Professor R. David Hawkins
Division of Medical Genetics, Department of Medicine
Department of Genome Sciences

Throughout the process of cellular differentiation, cell trajectories are shaped by different epigenomic and epitranscriptomic regulatory mechanisms. These trajectories eventually land cells into their terminal cell state, where they are no longer undergoing differentiation and have a stable gene expression profile. Throughout my thesis, I profiled gene expression of various cell states examining the impact of both natural variation between cell types and the effects of environmental perturbation. Specifically, I examined the natural variation found within chicken tissues and found a variety of alternatively spliced transcripts and differentially expressed genes between different cell and tissue types. To assess the effects of environmental perturbation, I examined the effects of a growing area of interest, the spaceflight environment, on mammalian cephalad cell types. From these experiments, I discovered that the combinatorial effect of spaceflight factors, specifically low-dose radiation and microgravity simulation, produce many more differentially expressed genes in murine brain tissue between the spaceflight simulated group and the control group than either low-dose radiation or microgravity simulation produced

in isolation. The results of this experiment demonstrated that there is a compounding effect of the different environmental factors of spaceflight on mammalian brain tissue. The second experiment I analyzed on the spaceflight environment were the gene expression profiles of retinal tissue from mice flown aboard the International Space Station. Compared to the ground control, mice flown in spaceflight had hundreds of differentially expressed genes, a handful of which overlapped with genes associated with the cone-rod degenerative disease, retinitis pigmentosa. After examining differential gene expression, due to both natural variation and the environmental perturbations caused by spaceflight, I examine how gene expression is regulated post-transcriptionally. I did this by performing direct-RNA sequencing using the Oxford Nanopore MinION and employing base-modification prediction software in order to determine the global RNA modification landscape across three distinct states of pluripotent stem cells. Doing this, I uncover that pluripotency-relevant genes are enriched for transcript modifications. I then used the known motif DRACH to specifically locate m^6A modifications. From this, I can identify uniquely modified genes in each pluripotent state. Collectively, these studies demonstrate the spectrum of transcriptional variation that can arise from natural and perturbed cell states, as well as the importance of incorporating mRNA modifications into analysis to understand the transitions and uniqueness between cellular states.

Acknowledgements

I learned more during my time in graduate school than any other period in my life. When I started the program, I relied on the expertise and patience of my friends, colleagues, and mentors to aid my development into a well-rounded molecular and computational biologist.

Thank you to:

Sanjay Srivatsan for being the first person to put a pipette in my hand during the Spring Break of my first year in the program, Cristina Valensisi for teaching me the complicated ins-and-outs of stem cell culture and differentiation, Robert Azad for helping me develop my benchwork technique, Arpit Mishra for working with me to refine Hi-C protocols, and Stephanie Battle helping me perfect a variety of benchwork techniques and for the guidance provided throughout graduate school.

The Institute for Stem Cell and Regenerative Medicine for further developing my understanding of stem cell biology and culture methods. Specifically, Savannah Cook, Chris Cavanaugh, and Jenn Hesson. In particular, I am appreciative of the support of Carol Ware, who distilled her expertise in stem cell biology clearly and precisely to a student from a completely different scientific background.

Bill Noble for the educational opportunities provided through the NIH Big Data in Genomics and Neuroscience (BDGN) training grant, which provided me the opportunity to refine my machine learning, statistics, and data science skills.

My mentors and collaborators at NASA Ames Research Center: Jon Galazka, Amber Paul, Candice Tahimic, Cassandra Juran, Sigrid Reinsch, Sylvain Costes, Amanda Saravia-Butler, and Matt McNulty. Also my collaborators on spaceflight-related projects at other institutions: Xiao “Vivien” Mao, Willian Abraham da Silveira, and Gary Hardiman.

The Interdepartmental Astrobiology Program for providing support and opportunities throughout graduate school. Particularly the enthusiasm of Victoria Meadows to include the Genome Sciences department in the program.

My committee: David Hawkins, Phil Green, Carol Ware, Bill Noble, Lea Starita, and Jodi Young, for the valuable feedback, inspiration, mentorship, and academic enrichment.

My advisor and mentor David Hawkins, who was able to strike the perfect balance between project guidance and independent development; for taking on the task of training a computational biologist to learn molecular biology benchwork; and for supporting me throughout all my training opportunities in genomics, data science, and astrobiology.

And lastly, my parents for the unconditional support and encouragement they have provided over the past 5 ½ years.

TABLE OF CONTENTS

Introduction

Every mammalian organism is composed of a multitude of cell types. Some of these cells are similar in functionality to other cells, such as the variance present on the spectrum of immune cells. Other cell types perform radically different functions, such as the difference in functionality of cells in the digestive tract versus cells of the reproductive system. Despite the broad range of functions that cells perform, with few exceptions, they all contain the exact same composition of DNA. The same bases, A/C/G/T, ordered in the exact same way, on the exact same chromosomes. Yet, from identical compositions of DNA vastly different cellular functions arise.

Cellular functions are established and regulated by epigenomic mechanisms. These mechanisms alter the conformation of DNA and use base modifications to affect the accessibility of DNA to transcriptional machinery, which in-turn regulates subsequent gene expression. Additionally, epitranscriptomic modifications can alter gene expression post-transcription. These modifications are deposited on multiple biotypes of RNA. With regard to mRNA, they can change RNA folding and impact translation efficiency. These different control mechanisms are used to guide a cell through differentiation and stabilize the cell in its terminal cell type.

In my thesis, I explore the natural variation in gene expression that exists between different cell types (Chapter 1), examples of how this natural variation can be perturbed when environmental conditions are altered (Chapters 2-4), and how the post-transcriptional, epitranscriptomic modification landscape has distinct variability during early developmental states (Chapter 5). Lastly, in the concluding remarks, I will argue why the method of epitranscriptome detection applied in Chapter 5 should be the standard method of RNA sequencing aboard the International Space Station.

Natural Variation of Gene Expression

Throughout the history of epigenetics, the process of cellular differentiation has been compared to a marble rolling down a hill. Depending on which grooves in the hill it hits during its journey, the marble will reach different terminal-positions. This metaphor, commonly referred to as Waddington's landscape (Waddington and Others 1940) provides a framework for understanding how pluripotent cells reach their terminal cell types. Cells start in a pluripotent state and through a series of divisions they bifurcate into distinct cell states, eventually reaching a terminal state where differentiation has ceased and they perform their intended function.The variation in gene expression between these terminal states is described here as natural variation.

The natural variation between different tissues has been well characterized in humans and mice (Lin et al., 2014). However, our understanding of gene expression in most mammalian and other vertebrate species is still lacking. The mammalian species that are best characterized are the species that are important for modeling and understanding human health and development. However, understanding a variety of other species can be indirectly important for improving human health.

Such understudied organisms include those that are a part of the food supply. In Chapter 1, I explore the natural variation in gene expression that arises in a spectrum of different cell and tissue types in domesticated *Gallus gallus,* the species of chicken whose meat and eggs are used extensively in the U.S. food supply. Even though the draft genome of *Gallus gallus* was published in 2004 (International Chicken Genome Sequencing Consortium 2004), much work remains to be done on understanding the differences in gene expression and the regulatory control mechanisms that establish cellular and tissue identity. This chapter is the first of a wider series of papers from the Hawkins lab that will continue to characterize gene expression and map the epigenomic regulatory features of the chicken genome.

The Impact of Spaceflight on Gene Expression

Gene expression profiles can also vary due to environmental variables. Environmental factors are able to perturb gene expression of terminal cell types and create unique gene expression profiles. The environmental factors I focus on in Chapters 2-4 are the factors imposed by spaceflight.

In recent years, there has been renewed interest in manned space exploration. The United States has had a continuous presence in low-Earth orbit since the construction of the International Space Station (ISS) in 2000. While the ISS was initially used for governmental research and technology development, the goals and priorities of the ISS have shifted in recent years. The ISS now has dedicated space to industry and academia for research and technology development. Additionally, they are prioritizing the commercialization of low-Earth orbit in the goal of creating an economically sustainable spaceflight sector.

Developments in private companies have further accelerated this progress. Manned rockets, once only capable of being created by the government, have now been successfully launched by SpaceX and Blue Origin. Space station modules are close to costing a fraction of their original price when the ISS was first constructed, attributable the progress of Bigelow Aerospace's expandable activity modules. There are also numerous companies refining the habitats and scientific equipment used in spaceflight, such as NanoRacks, Space Tango, Axiom Space, Made in Space, among many others. While technology development to facilitate space

utilization is accelerating rapidly, our understanding of the consequences of spaceflight on the human body is lagging behind.

Since the dawn of NASA's space program, physiological data has been collected from astronauts. A limited list of changes in physiology includes decreases in bone density (Stavnichuk et al., 2020), deterioration of muscle mass (Trappe et al., 2009; Burkhart, Allaire, and Bouxsein, 2019), changes in vision (Lee et al., 2020), and problems with neurocognitive abilities (De la Torre 2014). Even though the physiological consequences of spaceflight exposure are well-observed, the fundamental molecular mechanisms driving these changes are still poorly understood, hindering our ability to develop appropriate countermeasures and pharmaceutical solutions. To address this, a growing number of genomic experiments are being conducted aboard the ISS.

The renowned genomic study performed has been the NASA twins study (Garrett-Bakelman et al., 2019). This project was a landmark study conducted to characterize numerous genomic biomarkers, such as epigenetic profiles, telomere lengths, and gene expression changes between one twin who resided on the ISS for one year and his twin who served as a control back on Earth. However, this study has serious limitations. Much of the raw data from this study is not publicly available since the identity of the astronaut and his twin are well known. In addition, the study represents a sample size of 1, and, subsequently, does not have high statistical power. Also, given the limited capacity of the ISS, reaching a sample size of high significance is difficult. Despite these limitations, the twins study does present a strong case for why it is critical to better characterize the consequences of long-duration spaceflight on the human body and develop countermeasures before establishing permanent settlements aboard space stations or other planets. But these limitations also restrict the more rigorous scientific studies to experiments on non-human organisms.

In Chapter 2, I am seeking to understand the combinatorial effects of different spaceflight factors on murine gene expression in the brain using a ground-based study. Specifically, I am looking at the brain's ability to recover from two spaceflight factors individually and when combined together. After four months of recovery, these factors, microgravity and radiation, individually do not have a large impact on the brain's gene expression profile. However, in conjunction, these factors result in over one hundred differentially expressed genes. This chapter emphasizes insufficiency of ground based experiments that study spaceflight factors in isolation and stresses the need for more laboratory space in low-Earth orbit.

In Chapter 3, I move on to discuss changes in the gene expression of retinal tissue of mice flown in low-Earth orbit aboard the ISS for 35 days. Immunofluorescence staining of the retinal tissue showed disruption of cone photoreceptors and disorganization of the retinal

pigment epithelium. Gene expression analysis in the tissue showed 600 differentially expressed genes between the spaceflight-exposed mice and the ground control group. Of particular interest is the presence of multiple genes related to retinitis pigmentosa, a photoreceptor degenerative disease, that were differentially expressed in the group of mice exposed to spaceflight.

In Chapter 4, I address one of the largest hurdles in analyzing gene expression data from spaceflight: the array of differing variables and design decisions between spaceflight experiments. While there is no method to predict how different exposure lengths to spaceflight or different strains of organism used in an experiment are impacting gene expression, the analysis pipeline used on the data can be harmonized by creating a standard computational pipeline for all gene expression data generated on the ISS. This chapter is a description of the computational pipeline for short-read RNA-seq data that is implemented in the NASA GeneLab database. This pipeline emerged from a GeneLab workshop spanning many members of the spaceflight community from universities all over the world that was further refined and implemented at NASA. All RNA-seq data submitted to the GeneLab database will be run through this pipeline to help facilitate cross-experiment comparisons.

More limited than gene expression data in spaceflight are studies of epigenomic and epitranscriptomic changes. The GeneLab database has limited bisulfite-sequencing data available to characterize 5-methylcytosine modifications. However, ChIP-seq, ATAC-seq, Hi-C, and other methods of determining chromatin conformation and regulation remain to be generated.

The Epitranscriptome and Direct-RNA Sequencing

While gene expression is critical for determining cellular function and identity, regulation of gene expression does not end at transcription. Modifications can be placed onto RNA molecules post-transcriptionally and these modifications influence the metabolism and lifespan of the RNA molecule, particularly for mRNA.

Traditionally, RNA modifications have been challenging to detect and are limited to the detection of a single modification at a time. However, direct-RNA sequencing with the Oxford Nanopore MinION has been proposed as a method for simultaneously detecting all modifications on poly(A) RNA. In recent years, a small handful of tools have been developed in order to identify base modifications from the raw current signal collected by direct-RNA sequencing (Jenjaroenpun et al., 2021; Liu et al., 2019; Stoiber et al., 2016).

In Chapter 5, I apply one of these tools to detect base modifications across multiple states of pluripotent stem cells. Both the global epitranscriptome and the m^6A-specific landscape are reported. The m^6A modification is linked to transcript turnover (Batista et al., 2014) and is of

particular interest in pluripotent stem cells. Previous studies that knocked down m^6A writing proteins in various stem cell states have found differing effects on differentiation. One set of studies concluded that the m^6A modification is necessary for differentiation (Batista et al., 2014; Geula et al., 2015), while another set of studies found that it induced differentiation (Wen et al., 2018; Wu et al., 2019; Aguilo et al., 2015). My results indicate that different pluripotent states do, indeed, have unique m^6A landscapes, which could be a possible explanation for the conflicting results previously observed.

References

Aguilo, Francesca, Fan Zhang, Ana Sancho, Miguel Fidalgo, Serena Di Cecilia, Ajay Vashisht, Dung-Fang Lee, et al. 2015. "Coordination of m6A mRNA Methylation and Gene Transcription by ZFP217 Regulates Pluripotency and Reprogramming." *Cell Stem Cell* 17 (6): 689–704.

Batista, Pedro J., Benoit Molinie, Jinkai Wang, Kun Qu, Jiajing Zhang, Lingjie Li, Donna M. Bouley, et al. 2014. "m6A RNA Modification Controls Cell Fate Transition in Mammalian Embryonic Stem Cells." *Cell Stem Cell* 15 (6): 707–19.

Burkhart, Katelyn, Brett Allaire, and Mary L. Bouxsein. 2019. "Negative Effects of Long-Duration Spaceflight on Paraspinal Muscle Morphology." *Spine* 44 (12): 879–86.

De la Torre, Gabriel G. 2014. "Cognitive Neuroscience in Space." *Life* 4 (3): 281–94.

Garrett-Bakelman, Francine E., Manjula Darshi, Stefan J. Green, Ruben C. Gur, Ling Lin, Brandon R. Macias, Miles J. McKenna, et al. 2019. "The NASA Twins Study: A Multidimensional Analysis of a Year-Long Human Spaceflight." *Science* 364 (6436). https://doi.org/10.1126/science.aau8650.

Geula, Shay, Sharon Moshitch-Moshkovitz, Dan Dominissini, Abed Alfatah Mansour, Nitzan Kol, Mali Salmon-Divon, Vera Hershkovitz, et al. 2015. "Stem Cells. m6A mRNA Methylation Facilitates Resolution of Naïve Pluripotency toward Differentiation." *Science* 347 (6225): 1002–6.

International Chicken Genome Sequencing Consortium. 2004. "Sequence and Comparative Analysis of the Chicken Genome Provide Unique Perspectives on Vertebrate Evolution." *Nature* 432 (7018): 695–716.

Jenjaroenpun, Piroon, Thidathip Wongsurawat, Taylor D. Wadley, Trudy M. Wassenaar, Jun Liu, Qing Dai, Visanu Wanchai, et al. 2021. "Decoding the Epitranscriptional Landscape from Native RNA Sequences." *Nucleic Acids Research*. https://doi.org/10.1093/nar/gkaa620.

Lee, Andrew G., Thomas H. Mader, C. Robert Gibson, William Tarver, Pejman Rabiei, Roy F. Riascos, Laura A. Galdamez, and Tyson Brunstetter. 2020. "Spaceflight Associated Neuro-Ocular Syndrome (SANS) and the Neuro-Ophthalmologic Effects of Microgravity: A Review and an Update." *Npj Microgravity*. https://doi.org/10.1038/s41526-020-0097-9.

Lin, Shin, Yiing Lin, Joseph R. Nery, Mark A. Urich, Alessandra Breschi, Carrie A. Davis, Alexander Dobin, et al. 2014. "Comparison of the Transcriptional Landscapes between Human and Mouse Tissues." *Proceedings of the National Academy of Sciences of the United States of America* 111 (48): 17224–29.

Liu, Huanle, Oguzhan Begik, Morghan C. Lucas, Jose Miguel Ramirez, Christopher E. Mason, David Wiener, Schraga Schwartz, John S. Mattick, Martin A. Smith, and Eva Maria Novoa. 2019. "Accurate Detection of m6A RNA Modifications in Native RNA Sequences." *Nature Communications* 10 (1): 4079.

Stavnichuk, Mariya, Nicholas Mikolajewicz, Tatsuya Corlett, Martin Morris, and Svetlana V. Komarova. 2020. "A Systematic Review and Meta-Analysis of Bone Loss in Space Travelers." *Npj Microgravity*. https://doi.org/10.1038/s41526-020-0103-2.

Stoiber, M., J. Quick, R. Egan, J. E. Lee, and S. Celniker. 2016. “De Novo Identification of DNA Modifications Enabled by Genome-Guided Nanopore Signal Processing.” *BioRxiv*. https://www.biorxiv.org/content/10.1101/094672v1.abstract.

Trappe, Scott, David Costill, Philip Gallagher, Andrew Creer, Jim R. Peters, Harlan Evans, Danny A. Riley, and Robert H. Fitts. 2009. “Exercise in Space: Human Skeletal Muscle after 6 Months Aboard the International Space Station.” *Journal of Applied Physiology* 106 (4): 1159–68.

Waddington, Conrad Hal, and Others. 1940. “Organisers and Genes.” *Organisers and Genes.* https://www.cabdirect.org/cabdirect/abstract/19401601308.

Wen, Jing, Ruitu Lv, Honghui Ma, Hongjie Shen, Chenxi He, Jiahua Wang, Fangfang Jiao, et al. 2018. “Zc3h13 Regulates Nuclear RNA m6A Methylation and Mouse Embryonic Stem Cell Self-Renewal.” *Molecular Cell* 69 (6): 1028–38.e6.

Wu, Ruifan, Youhua Liu, Yuanling Zhao, Zhen Bi, Yongxi Yao, Qing Liu, Fengqin Wang, Yizhen Wang, and Xinxia Wang. 2019. “m6A Methylation Controls Pluripotency of Porcine Induced Pluripotent Stem Cells by Targeting SOCS3/JAK2/STAT3 Pathway in a YTHDF1/YTHDF2-Orchestrated Manner.” *Cell Death & Disease*. https://doi.org/10.1038/s41419-019-1417-4.

Chapter 1: Transcriptomes of an array of chicken ovary, intestinal, and immune cells and tissues

Chapter 1 is adapted with minimal modification from:

Transcriptomes of an array of chicken ovary, intestinal, and immune cells and tissues. Eliah G. Overbey, Theros T. Ng, *et al.* (2021); In Review, Frontiers in Genetics.

1.1 Abstract

While the chicken (*Gallus gallus*) is the most consumed agricultural animal worldwide, the chicken transcriptome remains understudied. We have characterized the transcriptome of 10 cell and tissue types from the chicken using RNA-seq, spanning intestinal tissues (ileum, jejunum, proximal cecum), immune cells (B cells, bursa, macrophages, monocytes, spleen T cells, thymus), and reproductive tissue (ovary). We detected 17,872 genes and 24,812 transcripts across all cell and tissue types, representing 73% and 63% of the genome respectively. Further quantification of RNA transcript biotypes revealed protein coding and lncRNAs specific to an individual cell/tissue type. Each cell/tissue type also has an average of around 1.4 isoforms per gene, however, they all have at least one gene with at least 11 isoforms. Differential expression analysis revealed a large number of differentially expressed genes between tissues of the same category (immune and intestinal). Many of these differentially expressed genes in immune cells were involved in cellular processes relating to differentiation and cell metabolism as well as basic functions of immune cells such as cell adhesion and signal transduction. The differential expressed genes of the different segments of the chicken intestine (jejunum, ileum, proximal cecum) correlated to the metabolic processes in nutrient digestion and absorption. These data should provide a valuable resource in understanding the chicken genome.

1.2 Introduction

In the United States, over 9 billion broiler chickens, which is estimated to be about 19 billion kilograms of chicken products, are produced per year (NCC, 2019). Egg production totaled to about 99.1 billion in 2019 in the US (UEP, 2019). Apart from the important role in food production, the chicken has been used as an animal model to benefit key areas in functional human research including immunology (Glick et al., 1956), vaccine development (Matthews, 2006), reproduction (Nap et al., 2003; Nap et al., 2004; Bédécarrats et al., 2016), and nutrition (Klasing, 1984; Shang et al., 2018). The process to improve the annotation of the chicken is on-going since it was first sequenced in 2004. As sequencing and data science technologies rapidly evolve, new tools allow for a more accurate representation of the chicken genome. The Functional Annotation of Animal Genomes (FAANG) project was launched to comprehensively

characterize the genome of farm animals to address the sustainable agriculture of farmed animals (Giuffra et al., 2019). The current study under the FAANG project focuses on the accurate annotation of the coding and long non-coding (LNC) RNA transcripts of various cells and tissues.

The chicken karyotype consists of 38 autosomes and two sex chromosomes (Z and W). The first drafted chicken genome was sequenced using whole-genome shotgun sequencing of a female Red Jungle Fowl, which is the closest wild variant of the domestic chicken and was 1.05 Gb in length (International Chicken Genome Sequencing Consortium, 2004) (Schmid et al., 2000). The current version of the chicken genome (Gallus_gallus-6.0; GCCA_000002315.5) was sequenced using the combined long single-molecule sequencing technology, and improved BAC and physical maps (Warren et al., 2017). This resulted in the increase of genome size to 1.21 Gb, accounting for micro-chromosomes that were not accounted for or incorrectly assembled in the previous version (Cheng and Burt, 2018). The coding and non-coding regions, as well as the regulatory elements, of the chicken genome is the current focus in annotation studies. Annotation of chicken genes is performed computationally from reference genomes of species that are better annotated. This method is successful in identifying conserved genes across species. However, it is challenging for non-conserved genes because of the relative physiology of the chicken compared to other species, in addition to different genome size, and differences in intron/exon organization between species (Shepard et al., 2009). Our annotation of the chicken genome has 16,779 protein coding genes (28,345 transcripts) and 7,577 lncRNA and other RNA biotypes (10,943 transcripts). Of the 39,288 unique transcripts, 72.1% are protein coding, 22.6% are lncRNAs, 2.9% are miRNAs and 2.4% are other RNA biotypes.

The transcriptome comprises coding RNA (mRNA) and non-coding RNA (ncRNA). While the central dogma has established that coding RNAs are translated into proteins, there continues to be a growing interest in the function of ncRNAs, some of which are not transcribed by RNA polymerase II (Mattick and Makunin, 2006). Recently, it was discovered that ncRNA plays a regulatory role in many biological processes (Zhang et al., 2009). Long non-coding RNAs (lncRNAs), which are non-protein-coding RNAs more than 200 nucleotides in length, play a role in post-transcriptional epigenetic regulation (Quinn and Chang, 2016). In chickens, lncRNA regulates a host of biological functions, including intramuscular adipogenesis (Zhang et al., 2017b;a), sperm motility (Liu et al., 2017), cholesterol synthesis (Muret et al., 2017), and embryonic development (Roeszler et al., 2012). In Avian leukovirus-J (ALV-J) infection, lncRNA regulates macrophages (Dai et al., 2019). and has also been implicated in a subtype of lncRNA, named long intergenic non-coding RNA, in Marek's disease (Han et al., 2017). Therefore, a comprehensive annotation of lncRNA expression in the chicken will reveal regulatory processes relevant to health and disease in an agriculturally important species.

In this study, we aimed to contribute to the catalog of transcriptomic differences of relevant chicken cells and tissues. We focused on multiple immune, intestinal, and

reproduction-related tissues and cells. Specifically, tissue-specific immune cells (lung macrophage, spleen T cells, peripheral monocytes and B-cells), immune organs (bursa and thymus), intestinal sections (jejunum, ileum, and proximal cecum), and ovary of the female reproductive tract were analyzed. The primary immune organs, the bursa and thymus, are the origin of B cells and T cells in chickens, respectively (Cooper et al., 1966). The proximal cecum is located in the intestine at the ileocecal junction between the ileum and colon, is also the secondary immune organ in chickens due to the presence of mucosal-associated lymphoid tissues (MALT), such as the cecal tonsils. The findings described here will be useful towards a complete annotation of chicken tissue and cellular transcriptomes.

1.3 Results

1.3.1 Sample Clustering and PCA

Ten cell and tissue types were profiled with RNA-sequencing with the goal of determining coding and primarily lncRNA expression, as well as isoform usage. All samples were compared to one another using Euclidean distance (Fig 1A) and principal component analysis (PCA) (Fig 1B) using the R package pcaExplorer (Marini, 2016). Replicates of the same tissue had the smallest Euclidean distance between one another (Fig 1A) and the highest Pearson correlation scores (Sup Figure 1). The 500 most variable genes (Fig 1B) were used for PCA. Samples appear to form 3 distinct clusters based on functional category: immune system (B cells, bursa, macrophage (lung), monocytes (blood), T cells (spleen), thymus), reproductive tissue (ovary), and intestinal tissue (jejunum, ileum, proximal cecum). To identify genes highly specific to tissue or cell types, Shannon's entropy was calculated for each gene across all cell types, obtaining a specificity score for each gene. The expression of the 2000 most specific genes were visualized in a heatmap (Fig 1C), revealing that macrophage cells have the most specific gene expression, while ileum tissue and monocytes have the least. When the next 2000 most specific genes are visualized (Sup Fig 2A) we begin to see less tissue specific expression and see genes that are expressed in a small subset of cell types, compared to the 1000 least specific genes (Sup Fig 2B), which show more uniform gene expression across all tissue and cell types. A UCSC browser shot of gene expression across all cell and tissue types shows uniformity of expression among some genes and variable expression among others (Fig 1D).

1.3.2 Transcriptome Coverage and Biotype Detection

Among all samples, 73.4% (17,872) of all known chicken genes and 63.2% (24,812) of all known transcripts from Ensembl annotations were detected (genome build GRCg6a) (Fig 2A). Tissue and cell type specific gene, transcript, and lncRNA counts are provided in Table 1.

Between 9,839 (monocyte) – 14,418 (thymus) genes and 11,522 (monocyte) – 17,794 (proximal cecum) transcripts were detected in each sample (Fig 2B). Out of the fifteen transcript biotypes (protein_coding, lncRNA, miRNA, pseudogene, misc_RNA, snoRNA, snRNA, scaRNA, rRNA, processed_pseudogene, IG_V_gene, Mt_rRNA, Mt_tRNA, ribozyme, sRNA) in the Gallus gallus reference annotation, fourteen were found in the each of the sample types. The largest number of transcripts detected were from protein coding RNA and lncRNA (Fig 2C). 28,345 (90.0%) protein coding and 3,939 (44.4%) lncRNA transcripts were detected among all samples. Even though our library preparation method did not enrich for small RNAs, a low level of these transcripts were detected (Fig 2D). Additionally, protein coding and lncRNA expression unique to each cell or tissue type was detected (Fig 2E; Sup Table 1). All cell and tissue types had a greater number of unique protein coding genes, except for the ovary tissue, which had a higher number of unique lncRNAs. Lung macrophage expressed the most unique protein coding genes (653), whereas jejunum tissue (28) and monocytes (19) expressed the fewest.

1.3.3 Isoform Characterization

Alternative splicing is a primary mechanism for diversifying protein expression. After constructing transcript isoforms from short-read sequencing, Shannon's entropy calculations revealed unique isoforms to each cell and tissue type were found among a set of 500 isoforms (Fig 3A, Sup Table 2). The highest number of unique isoforms were found in T cells. The lowest in monocytes, B cells, and ileum tissue. When expanded to view expressions of the top 1,000 isoforms with the highest specificity, isoforms are less specific to a single cell or tissue type (Sup Fig S2C). In contrast, when the 1,000 isoforms with the least entropy are observed, we see uniform expression among most cell and tissue types (Sup Fig S2D). Each cell and sample type has an average of 1.39 (monocyte) - 1.56 (ovary) isoforms per gene (Fig 3B). Histograms allow us to further visualize the distribution of isoform counts per gene in each tissue (Fig 3C, Sup Figure 3A). Most genes express only a single isoform of around 10,000 for each cell and tissue type. Around 2500 genes per cell type express two isoforms. A small subset of genes expressed more than 5 isoforms of a gene (Fig 3C and Sup Fig 3A insets). There were 590 genes with 5 or more isoforms expressed among all cell and tissue types. The gene with the most isoforms is ENSGALG00000032534 (no gene symbol), which has 24 isoforms in ovary tissue. They fall into GO biological process categories related to cell morphogenesis and structure (Sup Fig 3B). A UCSC browser shot of the gene *PDGFRB* (ENSGAL00000030613) visualizes differences in isoform expression of a single gene among different tissue and cell types (Fig 3D). Bursa, ileum, jejunum, ovary, proximal cecum, and thymus tissue express nearly all exons, whereas B cells, macrophage, monocytes, and T cells express a subset of exons.

1.3.4 Co-Expression on Forward and Reverse Strands

A subset of expression will occur within the same genomic coordinate range on strands opposite to one another. Co-expression of this kind can serve as a feedback mechanism to regulate expression of one another, particularly between lncRNAs and protein coding genes. An example of this is expression of the protein coding gene *FRMPD4* (ENSGALT00000049598) occurring on the strand opposite to lncRNA gene ENSGALT00000098634 (Fig 4A). The number of co-expressed pairs ranged from 371 (monocyte) to 621 (thymus) (Fig 4B, Sup Table 3). The majority of pairs were both protein coding genes for all cell and tissue types (range: 307-454) (Fig 4C). The next most common pairing were protein coding-lncRNA co-expresssion (range: 10-51). A small number of instances were lncRNA-lncRNA co-expression (range: 0-6). Also present were interactions between other biotypes (miRNA, pseudogene, misc_RNA, snoRNA, snRNA, scaRNA, rRNA, processed_pseudogene, IG_V_gene, Mt_rRNA, Mt_tRNA, ribozyme, sRNA) (range: 11-19).

1.3.5 DEG Analysis on Immune and Intestinal Samples

In addition to determining genes and isoforms highly enriched for cell- or tissue-specific expression, we identified genes differentially expressed between related cells or tissues. Differentially expressed genes (DEGs) were computed for six cell/tissue type comparisons using Deseq2. Three of these comparisons were among immune cell samples. There were 4911, 5907, and 3951 DEGs for the comparisons B cells vs monocytes, B cells vs bursa tissue, and bursa tissue vs thymus tissue respectively (Fig 5A, Sup Table 4). A weighted set cover analysis in WebGestalt (Liao et al., 2019) was performed to reduce redundancy and find the most representative GO biological process categories among sample comparisons (Fig 5B). The GO category “response to stress” was the only category shared among all three comparisons. When we compare the DEGs across all three comparisons, we find that there is a subset of genes that are shared across multiple sample comparisons, however, there is a sizable number of genes unique to each tissue comparison (Fig 5C, Sup Table 4). This was also reflected in similarities between enriched GO categories, which shared 33 categories between all three comparisons. Additionally, we see unique sets of genes among the top 10 DEGs for each comparison (Fig 5D).

DEG comparisons were also performed for three comparisons among intestinal samples. There were 3903, 2306, and 4270 DEGs for the comparisons of jejunum tissue vs ileum tissue, jejunum tissue vs proximal cecum tissue, and ileum tissue vs proximal tissue respectively (Fig 6A, Sup Table 5). A weighted set cover analysis was again performed (Fig 6B). There were no

overlaps of enriched GO categories in the weighted set cover or among the sets of all GO terms enriched for each cell type, despite seeing 332 differentially expressed genes shared between all tissue comparisons (Figure 6B, 6C, Sup Table 5). Among the sets of the top 10 differentially expressed genes for each tissue comparison, we observe the genes *APOA4* and *LCT* are present for tissue comparisons of jejunum vs ileum and jejunum vs proximal cecum (Fig 6D). Additionally, the tissue comparisons jejunum vs ileum and ileum vs proximal cecum share the five differentially expressed genes *MT-ND2, ND1, ND4, ND6,* and *SNORA73*.

1.4 Discussion

Side by side comparisons of transcriptomes were made for some of the immune cells and tissues, as well as intestinal tissues, in order to gain additional biological insight. B cells were compared to monocytes from peripheral blood, B cells with bursa, bursa with thymus. The 10 most significantly (p-value) differentially expressed genes were highlighted in results (Fig 5).Among the top ten most significantly different in the monocytes compared to B cells, were *CSF1R*, *GSTA3, LY86, S100A6, TGFBO*, and *VCAN*. Colony-stimulating factor-1 receptor (*CSF1R*) is a major stimulator of macrophage maturation from monocytes (Peng et al., 2020;Wu et al., 2020). Glutathione S-transferase α3 (*GSTA3*), for glutathione metabolism, is expressed in the macrophages against reactive oxygen species (ROSs) (McNeill et al., 2015). After phagocytosis of antigen or dead cells, macrophages release ROSs to destroy the ingested molecules through respiratory burst. Therefore, it is logical that monocytes have higher expression of *GSTA3* to control over-production of ROSs. Lymphocyte antigen 86 (*LY86*), also known as Myeloid Differentiating Protein-1 (*MD1*), activates toll-like receptors in innate immune cells (Candel et al., 2015). *S100A6* (calcyclin) has been implicated in cell differentiation and apoptosis (Donato et al., 2017). Transforming growth factor-β1 (*TGFβ1*) is produced by monocytes to regulate chemotaxis (McCartney-Francis et al., 1990;Sato et al., 2000). Versican (*VCAN*) is a chondroitin sulfate proteoglycan involved in cell proliferation (Zhang et al., 1998) and is produced by leukocytes to regulate inflammation (Wight et al., 2014). Due to the constant flux in monocyte development in the peripheral blood, it explains the higher expressions of *Ly86, S100A6, TGFβ1*, and *VCAN* in the bursa.

DENND5B, HVCN1, and *IKZF3*, and *POU2AF1* expression was significantly upregulated in the B cells compared to monocytes. The role of DENN Domain Containing 5B (*DENND5B*) on B cells is unclear. B cell antigen receptor (*BCR*) signaling requires the internalization of BCR with Hydrogen Voltage-Gated Channel 1 (*HVCN1*) to regulate ROS production (Capasso et al., 2010). The Ikaros Family of Zinc-finger Protein-3 (*IKZF3*) is involved in early B cell development and its expression in increased progressively throughout B cell development (Ferreirós-Vidal et al., 2013). The POU Class 2 Homeobox Associating Factor 1 (*POU2AF1*) promotes B cell development and maturation (Zhao et al., 2008).

APOA1, PTPRF, and *RARRES1* had higher expression in the B cells compared to Bursa in this study. The bursa of Fabricius is an unique organ near the cloaca of the birds for B cell development and production (Glick et al., 1956). *APOA1*, short for Apolipoprotein A-1, is a major component in high-density lipoprotein (HDL) for lipid transport in the plasma. Interestingly, *APOA1* was one of the most abundant proteins identified in the bursa in early embryonic development (Korte et al., 2013). However, the bursas sampled for this study were more mature, which might explain that the gene expression was lower. It is not clear what the role of *APOA1* in B cells might be. *PTPRF*, short for Protein Tyrosine Phosphatase Receptor Type F, regulates Wnt signaling, which mediates B cell differentiation (Qiang et al., 2003; Gan et al., 2020). RARRES1, short for Retinoic Acid Receptor Responder 1, also known as Tazarotene-induced gene 1 protein/RAR-responsive protein TIG1, facilitates retinoic acid synthesis from β-carotene (precursor of vitamin A) (Chung and Lo, 2007; Mihály et al., 2011). Vitamin A and retinoic acid are essential for B cell development and antibody production (Ross et al., 2011), as well as monocyte differentiation into macrophages (Gundra et al., 2017).

BHLHE41, short for Basic Helix-Loop-Helix Family Member E41, is a regulator of B cell development, which is consistent with our data showing that BHLHE41 is more expressed in bursa than more mature peripheral B cells (Kreslavsky et al., 2017). The Cytohesin 1 interacting protein (*CYTIP*) regulates lymphocyte cell adhesion (Boehm et al., 2003), an important function of B cells. cFos is involved in immune receptor interaction (Bush and Bishop, 2008) The transcription factor, *NR4A2*, limits B cell activation when the secondary T cell signaling is absent (Tan et al., 2020).

We identified several non-coding RNAs with higher expression in the bursa than B cells, particularly *Metazoa_SRP, SCARNA13*, and *U3*. *Metazoa_SRP* encodes for a signal recognition particle RNA that is predominantly studied in archaea, bacteria, fungi, and protozoa species (Rosenblad et al., 2004;Dumesic et al., 2015). Little is known about the *Metazoa_SRP* gene in animals but it is thought to be involved in the translocation of RNA between the endoplasmic reticulo-membrane and cytosol (Shan and Walter, 2005) and post-translational transport of proteins to the ER (Abell et al., 2004). *SCARNA-13*, short for small Cajal body-specific RNAs, are regulatory RNA that regulates gene expressions in the Cajal bodies by controlling small nucleolar RNA such as the *U3* (Richard et al., 2003;Allantaz et al., 2012).

In the comparison of the bursa and thymus DEGs, all of the DEGs had higher expression in the thymus. Particularly, *CD247* (Lundholm et al., 2010), *CD28* (Lenschow et al., 1996), *CD3E* (Call et al., 2002), *CD4* (Zhu et al., 2009), *DNTT* (Su et al., 2004; Su et al., 2005), *LCK* (Van Laethem et al., 2013), *LEF1* (Xing et al., 2019), *RAG1* (Xing et al., 2019), *TRAT1* (Mijušković et al., 2015) are all essential genes for T cell and thymic development, while the *BCL11B* transcription factor is involved in both B and T cell development (Avram and Califano, 2014).

In summary, many of these genes were mostly involved in cellular processes relating to differentiation and cell metabolism as well as basic functions of immune cells such as cell adhesion and signal transduction. This was to be expected, as there was no explicit immunological stimulus involved, the transcriptome rather represents the baseline activity at the time sampled. Nevertheless, it was notable that all top ten DEGs in the comparison between bursa and thymus for example were all upregulated in the thymus and related to T cell differentiation and maturation. Genes differentially regulated in B cell vs Bursa are mostly involved in B cell development and differentiation. Genes differentially regulated in B cells and monocytes are involved in specific functions of the cell types.

While the chicken ileum was previously profiled (Kuo et al., 2017),the jejunum and cecum were not studied previously. We included the top ten genes of the differential expression analyses between tissue types based on levels of significance. Hierarchical clustering showed clear discrimination between the different parts of the intestine (Fig 6A). Of the 3,903 DEGs of the jejunal and ileal cells, the number of genes involved in steroid metabolism are the most different between jejunal and ileal tissues. Lipid metabolism of fat in the diet requires steroid biosynthesis of molecules such as bile acid from the pancreas into the small intestine {Dawson, 2015 #71}(Dawson and Karpen, 2015). The bile acid emulsifies lipid molecules, which travels through the small intestine and allow fatty acids to be absorbed. Consistent with the observation in rats, absorption of steroidal hormones decreases throughout the small intestine {Nakayama, 1999 #72}(Nakayama et al., 1999). Unsurprisingly, bile acid absorption can be twice as high in the jejunum than in the ileum (Krag and Phillips, 1974; Aldini et al., 1996). This further confirms the higher lipid metabolism of the jejunum than the ileum in chickens (Tancharoenrat et al., 2014). Of the 2,306 DEGs of the jejunum and cecum, the number of genes involved in oxidation-reduction process, lipid metabolic process, and cell adhesion were the most different. The primary role of the jejunum is the digestion and absorption of nutrients. In contrast, the ceca are blind-sacs in the chicken intestine that play multiple roles in nutrients absorption including bacterial fermentation of small molecules and biosynthesis of short-chain fatty acids (propionic and butyric acids) (Clench, 1999). The proximal cecum contains the cecal tonsils, which are the largest gut-associated lymphoid tissues (GALT) in chickens that demonstrate protective immune responses in the intestinal tract (Heidari et al., 2015). Therefore, it is logical that the DEGs of these metabolic functions are more pronounced in the jejunum than in the cecum. Of the 4270 DEGs between the ileum and cecum, the DEGs corresponding to cellular respiration were the most different. This could be expected as bacteria fermentation produces high levels of short-chain fatty acids in the cecum, which can be used as energy by the intestinal cells (Murugesan et al., 2014). Due to the relative size and metabolic demands of the ileum compared to the cecum, much energy is needed from aerobic respiration and mitochondrial electron transport to produce adequate energy in the ileum.

Interestingly, in the comparisons of the jejunum to the ileum and to the proximal cecum revealed differential expression of *LCT*, the gene encoded for lactase production, which is lower in the jejunum compared to that in the ileum and proximal cecum (Fig 6D). Since chickens are not mammals, the expression of the lactase gene is perplexing. The expression of the lactase gene in chickens has been debated in the past (Hamilton and Mitchell, 1924). Several hypotheses had been proposed about the presence of the lactase gene in chickens. The presence of the lactase gene could be due to 1) bacterial fermentation of lactase in the intestine, 2) evolutionary artifacts, or 3) improper annotation of the gene in chickens that could have the same sequence but functionally different in the chicken compared to mammals. An early study using based on the disappearance of lactase *in vitro* showed that lactase was assimilated in the crop but not in the proventriculus or the intestine (Plimmer and Rosedale, 1922). However, the assumption of the disappearance of lactase as evident of lactose digestion is flawed because it does not account for the microbial degradation of lactose. Later, molecular cloning confirmed lactase expression in the chicken intestinal tract as well as in mussel (Freund et al., 1997). Based on our sequencing results, we cannot conclude whether this is due to evolutionary artifacts, evolutionary converged traits with separate lineages, or genes with the same sequence but with completely different functions, or otherwise.

Among the other DEGs from the intestinal tract, the majority of the genes had higher expression in the cranial than the caudal part of the intestine (i.e., the ileum had higher expression than jejunum, etc.), except *SNORA73, U3*, and *SCARNA13*. Similar to the analysis of the immune organs, *APOA4* was among the top DEGs in the comparisons between jejunum and ileum as well as jejunum and proximal cecum. In the intestine, *APOA4* is responsible for lipid metabolism (Tso et al., 2004;Wang et al., 2020) and lipid-soluble vitamin metabolism such as retinoic acid (vitamin A) (Hebiguchi et al., 2015). Coincidentally, both *APOA4* and retinoic acid-binding protein-2 (RBP2) had higher expressions in the ileum compared to the jejunum. Mitochondrial NADH dehydrogenase (*MT-ND2*), NADH dehydrogenase-1 (*ND1*), NADH dehydrogenase-4 (*ND4*), NADH dehydrogenase-5 (*ND5*), and NADH dehydrogenase-6 (*ND6*) relate to the electron transport chain that generates cellular energy in the form of ATP through oxidative respiration (Weiss et al., 1991). These energy metabolic genes had higher expression in the ileum compared to the jejunum and in the proximal cecum compared to the ileum, suggesting higher energy production through aerobic respiration in these tissues. The Transmembrane Serine Protease 15 (*TMPRSS15*) is an enteropeptidase secreted from the pancreas that catabolizes trypsinogen to trypsin and chymotrypsinogen to procarboxypeptidase for protein digestion in the intestine (Zheng et al., 2009). Expression of *TMPRSS15* was higher in the ileum than the jejunum, suggesting the increasing rate of protein digestion throughout the small intestine.

Most of the top DEGs in proximal cecum comparisons had higher expression in the proximal cecum. Several of these genes are involved in lipid metabolism, including *APOA4* and *APOB* (major components of lipoproteins) (Schianca et al., 2011), Beta-carotene Oxygenase-1

(*BCO1*) (lipid-soluble vitamin A metabolism), *CUBN* (lipoprotein endocytosis) (Christensen and Birn, 2002), and *SLC26A9* (bile metabolism) (Li et al., 2016). Adenosine deaminase (*ADA*) is involved in purine metabolism for nucleotide synthesis (Ikehara and Fukui, 1974). Bacteria in the intestine are essential for vitamin absorption for the host (LeBlanc et al., 2013). Therefore, it is consistent that *TM4SF4*, which is involved in thiamine (vitamin B1) metabolism, had higher expression in the proximal cecum. The DEAD/DEAD-Box Helicase-60 is involved in innate immunity (Perčulija and Ouyang, 2019). The mucosal-associated lymphoid tissue (MALT) in the proximal cecum is the secondary lymphoid organ of the chicken, which could explain the higher expression of the DEAD/DEAD-Box Helicase-60 (*DDX60*) in the cecum.

In summary, we were able to correlate most of the differential expressed genes in the intestine to mostly metabolic processes related to nutrient digestion and absorption. Several genes in the distal part of the intestine were particularly implicated in vitamin metabolism. This was not surprising because vitamin metabolism requires the microbiota which are more abundant in the distal intestines. Genes involved in energy metabolism are also abundant in the cecum, which suggests that microbial contribution of energy production in the intestine is especially important.

In the current study, whole transcriptome RNA-seq of immune, intestinal, and reproductive cells and tissues were sequenced. The current Ensembl chicken annotation (GRCg6a, genome-build-accession NCBI:GCA_000002315.5), contains 16,779 protein coding genes, 7,577 non-coding genes, and 39,288 gene transcripts. Of the non-coding genes, 5,504 were long non-coding genes. From 10 diverse cell and tissue types, we recovered 73% of annotated genes and 63% of known transcripts. Of annotated genes, 90% of coding genes are expressed in the 10 cell and tissue types studied here, while only 44% of annotated lncRNAs are expressed. The potential regulatory role of lncRNAs may explain the limited expression, and suggest a more cell- or tissue-specific role. We found that biosamples often expressed hundreds of cell- or tissue-specific coding genes and lncRNAs. While many genes are commonly expressed in multiple samples, we also determined that over 500 isoforms of genes are uniquely expressed. Surprisingly, each cell and tissue type only expressed an average of 1-2 gene isoforms; however, each biosample type had at least one gene with 11 or more isoforms expressed in the cell or tissue type. Analysis of differentially expressed genes revealed biological processes that are consistent with a function in the cells or tissues of interest. Continued investigation of these genes should further our understanding of disease susceptibility/resistance, feed conversion and egg production. Collectively, these data provide a deeper understanding of the chicken transcriptome in a cell- and tissue-specific manner. Additional samples from the FAANG and greater community will continue to advance efforts towards a comprehensive catalogue of the chicken transcriptome.

1.5 Materials and Methods

Experimental Animals

The animal procedure was approved and conducted according to guidelines established by the Western University of Health Sciences, Pomona, California (WesternU) Institutional Animal Care and Use Committee. The F1 crosses of Line 6 and Line 7 from the Avian Disease and Oncology Laboratory (ADOL) were used in this study (Stone, 1975; Briles et al., 1977; Bacon et al., 2000). The two lines have identical major histocompatibility complex (MHC) B*2 haplotype, but present different disease susceptibility to Marek's Disease Virus (Line 6_3: MDV-resistant and Line 7_2: MDV-susceptible) (Liu et al., 2001). The F1 crosses of these lines have been used in other annotation studies by the FAANG consortium; therefore, it is used in this study to allow for better comparison of the data. The chickens were held in open cages in the vivarium of the University Research Center at Western University. In addition to daily health monitoring, fresh food and water were provided *ad libitum*. Room temperature was adjusted to and maintained at 32°C until 3 weeks of age. To minimize the risk of pecking disorders, chicks were kept under restricted lighting conditions throughout the study. Peripheral blood was collected from jugular or wing web veins. Experimental animals were euthanized by insufflation of isoflurane.

Sample Collection

All assays were performed in at least duplicates. Immune tissue (thymus and bursa), intestinal tissues (jejunum, ileum, and proximal cecum), and reproductive tissue (ovary) were collected and flash-frozen in liquid nitrogen for later use. Tissue immune cells (lung macrophage, and spleen CD3+ T cells) were collected from the organs homogenized and filtered through 70 μm nylon cell strainers.

Tissue macrophages and T cells were extracted using magnetic beads (Dynabeads FlowComp Flexi, Invitrogen, Carlsbad, CA) coated with biotinylated-mouse-anti-chicken-monocyte/macrophage-monoclonal antibodies (Clone KUL-1, Cat. No. 8420-08, SouthernBiotech) and biotinylated-mouse-anti-chicken-CD3-monoclonal antibodies (Clone AV-20, Cat. No. 8200-08, SouthernBiotech, Birmingham, AL), respectively.

Peripheral blood B cells were collected from the blood (Collisson et al., 2017). Briefly, the blood was diluted in an equal volume of PBS and layered slowly over Ficoll-Histopaque (1.083 g/mL) (Sigma-Aldrich, St. Louis, MO), and then centrifuged for 35 minutes (400 x g at 23°C with brake off). The interface containing the peripheral blood mononuclear cells (PBMCs) and B cells were collected. Peripheral blood B cells were extracted using magnetic beads (Dynabeads Pan Mouse IgG, Invitrogen, Waltham, MA) coated with

unlabeled-mouse-anti-chicken-Bu-1a/b-monoclonal antibodies (Clone AV-20, Cat. No. MCA5764, BioRad, Hercules, CA).

Peripheral blood monocytes were collected as described previously (Dawes et al., 2014). Briefly, after density gradient separation using Ficoll-Histopaque as described above, the PBMCs were washed twice in PBS, counted, and then plated in cell culture plate containing RPMI with 10% heat-inactivated fetal bovine serum; nonessential amino acids (0.1 mM/mL; Invitrogen), L-glutamine (2 mM; Sigma-Aldrich), penicillin (50 U/mL; Invitrogen), and Streptomycin (50 µg/mL; Invitrogen). The cells were incubated (37°C, 5% CO_2) for 3 hours. Non-adherent cells were removed with warm PBS. Adherent peripheral blood monocytes were collected by adding Trizol directly to cells for RNA extraction immediately.

RNA Extraction and Library Construction

Total RNA from tissues and immune cells was collected using a modified Trizol/Chloroform method. Briefly, a second chloroform phase extraction and a second ethanol wash were included in the modified method. The total RNA from tissues was purified, and DNase treated using the Direct-zol RNA Miniprep Plus (Zymo Research, Irvine, CA). The total RNA from immune cells (lung macrophage, spleen T cells, peripheral blood B cells, and peripheral blood monocytes) were not purified using the Direct-zol RNA miniprep Plus due to the lower concentration of the immune cell RNA compare to the tissue RNA. The total RNA from immune cells was DNase treated after extraction. Quality control of the total RNA was performed fluorometrically using the Qubit RNA HS Assay Kit and Qubit 3 (ThermoFisher Scientific, Waltham, MA) and RNA 6000 Nano Kit and Bioanalyzer 2100 (Agilent, Santa Clara, CA). Total RNA with RNA integrity number (RIN) above 8.0 were used in the library generation process using the Zymo-Seq RiboFree Total RNA Library Kit (Zymo Research). ERCC RNA Spike-In Controls (Invitrogen) were used to create a standard baseline measurement of RNA. Ribosomal-RNA (rRNA), globin, and overrepresented transcripts were removed, and sequencing adaptor ligation of the cDNA was removed by size selection and PCR enrichment. Libraries were barcoded with P5 and P7 index sequences according to the manufacturer's protocol.

RNA-Sequencing

Libraries were pooled and sequenced on HiseqX-PE150 by Novogene Bioinformatics Technology Co (Beijing, China). Libraries were sequenced to an average depth of 43.7 million paired reads per library.

Bioinformatics Analyses of RNA-sequencing Data

Raw reads were trimmed with TrimGalore (v0.4.1){Martin, 2011 #96}(Martin, 2011). Trimmed reads were mapped and quantified using STAR (v2.6.1c) and RSEM (v1.3.1) using the function rsem-calculate-expression and the reference file GRCg6a, genome-build-accession NCBI:GCA_000002315.5 (Li and Dewey, 2011). Transcriptomes were assembled using StringTie (v2.1.4) and gffCompare (v0.11.6) (Pertea and Pertea, 2020). Counts of genes and transcripts from Figs 2, 3B-C were obtained from the output of gffCompare.

Euclidean distance, pairwise correlations, and PCA plots were generated by pcaExplorer (Marini and Binder, 2019). PCA was performed using the top 500 most variable genes, used the gene counts from the RSEM quantification, and the gene counts were first normalized with DESeq2 (v1.30.0) (Love et al., 2014). Heatmaps were generated with Morpheus (https://software.broadinstitute.org/morpheus) (Gould, 2016). Shannon's entropy calculations were performed with the BioQC function entropyDiversity (Zhang et al., 2017a). Count matrices inputted to BioQC were normalized with DESeq2 and only genes and isoforms with an average TPM greater than 0.5 across replicates from the RSEM quantified counts were included. Isoform entropy had an additional filter, requiring that the isoform's gene be expressed in at least two cell types. Browser shots were generated using the UCSC genome browser (Kent et al., 2002). BigWig files for the UCSC genome browser were generated from the mapped bam files using deepTools bamCoverage (v3.5.0) (Ramírez et al., 2014).

Differential gene expression was calculated using DESeq2 (v1.30.0) (Love et al., 2014b). Genes with an adjusted p-value less than 0.05 were considered differentially expressed. GO biological processes were calculated using WebGestalt (Liao et al., 2019) with an FDR threshold of 0.05 for determining GO category overrepresentation. The exception for this is the GO analysis in Sup Fig 3B, which lists the top 10 enriched GO terms. WebGestalt was run with the basic parameters "Gallus gallus", "Over-Representation Analysis (ORA)", "geneontology and "Biological Process". "genome" was selected as the reference set. Figs 5C, 6C, and 3B directly use these GO terms. Figs 5B and 6B display the weighted set cover, which reduces redundancy of the categories displayed. Full GO categories corresponding to the weighted set covers are provided in Supplemental Tables 4 and 5. Venn diagrams were generated with Intervene (Khan and Mathelier, 2017).

1.6 Figures

Figure 1

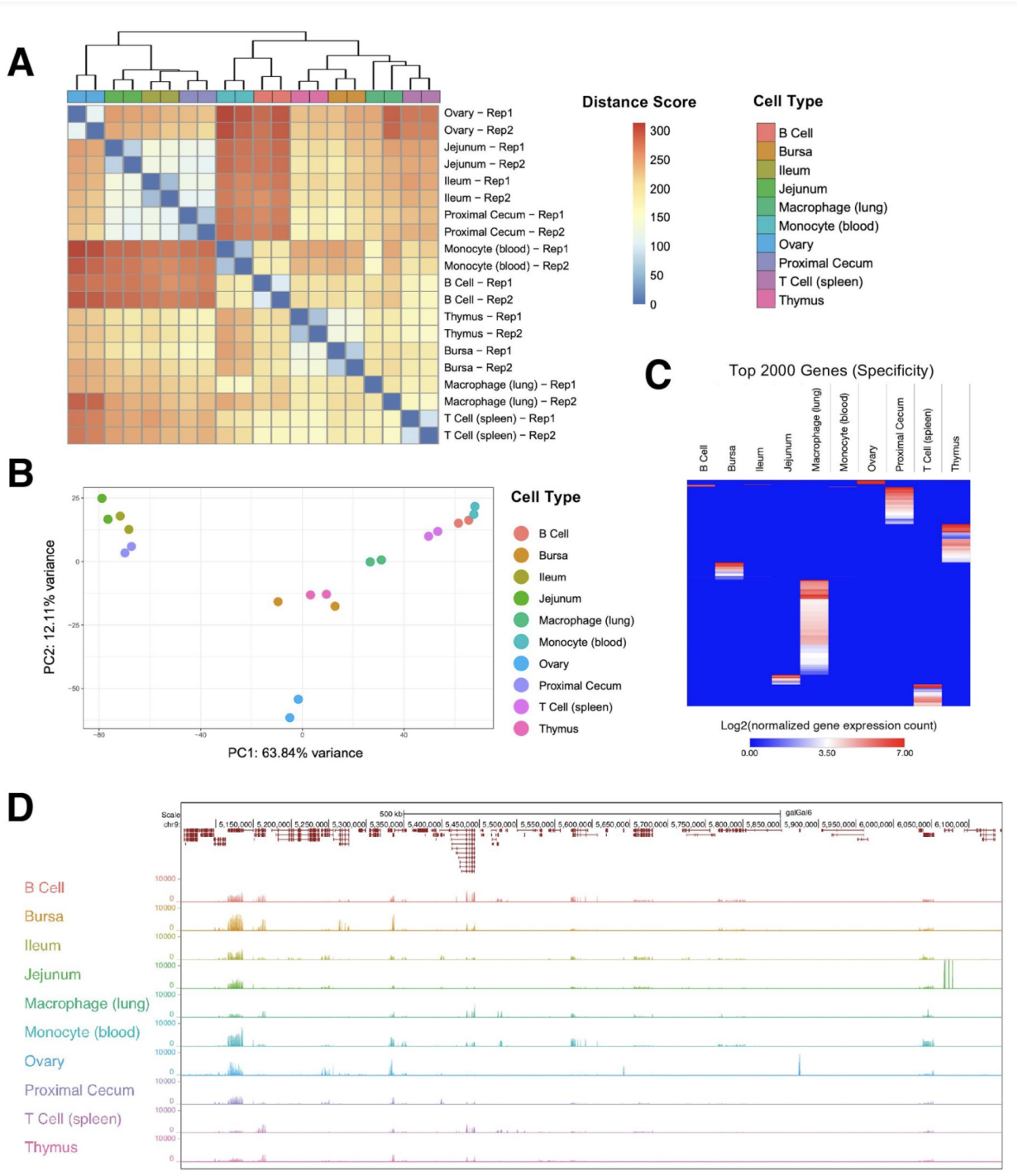


Figure 1: Overview of Tissue RNA-Sequencing Results. **(A)** Sample to sample distance heatmap quantifying the Euclidean distance between each sample. **(B)** Principal component analysis using the top 500 most variable genes for all samples. **(C)** Expression of the 2000 genes with the highest Shannon's entropy values. Rows sorted using Euclidean distance. **(D)** UCSC browser shot of RNA-seq data showing variable expression among samples.

Figure 2

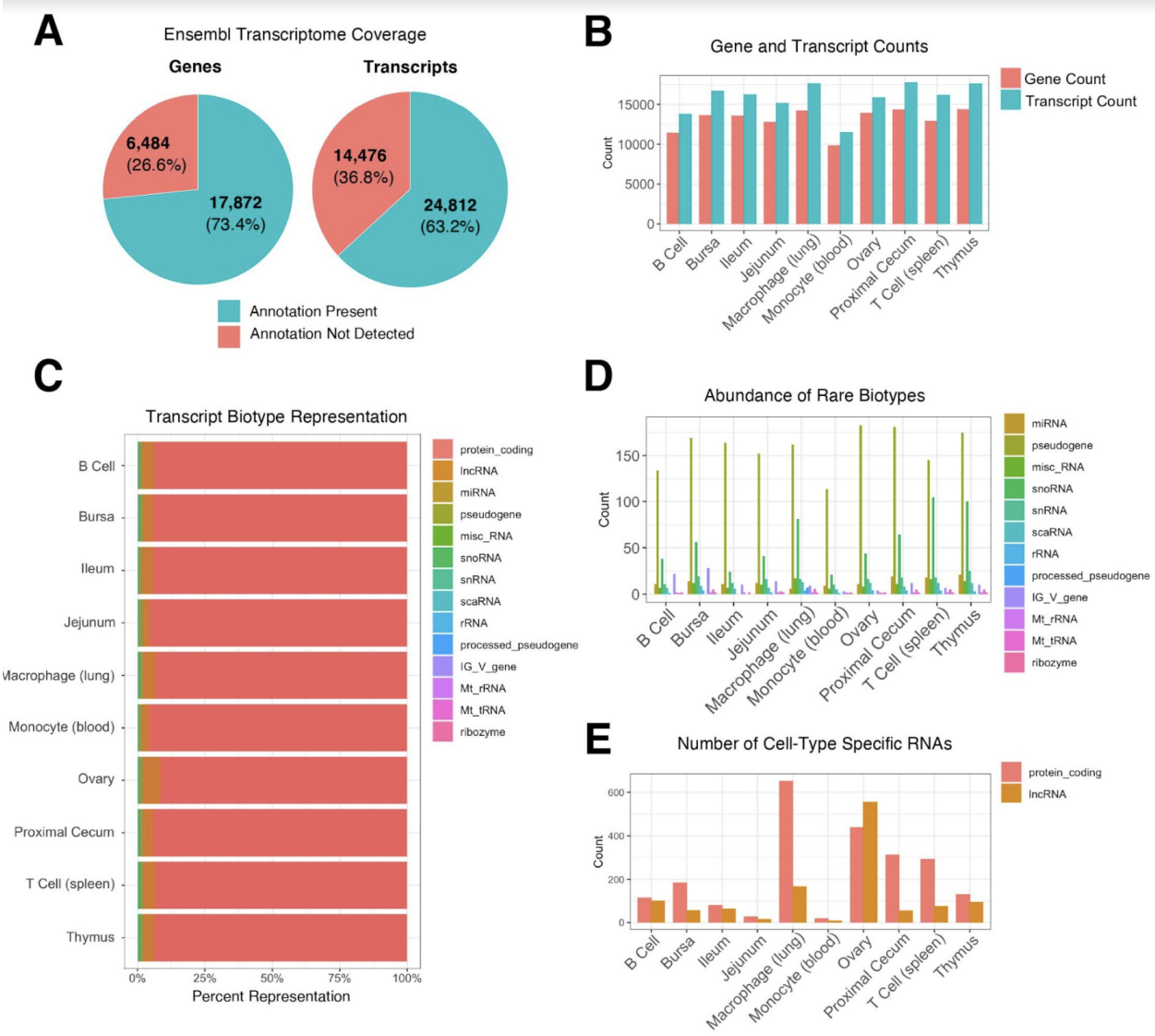


Figure 2: Gene and Transcript Characterization. (A) Percentage of annotated chicken genes and transcripts detected across all samples. There are 24,356 genes and 39,288 transcripts in the *Gallus gallus* GRCg6a Ensembl annotation. **(B)** Number of genes and transcripts detected per cell type. Gene counts range between 9,839 [Monocyte (blood)] – 14,418 [Thymus]. Transcript counts range between 11,522 [Monocyte (blood)] – 17,794 [Proximal Cecum]. **(C)** Breakdown of transcript types detected per cell type, by percentage. **(D)** Counts of low abundance transcript

biotypes with less than 3% representation (all transcript biotypes, except lncRNA and protein coding RNA). **(E)** Number of protein coding RNAs and lncRNAs unique to each sample type.

Figure 3

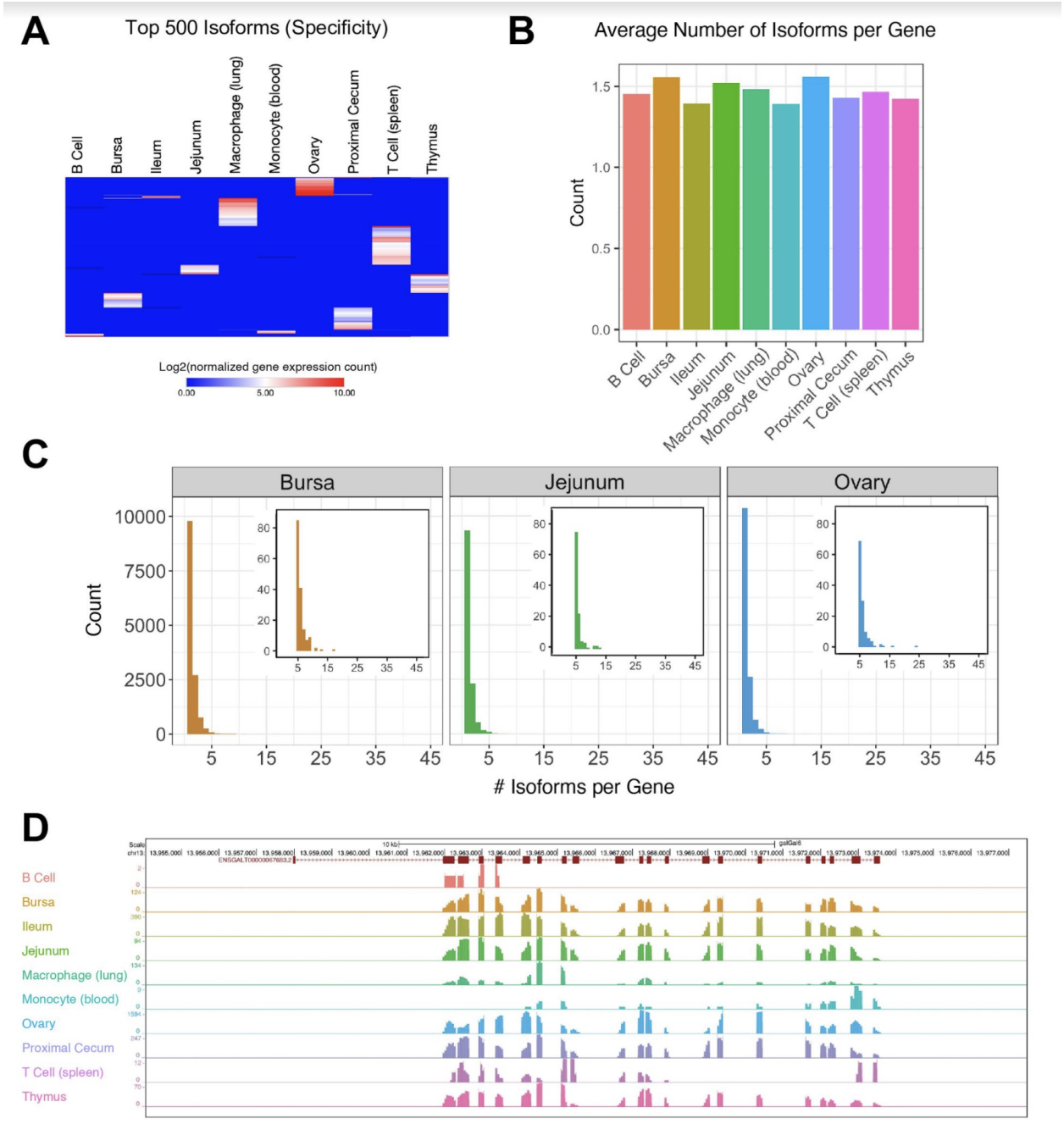


Figure 3: Isoform Characterization. **(A)** Average number of isoforms per gene for each cell type. **(B)** Expression of the 500 isoforms with the highest Shannon's entropy values. Rows sorted using Euclidean distance. Isoforms have been filtered for genes that have a TPM of at least 0.5 in at least two cell types. **(C)** Histogram of isoform counts per gene. The cutout plot in the upper-right corner is a zoomed-in section for 5+ isoforms per gene. **(D)** Browser shot of the single annotated isoform for the gene *PDGFRB* (ENSGAL00000030613) emphasizing splice isoform differences between cell types.

Figure 4

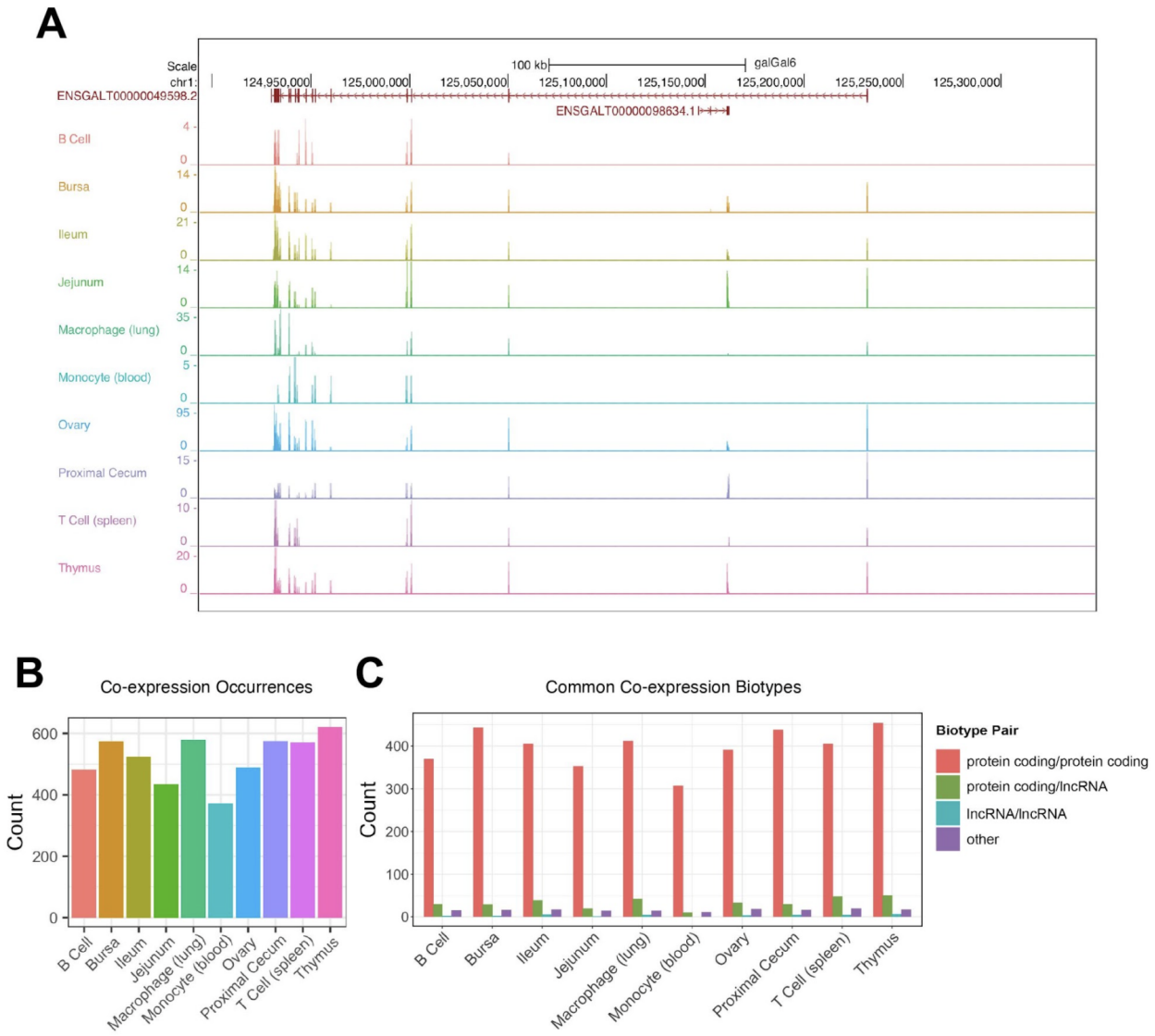


Figure 4: Forward-Reverse Strand Co-expression. **(A)** Example of co-expressed transcripts on the forward and reverse strands. ENSGALT00000049598/*FRMPD4* (reverse strand) and ENSGALG00000098634 (forward strand) are overlapping in their genomic coordinates. **(B)** Number of co-expression occurrences in each tissue type. **(C)** Co-expression counts of protein coding RNAs and lncRNAs.

Figure 5

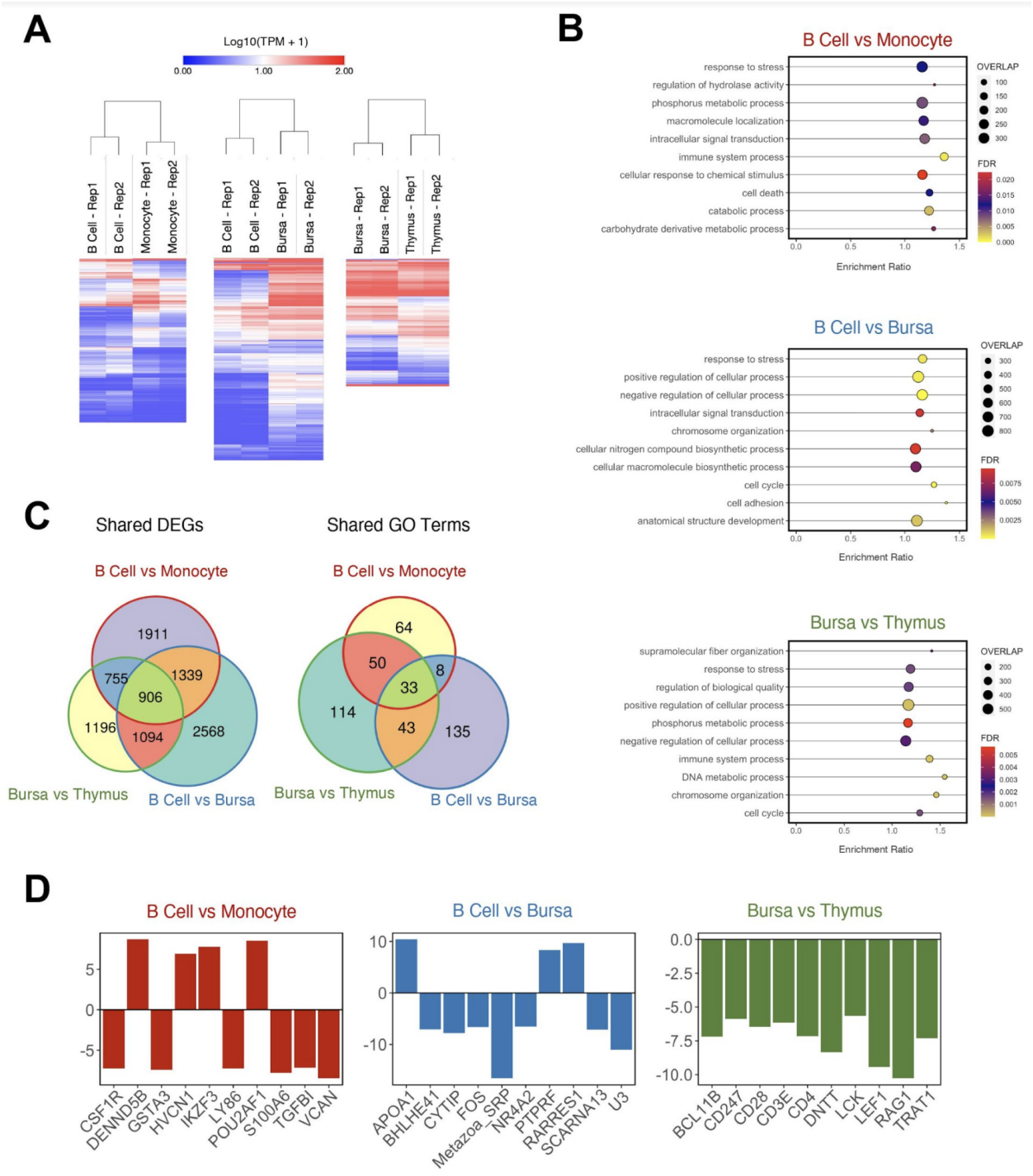


Figure 5: Differentially Expressed Gene (DEG) Analysis on Immune Samples. (A) Heatmaps of DEGs in three cell type comparisons: B Cells vs Monocytes (4911 DEGs), B Cells vs Bursa Tissue (5907 DEGs), Bursa Tissue vs Thymus Tissue (3951 DEGs). Samples were clustered both

by column and by row using Euclidean distance based on log transformed TPM value. **(B)** Enriched GO Biological Process categories for each sample using weighted set cover filtering in WebGestalt. “Overlap” quantifies the number of DEGs present in that GO set. **(C)** The numbers of DEGs overlapping between sample comparisons. **(D)** Log2 fold-change of top 10 DEGs for each sample comparison.

Figure 6

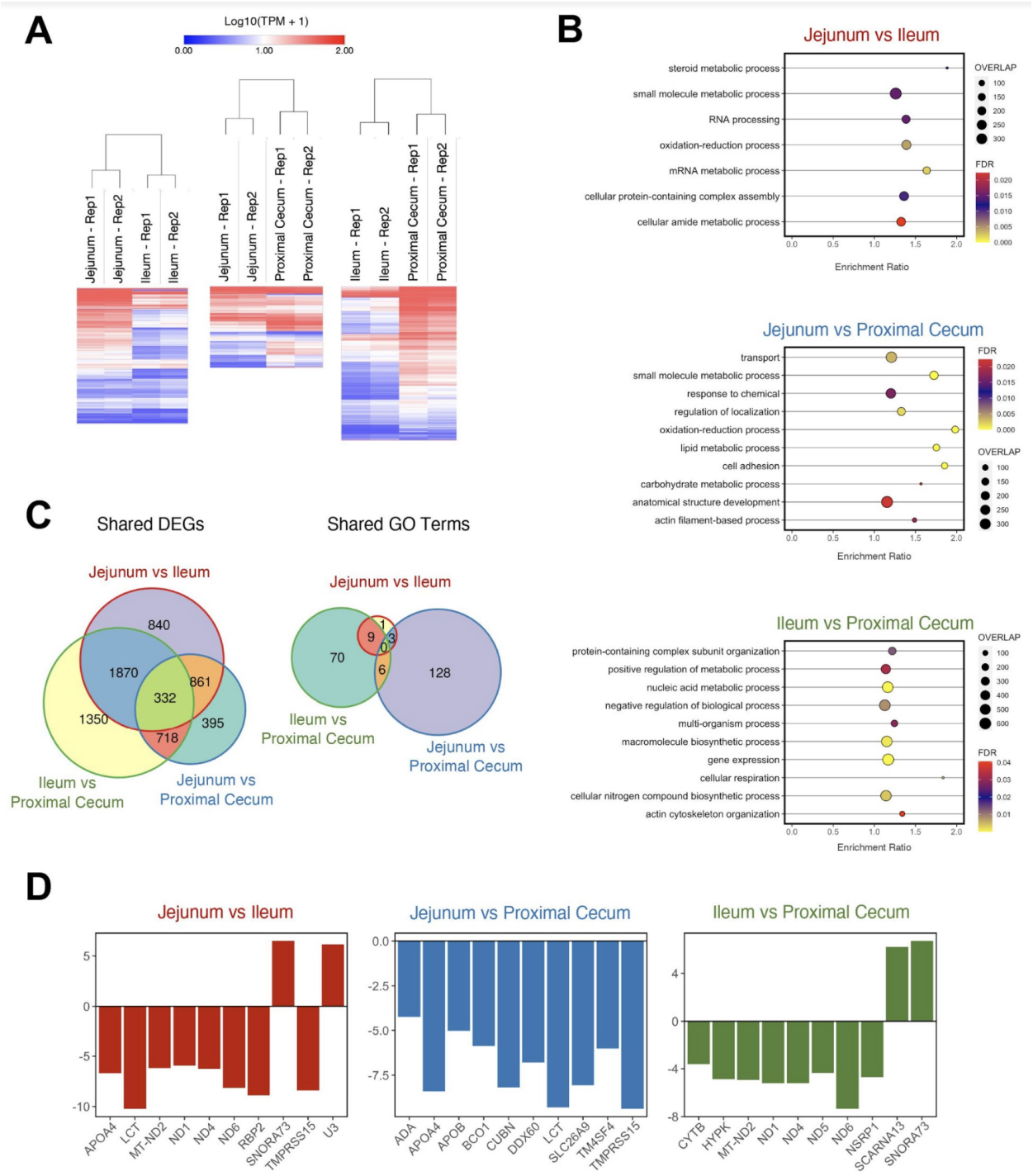
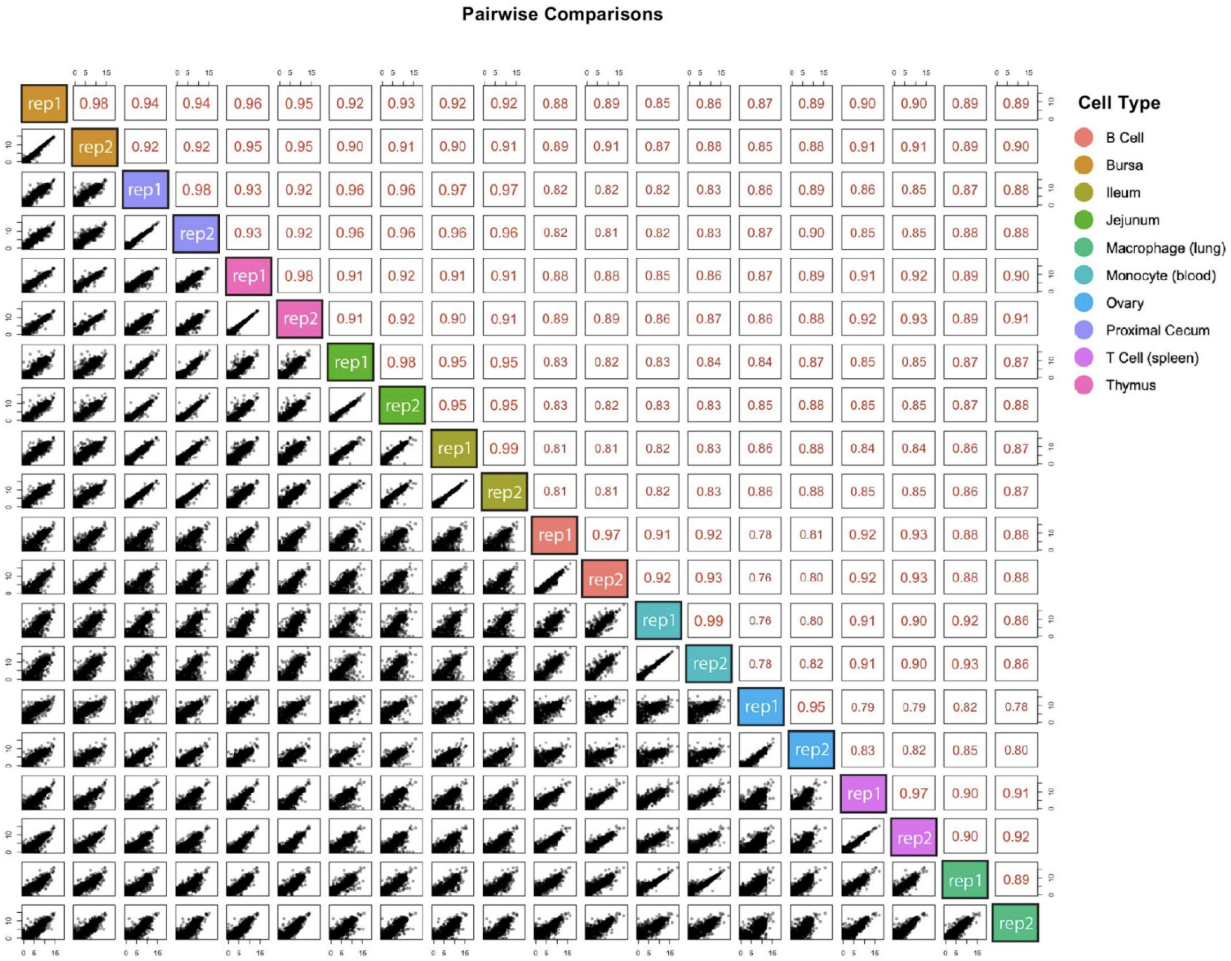


Figure 6: Differentially Expressed Gene (DEG) Analysis on Intestinal Samples. (A) Heatmaps of DEGs in three cell type comparisons: Jejunum vs Ileum (3903 DEGs), Jejunum vs Proximal Cecum (2306 DEGs), Ileum vs Proximal Cecum (4270 DEGs). Samples were clustered both by column and by row using Euclidean distance based on log transformed TPM value. **(B)** Enriched GO Biological Process categories for each sample using weighted set cover filtering in

WebGestalt. “Overlap” quantifies the number of DEGs present in that GO set. **(C)** The numbers of DEGs overlapping between sample comparisons. **(D)** Log2 fold-change of top 10 DEGs for each sample comparison.

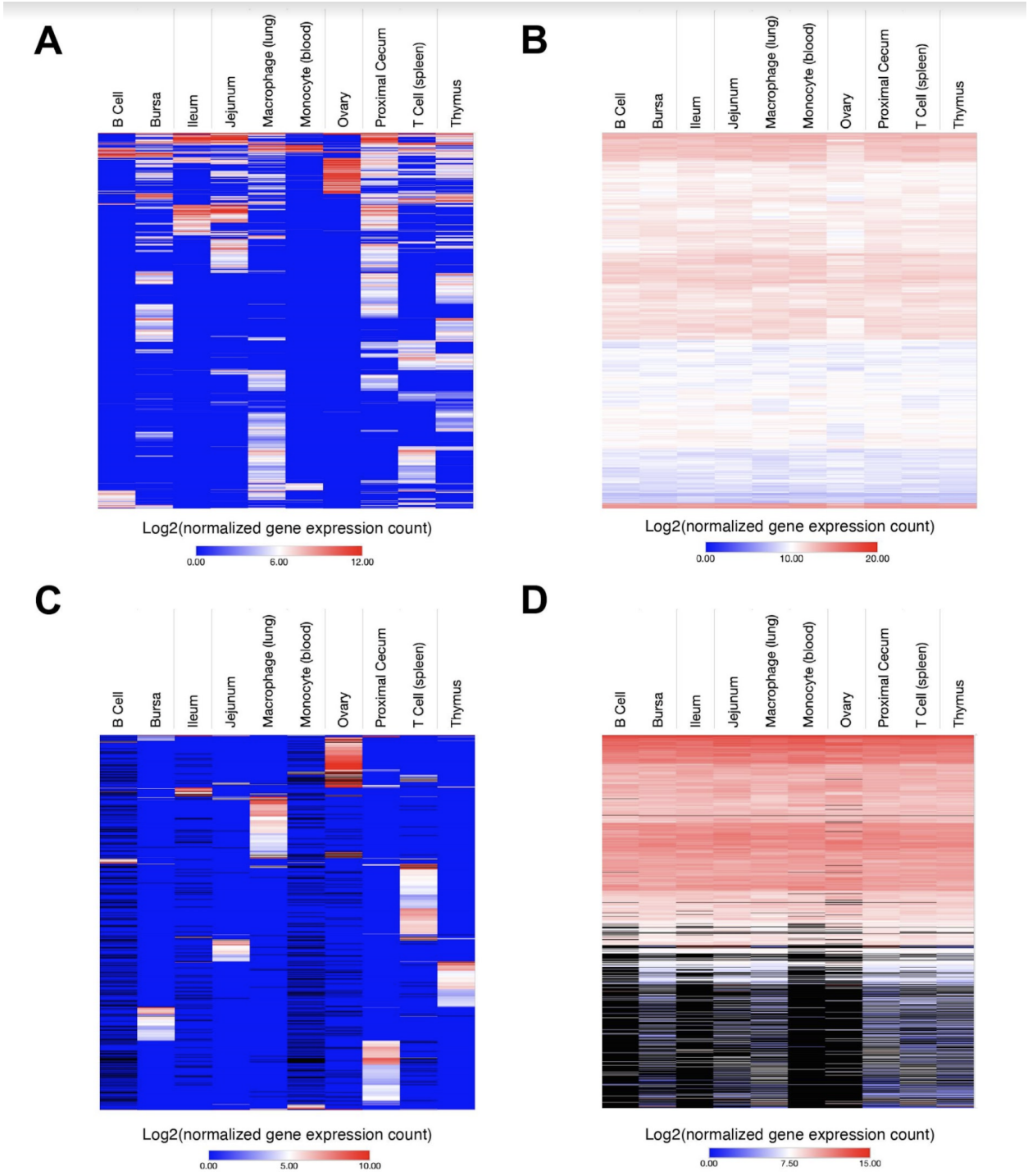
1.7 Supplemental Figures

Supplemental Figure 1



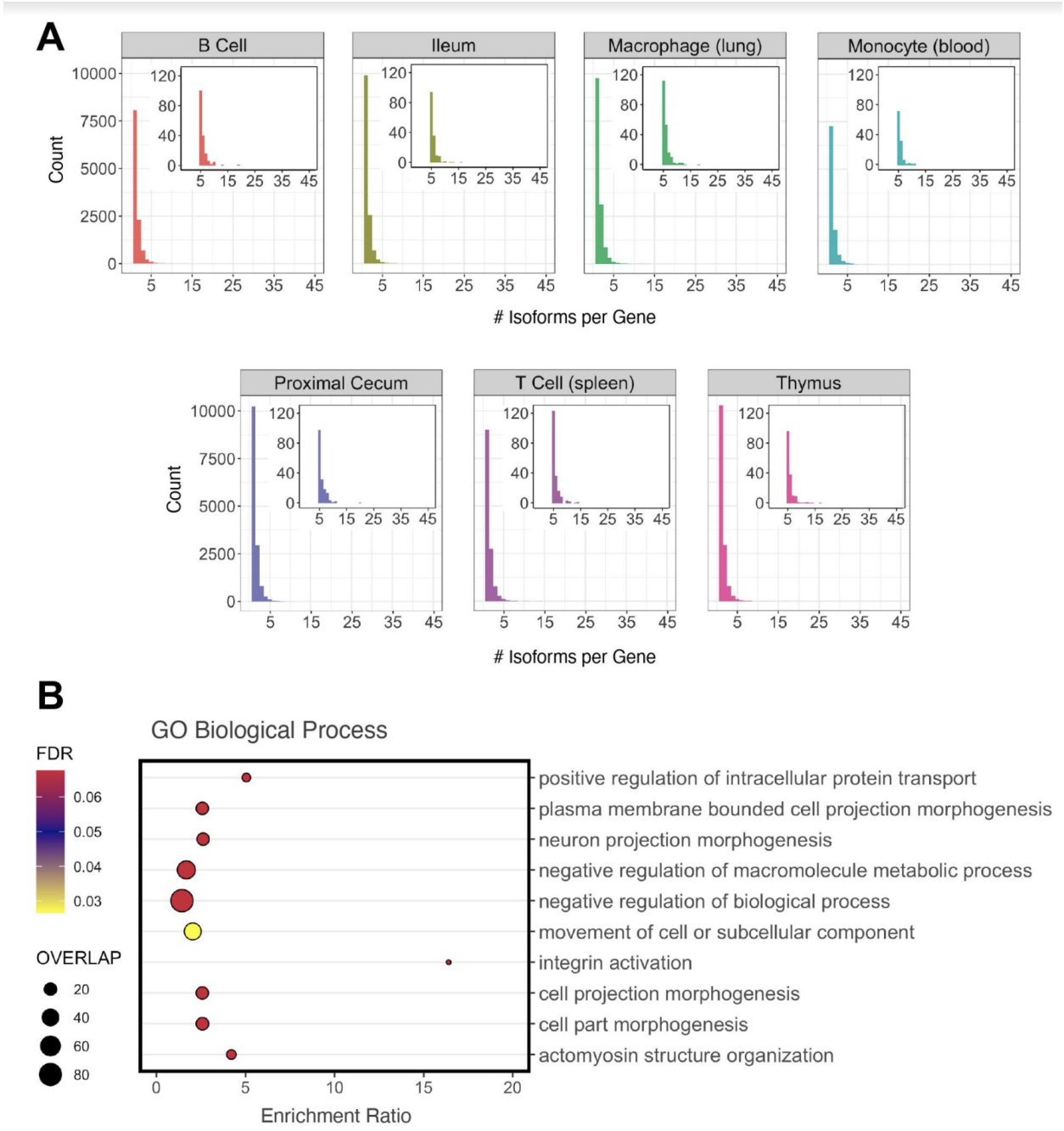
Supplemental Figure 1: Pairwise Correlations. Pearson correlation plot for each pairwise comparison generated with pcaExplorer. Plots use a subset of 1000 genes and uses log2 normalized gene counts for plot axes and values, respectively.

Supplemental Figure 2



Supplemental Figure 2: Gene and Isoform Expression Specificity. (A) Top 2000-4000 most specific genes. **(B)** The 1000 least specific genes. **(C)** The top 1000 most specific isoforms. **(D)** 1000 least specific isoforms. Rows sorted by Euclidean distance. **(C-D)** Isoforms have been filtered for genes that have a TPM of at least 0.5 in at least two cell types. Matrix entries that have no expression of that isoform's gene are colored black.

Supplemental Figure 3



Supplemental Figure 3: Isoform Histograms. (A) Histogram of isoform counts per gene for tissues not included in Fig 3B. The cutout plot in the upper-right corner is a zoomed-in section for 5+ isoforms per gene. **(B)** GO Biological Process analysis for genes with five or more isoforms.

1.8 Tables

Table 1.
Number of transcripts, genes, and lncRNA by tissue.

Type	Tissue	# of Transcripts	# of Genes	# of lncRNA
Immune	B cell	16,170	12,94	648
	Bursa	113,825	11,442	608
	Monocyte (blood)	17,794	14,380	316
	Macrophage (lung)	11,522	9,839	868
	T cell (spleen)	16,703	13,642	731
	Thymus	17,639	14,240	757
Intestinal	Jejunum	15,172	12.820	398
	Ileum	15,924	13,940	710
	Proximal cecum	17,619	14,418	690
Reproductive	Ovary	16,255	13,585	1,063

1.9 Author Contributions

EO performed data analysis, figure generation, and manuscript writing. TTN performed sample isolation and preparation, data interpretation and manuscript writing. PC contributed to sample isolation. LG, PS, and ST led the tissue isolation. YD and RDH conceived the study, guided data analysis, contributed to data interpretation and manuscript writing.

1.10 Data Availability

Data is publically available at GEO (GSE166257).

1.11 References

Abell, B.M., Pool, M.R., Schlenker, O., Sinning, I., and High, S. (2004). Signal recognition particle mediates post-translational targeting in eukaryotes. The EMBO Journal 23, 2755-2764.

Aldini, R., Montagnani, M., Roda, A., Hrelia, S., Biagi, P., and Roda, E. (1996). Intestinal absorption of bile acids in the rabbit: different transport rates in jejunum and ileum. *Gastroenterology* 110**,** 459-468.

Allantaz, F., Cheng, D.T., Bergauer, T., Ravindran, P., Rossier, M.F., Ebeling, M., Badi, L., Reis, B., Bitter, H., and D'asaro, M. (2012). Expression profiling of human immune cell subsets identifies miRNA-mRNA regulatory relationships correlated with cell type specific expression. *PloS One* 7**,** e29979.

Avram, D., and Califano, D. (2014). The multifaceted roles of Bcl11b in thymic and peripheral T cells: impact on immune diseases. *The Journal of Immunology* 193**,** 2059-2065.

Bacon, L., Hunt, H., and Cheng, H. (2000). A review of the development of chicken lines to resolve genes determining resistance to diseases. *Poultry Science* 79**,** 1082-1093.

Bédécarrats, G.Y., Baxter, M., and Sparling, B. (2016). An updated model to describe the neuroendocrine control of reproduction in chickens. *General and Comparative Endocrinology* 227**,** 58-63.

Boehm, T., Hofer, S., Winklehner, P., Kellersch, B., Geiger, C., Trockenbacher, A., Neyer, S., Fiegl, H., Ebner, S., and Ivarsson, L. (2003). Attenuation of cell adhesion in lymphocytes is regulated by CYTIP, a protein which mediates signal complex sequestration. *The EMBO Journal* 22**,** 1014-1024.

Briles, W.E., Stone, H.A., and Cole, R. (1977). Marek's disease: effects of B histocompatibility alloalleles in resistant and susceptible chicken lines. *Science* 195**,** 193-195.

Bush, T.J.V., and Bishop, G.A. (2008). TLR7 and CD40 cooperate in IL‐6 production via enhanced JNK and AP‐1 activation. *European journal of immunology* 38**,** 400-409.

Call, M.E., Pyrdol, J., Wiedmann, M., and Wucherpfennig, K.W. (2002). The organizing principle in the formation of the T cell receptor-CD3 complex. *Cell* 111**,** 967-979.

Candel, S., Sepulcre, M.P., Espín-Palazón, R., Tyrkalska, S.D., De Oliveira, S., Meseguer, J., and Mulero, V. (2015). Md1 and Rp105 regulate innate immunity and viral resistance in zebrafish. *Developmental & Comparative Immunology* 50**,** 155-165.

Capasso, M., Bhamrah, M.K., Henley, T., Boyd, R.S., Langlais, C., Cain, K., Dinsdale, D., Pulford, K., Khan, M., and Musset, B. (2010). HVCN1 modulates BCR signal strength via regulation of BCR-dependent generation of reactive oxygen species. *Nature Immunology* 11**,** 265.

Cheng, Y., and Burt, D.W. (2018). Chicken genomics. *International Journal of Developmental Biology* 62**,** 265-271.

Christensen, E.I., and Birn, H. (2002). Megalin and cubilin: multifunctional endocytic receptors. *Nature reviews Molecular cell biology* 3**,** 258-267.

Chung, G., and Lo, K. (2007). RARRES1 (retinoic acid receptor responder (tazarotene induced) 1). *Atlas of Genetics and Cytogenetics in Oncology and Haematology*.

Clench, M.H. (1999). The avian cecum: update and motility review. *Journal of experimental Zoology* 283**,** 441-447.

Collisson, E., Griggs, L., and Drechsler, Y. (2017). Macrophages from disease resistant B2 haplotype chickens activate T lymphocytes more effectively than macrophages from disease susceptible B19 birds. *Developmental & Comparative Immunology* 67**,** 249-256.

Cooper, M.D., Peterson, R.D., South, M.A., and Good, R.A. (1966). The functions of the thymus system and the bursa system in the chicken. *Journal of Experimental Medicine* 123**,** 75-102.

Dai, M., Feng, M., Xie, T., and Zhang, X. (2019). Long non-coding RNA and MicroRNA profiling provides comprehensive insight into non-coding RNA involved host immune responses in ALV-J-infected chicken primary macrophage. *Developmental & Comparative Immunology* 100**,** 103414.

Dawes, M.E., Griggs, L.M., Collisson, E.W., Briles, W.E., and Drechsler, Y. (2014). Dramatic differences in the response of macrophages from B2 and B19 MHC-defined haplotypes to interferon gamma and polyinosinic: polycytidylic acid stimulation. *Poultry Science* 93**,** 830-838.

Dawson, P.A., and Karpen, S.J. (2015). Intestinal transport and metabolism of bile acids. *Journal of Lipid Research* 56**,** 1085-1099.

Donato, R., Sorci, G., and Giambanco, I. (2017). S100A6 protein: functional roles. *Cellular and Molecular Life Sciences* 74**,** 2749-2760.

Dumesic, P.A., Rosenblad, M.A., Samuelsson, T., Nguyen, T., Moresco, J.J., Yates Iii, J.R., and Madhani, H.D. (2015). Noncanoncial signal recognition particle RNAs in a major eukaryotic phylum revealed by purification of SRP from the human pathogen Cryptococcus neoformans. *Nucleic Acids Research* 43**,** 9017-9027.

Ferreirós-Vidal, I., Carroll, T., Taylor, B., Terry, A., Liang, Z., Bruno, L., Dharmalingam, G., Khadayate, S., Cobb, B.S., and Smale, S.T. (2013). Genome-wide identification of Ikaros targets elucidates its contribution to mouse B-cell lineage specification and pre-B–cell differentiation. *Blood, The Journal of the American Society of Hematology* 121**,** 1769-1782.

Freund, J.-N., Jost, B., Loreztz, O., and Duluc, I. (1997). Identification of homologues of the mammalian intestinal lactase gene in non-mammals (birds and molluscs). *Biochemical Journal* 322**,** 491-498.

Gan, T., Stevens, A.T., Xiong, X., Wen, Y.-A., Farmer, T.N., Li, A.T., Stevens, P.D., Golshani, S., Weiss, H.L., and Evers, B.M. (2020). Inhibition of protein tyrosine phosphatase receptor type F suppresses Wnt signaling in colorectal cancer. *Oncogene* 39**,** 6789-6801.

Giuffra, E., Tuggle, C.K., and Consortium, F. (2019). Functional annotation of animal genomes (FAANG): current achievements and roadmap. *Annual review of animal biosciences* 7**,** 65-88.

Glick, B., Chang, T.S., and Jaap, R.G. (1956). The bursa of Fabricius and antibody production. *Poultry Science* 35**,** 224-225.

Gould, J. (2016). "Morpheus: Versatile matrix visualization and analysis software".).

Gundra, U.M., Girgis, N.M., Gonzalez, M.A., San Tang, M., Van Der Zande, H.J., Lin, J.-D., Ouimet, M., Ma, L.J., Poles, J., and Vozhilla, N. (2017). Vitamin A mediates conversion of monocyte-derived macrophages into tissue-resident macrophages during alternative activation. *Nature Immunology* 18**,** 642-653.

Hamilton, T., and Mitchell, H. (1924). The occurrence of lactase in the alimentary tract of the chicken. *Journal of Agriculture Research* 27**,** 605-608.

Han, B., He, Y., Zhang, L., Ding, Y., Lian, L., Zhao, C., Song, J., and Yang, N. (2017). Long intergenic non-coding RNA GALMD3 in chicken Marek's disease. *Scientific Reports* 7.

Hebiguchi, T., Mezaki, Y., Morii, M., Watanabe, R., Yoshikawa, K., Miura, M., Imai, K., Senoo, H., and Yoshino, H. (2015). Massive bowel resection upregulates the intestinal mRNA expression levels of cellular retinol-binding protein II and apolipoprotein A-IV and alters the intestinal vitamin A status in rats. *International Journal of Molecular Medicine* 35**,** 724-730.

Heidari, M., Fitzgerald, S.D., and Zhang, H. (2015). Immune Responses in Cecal Tonsils of Marek's Disease Virus–Infected Chickens. *Avian Diseases* 59**,** 213-226.

Ikehara, M., and Fukui, T. (1974). Studies of nucleosides and nucleotides.: LVIII. Deamination of adenosine analogs with calf intestine adenosine deaminase. *Biochimica et Biophysica Acta (BBA)-General Subjects* 338**,** 512-519.

Kent, W.J., Sugnet, C.W., Furey, T.S., Roskin, K.M., Pringle, T.H., Zahler, A.M., and Haussler, D. (2002). The human genome browser at UCSC. *Genome Research* 12**,** 996-1006.

Khan, A., and Mathelier, A. (2017). Intervene: a tool for intersection and visualization of multiple gene or genomic region sets. *BMC Bioinformatics* 18**,** 1-8.

Klasing, K.C. (1984). Effect of inflammatory agents and interleukin 1 on iron and zinc metabolism. *American Journal of Physiology-Regulatory, Integrative and Comparative Physiology* 247**,** R901-R904.

Korte, J., Fröhlich, T., Kohn, M., Kaspers, B., Arnold, G.J., and Härtle, S. (2013). 2D DIGE analysis of the bursa of Fabricius reveals characteristic proteome profiles for different stages of chicken B‐cell development. *Proteomics* 13**,** 119-133.

Krag, E., and Phillips, S.F. (1974). Active and passive bile acid absorption in man. Perfusion studies of the ileum and jejunum. *The Journal of Clinical Investigation* 53**,** 1686-1694.

Kreslavsky, T., Vilagos, B., Tagoh, H., Poliakova, D.K., Schwickert, T.A., Wöhner, M., Jaritz, M., Weiss, S., Taneja, R., and Rossner, M.J. (2017). Essential role for the transcription factor Bhlhe41 in regulating the development, self-renewal and BCR repertoire of B-1a cells. *Nature immunology* 18**,** 442-455.

Kuo, R.I., Tseng, E., Eory, L., Paton, I.R., Archibald, A.L., and Burt, D.W. (2017). Normalized long read RNA sequencing in chicken reveals transcriptome complexity similar to human. *BMC Genomics* 18.

Leblanc, J.G., Milani, C., De Giori, G.S., Sesma, F., Van Sinderen, D., and Ventura, M. (2013). Bacteria as vitamin suppliers to their host: a gut microbiota perspective. *Current Opinion in Biotechnology* 24**,** 160-168.

Lenschow, D.J., Walunas, T.L., and Bluestone, J.A. (1996). CD28/B7 system of T cell costimulation. *Annual Review of Immunology* 14**,** 233-258.

Li, B., and Dewey, C.N. (2011). RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC bioinformatics* 12**,** 1-16.

Li, T., Riederer, B., Liu, X., Pallagi, P., Singh, A.K., Soleimani, M., and Seidler, U. (2016). Tu1491 loss of Slc26a9 anion transporter results in reduced pancreatic fluid secretion in young female mice. *Gastroenterology* 150**,** S915.

Liao, Y., Wang, J., Jaehnig, E.J., Shi, Z., and Zhang, B. (2019). WebGestalt 2019: gene set analysis toolkit with revamped UIs and APIs. *Nucleic Acids Research* 47**,** W199-W205.

Liu, H.C., Cheng, H., Tirunagaru, V., Sofer, L., and Burnside, J. (2001). A strategy to identify positional candidate genes conferring Marek's disease resistance by integrating DNA microarrays and genetic mapping. *Animal genetics* 32**,** 351-359.

Liu, Y., Sun, Y., Li, Y., Bai, H., Xue, F., Xu, S., Xu, H., Shi, L., Yang, N., and Chen, J. (2017). Analyses of long non-coding RNA and mRNA profiling using RNA sequencing in chicken testis with extreme sperm motility. *Scientific Reports* 7**,** 1-8.

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

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

Lundholm, M., Mayans, S., Motta, V., Löfgren-Burström, A., Danska, J., and Holmberg, D. (2010). Variation in the CD3ζ (Cd247) gene correlates with altered T cell activation and is associated with autoimmune diabetes. *The Journal of Immunology* 184**,** 5537-5544.

Marini, F. (2016). "pcaExplorer: interactive visualization of RNA-seq data using a principal components approach".).

Marini, F., and Binder, H. (2019). pcaExplorer: an R/Bioconductor package for interacting with RNA-seq principal components. *BMC Bioinformatics* 20**,** 1-8.

Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet. Journal* 17**,** 10-12.

Matthews, J.T. (2006). Egg-based production of influenza vaccine: 30 years of commercial experience. *Bridge-Washington-National Academy of Engineering* 36**,** 17.

Mattick, J.S., and Makunin, I.V. (2006). Non-coding RNA. *Human Molecular Genetics* 15**,** R17-R29.

Mccartney-Francis, N., Mizel, D., Wong, H., Wahl, L., and Wahl, S. (1990). TGF-β regulates production of growth factors and TGF-β by human peripheral blood monocytes. *Growth Factors* 4**,** 27-35.

Mcneill, E., Crabtree, M.J., Sahgal, N., Patel, J., Chuaiphichai, S., Iqbal, A.J., Hale, A.B., Greaves, D.R., and Channon, K.M. (2015). Regulation of iNOS function and cellular redox state by macrophage Gch1 reveals specific requirements for tetrahydrobiopterin in NRF2 activation. *Free Radical Biology and Medicine* 79**,** 206-216.

Mihály, J., Gamlieli, A., Worm, M., and Rühl, R. (2011). Decreased retinoid concentration and retinoid signalling pathways in human atopic dermatitis. *Experimental Dermatology* 20**,** 326-330.

Mijušković, M., Chou, Y.-F., Gigi, V., Lindsay, C.R., Shestova, O., Lewis, S.M., and Roth, D.B. (2015). Off-target V (D) J recombination drives lymphomagenesis and is escalated by loss of the Rag2 C terminus. *Cell* 12**,** 1842-1852.

Muret, K., Klopp, C., Wucher, V., Esquerré, D., Legeai, F., Lecerf, F., Désert, C., Boutin, M., Jehl, F., and Acloque, H. (2017). Long noncoding RNA repertoire in chicken liver and adipose tissue. *Genetics Selection Evolution* 49**,** 1-17.

Murugesan, G.R., Romero, L.F., and Persia, M.E. (2014). Effects of protease, phytase and a Bacillus sp. direct-fed microbial on nutrient and energy digestibility, ileal brush border digestive enzyme activity and cecal short-chain fatty acid concentration in broiler chickens. *PloS One* 9**,** e101888.

Nakayama, A., Eguchi, O., Hatakeyama, M., Saitoh, H., and Takada, M. (1999). Different absorption behaviors among steriod hormones due to possible interaction with P-glycoprotein in the rat small intestine. *Biological & Pharmaceutical Bulletin* 22**,** 535-538.

Nap, A.W., Dunselman, G.A., De Goeij, A.F., Evers, J.L., and Groothuis, P.G. (2004). Inhibiting MMP activity prevents the development of endometriosis in the chicken chorioallantoic membrane model. *Human Reproduction* 19**,** 2180-2187.

Nap, A.W., Groothuis, P.G., Demir, A.Y., Maas, J.W., Dunselman, G.A., De Goeij, A.F., and Evers, J.L. (2003). Tissue integrity is essential for ectopic implantation of human endometrium in the chicken chorioallantoic membrane. *Human Reproduction* 18**,** 30-34.

Ncc (2019). "Broiler chicken industry key facts 2019". (: National Chicken Council).

Peng, L., Van Den Biggelaar, R.H., Jansen, C.A., Haagsman, H.P., and Veldhuizen, E.J. (2020). A method to differentiate chicken monocytes into macrophages with proinflammatory properties. *Immunobiology* 225**,** 152004.

Perčulija, V., and Ouyang, S. (2019). "Diverse roles of DEAD/DEAH-Box helicases in innate immunity and diseases," in *Helicases from All Domains of Life*. Elsevier), 141-171.

Pertea, G., and Pertea, M. (2020). GFF Utilities: GffRead and GffCompare [version 1; peer review: 2.

Plimmer, R.H.A., and Rosedale, J.L. (1922). Distribution of enzymes in the alimentary canal of the chicken. *Biochemical Journal* 16**,** 23-26.

Qiang, Y.-W., Endo, Y., Rubin, J.S., and Rudikoff, S. (2003). Wnt signaling in B-cell neoplasia. *Oncogene* 22**,** 1536-1545.

Quinn, J.J., and Chang, H.Y. (2016). Unique features of long non-coding RNA biogenesis and function. *Nature Reviews Genetics* 17**,** 47.

Richard, P., Darzacq, X., Bertrand, E., Jády, B.E., Verheggen, C., and Kiss, T. (2003). A common sequence motif determines the Cajal body‐specific localization of box H/ACA scaRNAs. *The EMBO Journal* 22**,** 4283-4293.

Roeszler, K.N., Itman, C., Sinclair, A.H., and Smith, C.A. (2012). The long non-coding RNA, MHM, plays a role in chicken embryonic development, including gonadogenesis. *Developmental Biology* 366**,** 317-326.

Rosenblad, M.A., Zwieb, C., and Samuelsson, T. (2004). Identification and comparative analysis of components from the signal recognition particle in protozoa and fungi. *BMC Genomics* 5**,** 5.

Ross, A.C., Chen, Q., and Ma, Y. (2011). "Vitamin A and retinoic acid in the regulation of B-cell development and antibody production," in *Vitamins & Hormones*. Elsevier), 103-126.

Sato, K., Kawasaki, H., Nagayama, H., Enomoto, M., Morimoto, C., Tadokoro, K., Juji, T., and Takahashi, T.A. (2000). TGF-β1 reciprocally controls chemotaxis of human peripheral blood monocyte-derived dendritic cells via chemokine receptors. *The Journal of Immunology* 164**,** 2285-2295.

Schianca, G.C., Pedrazzoli, R., Onolfo, S., Colli, E., Cornetti, E., Bergamasco, L., Fra, G., and Bartoli, E. (2011). ApoB/apoA-I ratio is better than LDL-C in detecting cardiovascular risk. *Nutrition, Metabolism and Cardiovascular Diseases* 21**,** 406-411.

Schmid, M., Nanda, I., Guttenbach, M., Steinlein, C., Hoehn, M., Schartl, M., Haaf, T., Weigend, S., Fries, R., and Buerstedde, J. (2000). First report on chicken genes and chromosomes 2000. *Cytogenetic and Genome Research* 90**,** 169-218.

Shan, S.-O., and Walter, P. (2005). Co-translational protein targeting by the signal recognition particle. *FEBS Letters* 579**,** 921-926.

Shang, Y., Kumar, S., Oakley, B., and Kim, W.K. (2018). Chicken gut microbiota: importance and detection technology. *Frontiers in Veterinary Science* 5**,** 254.

Shepard, S., Mccreary, M., and Fedorov, A. (2009). The Peculiarities of large intron splicing in animals. *PloS One* 4**,** e7853.

Stone, H.A. (1975). *Use of highly inbred chickens in research.* Agricultural Research Service, US Department of Agriculture.

Su, R.-C., Brown, K.E., Saaber, S., Fisher, A.G., Merkenschlager, M., and Smale, S.T. (2004). Dynamic assembly of silent chromatin during thymocyte maturation. *Nature Genetics* 36**,** 502-506.

Su, R.-C., Sridharan, R., and Smale, S.T. (Year). "Assembly of silent chromatin during thymocyte development", in: *Seminars in Immunology*: Elsevier), 129-140.

Tan, C., Hiwa, R., Mueller, J.L., Vykunta, V., Hibiya, K., Noviski, M., Huizar, J., Brooks, J.F., Garcia, J., and Heyn, C. (2020). NR4A nuclear receptors restrain B cell responses to antigen when second signals are absent or limiting. *Nature Immunology* 21**,** 1267-1279.

Tancharoenrat, P., Ravindran, V., Zaefarian, F., and Ravindran, G. (2014). Digestion of fat and fatty acids along the gastrointestinal tract of broiler chickens. *Poultry Science* 93**,** 371-379.

Tso, P., Sun, W., and Liu, M. (2004). Gastrointestinal satiety signals IV. Apolipoprotein A-IV. *American Journal of Physiology-Gastrointestinal and Liver Physiology* 286**,** G885-G890.

Uep (2019). *Facts and stats* [Online]. United Egg Producers, Johns Creek, GA, USA. . Available: https://unitedegg.com/facts-stats/#:~:text=U.S.%20table%20egg%20production%20totaled,produces%20294%20eggs%20per%20year. [Accessed].

Van Laethem, F., Tikhonova, A.N., Pobezinsky, L.A., Tai, X., Kimura, M.Y., Le Saout, C., Guinter, T.I., Adams, A., Sharrow, S.O., and Bernhardt, G. (2013). Lck availability during thymic selection determines the recognition specificity of the T cell repertoire. *Cell* 154**,** 1326-1341.

Wang, M., Mao, Y., Wang, B., Wang, S., Lu, H., Ying, L., and Li, Y. (2020). Quercetin improving lipid metabolism by regulating lipid metabolism pathway of ileum mucosa in broilers. *Oxidative Medicine and Cellular Longevity* 2020.

Warren, W.C., Hillier, L.W., Tomlinson, C., Minx, P., Kremitzki, M., Graves, T., Markovic, C., Bouk, N., Pruitt, K.D., and Thibaud-Nissen, F. (2017). A new chicken genome assembly provides insight into avian genome structure. *G3: Genes, Genomes, Genetics* 7**,** 109-117.

Weiss, H., Friedrich, T., Hofhaus, G., and Preis, D. (1991). "The respiratory-chain NADH dehydrogenase (complex I) of mitochondria," in *EJB Reviews 1991*. Springer), 55-68.

Wight, T.N., Kang, I., and Merrilees, M.J. (2014). Versican and the control of inflammation. *Matrix Biology* 35**,** 152-161.

Wu, Z., Harne, R., Chintoan-Uta, C., Hu, T.-J., Wallace, R., Maccallum, A., Stevens, M.P., Kaiser, P., Balic, A., and Hume, D.A. (2020). Regulation and function of macrophage colony-stimulating factor (CSF1) in the chicken immune system. *Developmental & Comparative Immunology* 105**,** 103586.

Xing, S., Gai, K., Li, X., Shao, P., Zeng, Z., Zhao, X., Zhao, X., Chen, X., Paradee, W.J., and Meyerholz, D.K. (2019). Tcf1 and Lef1 are required for the immunosuppressive function of regulatory T cells. *Journal of Experimental Medicine* 216**,** 847-866.

Zhang, J.D., Hatje, K., Sturm, G., Broger, C., Ebeling, M., Burtin, M., Terzi, F., Pomposiello, S.I., and Badi, L. (2017a). Detect tissue heterogeneity in gene expression data with BioQC. *BMC Genomics* 18**,** 1-9.

Zhang, T., Zhang, X., Han, K., Zhang, G., Wang, J., Xie, K., Xue, Q., and Fan, X. (2017b). Analysis of long noncoding RNA and mRNA using RNA sequencing during the differentiation of intramuscular preadipocytes in chicken. *PloS One* 12**,** e0172389.

Zhang, T., Zhang, X., Han, K., Zhang, G., Wang, J., Xie, K., Xue, Q., and Fan, X. (2017c). Analysis of long noncoding RNA and mRNA using RNA sequencing during the differentiation of intramuscular preadipocytes in chicken. 12**,** e0172389.

Zhang, Y., Cao, L., Yang, B.L., and Yang, B.B. (1998). The G3 domain of versican enhances cell proliferation via epidermial growth factor-like motifs. *Journal of Biological Chemistry* 273**,** 21342-21351.

Zhang, Y., Wang, J., Huang, S., Zhu, X., Liu, J., Yang, N., Song, D., Wu, R., Deng, W., and Skogerbø, G. (2009). Systematic identification and characterization of chicken (Gallus gallus) ncRNAs. *Nucleic Acids Research* 37**,** 6562-6574.

Zhao, C., Inoue, J., Imoto, I., Otsuki, T., Iida, S., Ueda, R., and Inazawa, J. (2008). POU2AF1, an amplification target at 11q23, promotes growth of multiple myeloma cells by directly regulating expression of a B-cell maturation factor, TNFRSF17. *Oncogene* 27**,** 63-75.

Zheng, X.L., Kitamoto, Y., and Sadler, J.E. (2009). Enteropeptidase, a type II transmembrane serine protease. *Front Biosci (Elite Ed)* 1**,** 242-249.

Zhu, J., Yamane, H., and Paul, W.E. (2009). Differentiation of effector CD4 T cell populations. *Annual Review of Immunology* 28**,** 445-489.

Chapter 2: Combinatorial study of spaceflight factors - cephalad fluid-shift and gamma radiation.

Chapter 2 is adapted with minimal modification from:

Mice Exposed to Combined Chronic Low-Dose Irradiation and Modeled Microgravity Develop Long-Term Neurological Sequelae. Eliah G. Overbey, Amber M. Paul, *et al.* IJMS. 2019.

2.1 Abstract

Spaceflight poses many challenges for humans. Ground-based analogs typically focus on single parameters of spaceflight and their associated acute effects. This study assesses the long-term transcriptional effects following single and combination spaceflight analog conditions using the mouse model: simulated microgravity via hindlimb unloading (HLU) and/or low-dose γ-ray irradiation (LDR) for 21 days, followed by 4 months of readaptation. Changes in gene expression and epigenetic modifications in brain samples during readaptation were analyzed by whole transcriptome shotgun sequencing (RNA-seq) and reduced representation bisulfite sequencing (RRBS). The results showed minimal gene expression and cytosine methylation alterations at 4 months readaptation within single treatment conditions of HLU or LDR. In contrast, following combined HLU+LDR, gene expression and promoter methylation analyses showed multiple altered pathways involved in neurogenesis and neuroplasticity, the regulation of neuropeptides, and cellular signaling. In brief, neurological readaptation following combined chronic LDR and HLU is a dynamic process that involves pathways that regulate neuronal function and structure and may lead to late onset neurological sequelae.

2.2 Introduction

The central nervous system (CNS) is vulnerable to irradiation (Anna Kovalchuk and Kolb, 2017a) and fluid shifts (Lawley et al., 2017) experienced during short- and long-term spaceflight. An important goal for the National Aeronautics and Space Administration (NASA) is to identify the effects of spaceflight-like conditions on the CNS to better prepare astronauts during long-duration missions to the Moon and Mars. During missions beyond low-earth orbit (LEO), astronauts will be continuously exposed to low-dose ionizing radiation (LDR). While high linear energy transfer (LET) galactic cosmic radiation (GCR) and low LET solar particle event (SPE) radiation contribute large portions of the radiation dose accumulated by astronaut crew members (Chancellor, Scott, and Sutton 2014), interactions between GCR and SPE particles and spacecraft release secondary radiation including γ-rays that can deliver a significant fraction of the total mission does (Seawright et al., 2017). In addition, altered gravity (hypergravity experienced at

launch/landing and microgravity experienced in-flight) affects homeostatic fluid-shifting in the CNS. Studies examining readaptation to Earth's gravity are of equal importance, as recovery from spaceflight exposure may require medical intervention, in particular within organs known to be sensitive to irradiation and fluid shifts. Astronauts in spaceflight experience impairments in neurocognition (as measured by altered decision making and problem solving) and neurobehavior (as measured by impaired visual perceptions, motor controls and sleep–wake cycles) (De la Torre 2014). The NASA Twins study offers some additional insight into post-flight cognitive decline that persists up to 6 months post-flight (end of experimental sampling) (Garrett-Bakelman et al., 2019). Longitudinal diffusion magnetic resonance imaging (MRI) data collected from astronauts pre- and post-flight (shuttle missions <30 days and ISS mission <200 days) revealed whole brain shifting within the skull and white matter disruptions, suggesting spaceflight leads to long-term alterations of brain structure in humans (J. K. Lee et al., 2019). Rodent models offer an appropriate alternative to assess the effects of spaceflight-like conditions on brain tissue. Recently, the Rodent Research (RR)-1 mission flew mice to the International Space Station (ISS) for 33 days and measured neurobehavioral outcomes through video monitoring, which indicated 16-week old (at launch) females displayed unique group hyperactivity behavior at 2 weeks post-launch, implying neurological disruptions occur in flight (Ronca et al., 2019). Moreover, long-term readaptation to chronic LDR and hindlimb unloading (HLU) (a ground-based analog of spaceflight) resulted in increased aquaporin-4 (AQP4) protein expression, oxidative stress damage, apoptosis, blood–brain barrier (BBB) compromise and an increase in neurobehavioral risk-taking behavior (Bellone et al., 2016; Xiao Wen Mao et al., 2016, 2017). The effects of acute radiation doses on the CNS have been well documented (Smart 2017). However, responses to chronic LDR exposure, as encountered on space missions (<0.04 Gy/y), are not well investigated (Cucinotta and Cacao 2019), particularly in a pan-transcriptome context. Moreover, long-term readaptive outcomes on neurohealth following spaceflight-like conditions are less studied. Due to the need to better understand the transcriptional changes associated with chronic LDR in combination with reduced gravity (unloading) on neurohealth, mice were exposed to chronic LDR γ-ray irradiation (0.04 Gy), HLU, or a combination of HLU+LDR for 21 days, followed by a 4-month readaptation period. Whole transcriptome shotgun sequencing (RNA-seq) data were analyzed for differentially expressed genes (DEG), while data from reduced representation bisulfite sequencing (RRBS) were analyzed to determine differentially methylated promoters (DMP) ("GLDS-202: Low Dose (0.04 Gy) Irradiation (LDR) and Hindlimb Unloading (HLU) Microgravity in Mice: Brain Transcriptomic and Epigenomic Data" n.d.). Overall, we found minimal gene alterations at 4 months of readaptation within single treatment conditions of HLU or LDR, namely, in pathways related to the reduced expression of translational machinery, while the combination of HLU+LDR resulted in a wide panel of altered DEG and DMP profiles for tight junctions, aquaporins, neurogenesis markers and neuropeptides. Therefore, exposure to spaceflight-like conditions may lead to mild, long-term neurological

consequences, while readaptation is an active process involving neuroplasticity and repair that is cooperatively engaged post-HLU+LDR exposure to maintain neural homeostasis.

2.3 Results

2.3.1 *Experimental Design and Mouse Body Weights Across Experimental and Readaptation Timepoints*

To assess the effects of prolonged spaceflight-like conditions on central nervous system (CNS) health following exposure, 6-month-old female *C57BL/6J* mice were either irradiated with low-dose γ-ray irradiation (0.04 Gy, LDR), hindlimb unloaded (HLU) or a combination of the two (HLU+LDR) for 21 days, followed by a 4-month readaptation period (Figure 1A). While the LDR γ-ray irradiation used does not recapitulate the entire space radiation spectrum, it constitutes a significant fraction and poses a hazard to astronauts' health (Chatterjee et al., 2019).

Total body weights were monitored from the initiation of the experimental conditions (0 days), throughout the duration of exposure (up to 21 days), and during the readaptation period (4 months, 4M). Mice were weighted weekly at the same time as the 21-day exposure and monthly for the 4-month readaptation phase. Baseline weight measurements were taken on day 0, prior to the initiation of HLU and LDR exposure. While there was no difference with LDR alone compared to control animals, a trend for decreased body weight in HLU and HLU+LDR combination conditions versus controls was noted, while statistical significance was observed at 7 days post-experimental start in HLU only (Fig 1B). Furthermore, during the readaptation period, decreased weights observed in HLU and in HLU+LDR animals recovered to control values, indicating physiological readaptation following exposure to spaceflight-like conditions.

Modeled microgravity in combination with chronic low-dose irradiation results in increased differential gene expression in brains at 4 months post-exposure. To determine the readaptation effects on the CNS at 4 months post-exposure, whole brains were isolated and mid-sagittal sectioned into two hemispheres. The right hemisphere was further sectioned mid-coronal and the caudal cortex (including hind- and midbrain) regions were collected for RNA-seq analysis. Collected brains were free of cerebellum and brainstem contamination. Principal component analysis (PCA) of RNA-seq data revealed clustering of samples by experimental conditions (Fig 2A), while volcano plots showed limited DEG ($p < 0.05$, Log2 fold change > 0.263) in all but the combination HLU+LDR (Fig 2B).

Brain readaptation following exposure to LDR alone displayed minimal (5) DEG with large fold change inductions only in genes of unknown function (Fig 3A, Table 1). Brain readaptation following exposure to HLU alone displayed 11 DEG with known functions primarily associated with reduced protein synthesis and/or functions mediated by ribosomes (Fig 3B, Table 1). In contrast to single exposures, brain readaptation following exposure to a combination of HLU+LDR displayed numerous (270) DEG associated with various cellular functions including tight junctions, aquaporins, neurogenesis and neuropeptide production (Fig 3C, Table 1). These results showed minimal transcriptional changes in single treatment groups at 4 months post-exposure, while the combined exposure to HLU+LDR resulted in multiple DEG.

2.3.2 LDR, HLU and Combined HLU+LDR Exposures Displayed Minimal Gene and Promoter Methylation Overlap at 4 Months Readaptation

Given that there were significant changes in gene expression in the combined exposure to HLU+LDR mice at the 4-month time point, we next determined whether these changes in gene expression reflected modifications to cytosine methylation using RRBS analysis. RRBS revealed 137, 71 and 170 differentially methylated promoters (DMPs) in the LDR, HLU, and HLU+LDR animals, respectively (Table 1). To find overlap across assays and experimental conditions, gene sets were created from DEG (upregulated/downregulated) and genes corresponding to DMPs (hypermethylated/hypomethylated). Overlapping genes between these sets were determined

(Table 2) and summarized in an UpSet diagram (Khan and Mathelier, 2017) (Fig 4). Overall, there was little overlap between assays and experimental conditions (grey bars/dots/line). Most overlap was observed between DMPs from different experimental conditions (red bars/dots/lines). Overlap between DEGs from different experimental conditions was minimal (blue bars/dots/lines). Overlap between DEGs and DMPs from the same experimental condition or from different experimental conditions were also minimal (black bars/dots/lines). Collectively, these results indicated most DEGs and DMPs are unique to their experimental condition with minimal overlap between exposure groups.

3.3.3 Readaptation at 4 Months Following Exposure to Combined HLU+LDR Involves Gene Expression Changes in Multiple Brain-Associated Pathways

To determine the associated mechanisms involved in readaptation and neuroplasticity from simulated spaceflight combined HLU+LDR exposure, we used WebGestalt to detect enriched gene ontology (GO) terms in sets of up- and downregulated genes in post-HLU+LDR animals (Fig 5A, B). The downregulated set was enriched for GO terms associated with neurogenesis, transmembrane transport, cell signaling, and neurosynaptic and brain architecture deficits (Fig 5A). The upregulated DEG set was enriched for terms associated with cell proliferation, neurogenesis, and hormonal regulation (Fig 5B). Mouse Reactome pathway analyses of DEGs in the post-HLU+LDR animals indicated biological processes associated with signal transduction, neurodevelopment and neurohomeostasis (Sup Fig 1A). Human Reactome pathway translational analyses identified human orthologs involved in neurotransmission and cellular signaling (Sup Fig 1B). Collectively, brain-readaptation post-combination HLU+LDR resulted in altered expression in pathways governing biological processes associated with neuroplasticity, neurogenesis, neurostructural reorganization and neuropeptide production.

2.4 Discussion

Herein, we identified that chronic exposure (21 days) to single or combined HLU and LDR resulted in long-term transcriptional alterations in CNS brain tissue in female mice at 4 months post-exposure. The gene expression changes observed here imply reduced transcriptional machinery, increased neurogenesis and neuropeptide production, and dysregulated cell structure and cell signaling. These studies offer translational value into pathways regulated in humans, as gene ontology analyses identified similar genetic overlap between mice and human profiles, in particular in pathways involved in neurodevelopmental homeostasis and signal transduction. Collectively, brain-related transcriptional changes that are dynamic in readaptation from exposure to spaceflight-like conditions may lead to long-term neurological consequences. Indeed, 21 days of exposure to combined HLU+LDR did not result in immediate

neurobehavioral deficits (1 week post-exposure); however, increased risk-taking behavior was observed at 9 months post-exposure (Bellone et al., 2016), indicating delayed neurological impairments in a combination of HLU+LDR, that was not apparent in HLU or LDR only.

To assess CNS effects, 4-month post-experimental conditions mid/hindbrain (sans cerebellum and brainstem) were collected, sectioned, and total RNA was isolated for RNA-seq analysis. This portion of the brain houses the hippocampus, basal ganglia, substantia nigra, and thalamus, and is a major signaling hub for the neuroendocrine system (hypothalamus and pituitary gland). Mid/hindbrain sections were assessed to localize the transcriptional profile to a region that is central to neurogenesis (hippocampus) and hypothalamic–pituitary axis (HPA) responses, in order to better assess neurofunctional output without potential contamination from the cerebral cortex or cerebellum. The results displayed robust DEG alterations within combination HLU+LDR at 4 months readaptation compared to controls, and marginal DEG alterations in single HLU or LDR at 4 months readaptation. At 4 months post-HLU exposure, DEGs including *EphrB3* and *Sh2d5* were altered, suggesting cell morphological and structural changes are active at this time point. At 4 months post-LDR exposure, only five DEGs were found, four of which have an unknown function. One DEG, *Rps13*, was reduced, suggesting impaired protein synthesis and/or other functional outputs mediated by ribosomes. Interestingly, *Rps13* was also reduced in HLU alone at 4 months readaptation along with other protein synthesis-related genes, including *Hist1h2bc* and *Rpl36a-ps2*, collectively suggesting fundamental translational pathways are altered during chronic HLU and LDR during readaptation. In line with this, another group has reported that chronic LDR whole brain irradiation resulted in protein synthesis impairment (A. Kovalchuk and Kolb 2017) that can affect adult neurogenesis (Mizumatsu et al., 2003). Interestingly, no notable protein synthesis pathway deficits were identified in combined HLU+LDR exposures, suggesting distinct responses are generated within each experimental condition at readaptation. Moreover, very limited DEG overlap was observed between experimental conditions, further validating unique responses to each condition and highlighting the importance of implementing combined exposure studies to better assess spaceflight-like readaptation effects on the nervous system. Other spaceflight-like factors such as hypoxia and social isolation were not assessed in this study and merit further investigation. Nonetheless, the usefulness of this model provides a scaffold for future studies to assess the long-term effects of combination HLU+LDR on the health of the nervous system.

Brain structural damage through disruption of the blood–brain barrier (BBB) is associated with neurodegeneration and subsequent behavioral dysfunctions (Najjar et al., 2013). Our previous study showed that aquaporin-4 (AQP4), a water-channel protein involved in brain water homeostasis, was up-regulated following combined exposure to irradiation and unloading (Bellone et al., 2016), suggesting a disturbance in BBB integrity may lead to edema (Song et al., 2016). Furthermore, AQP4 induction was seen in concert with subtle behavioral changes at 9 months post-exposure (Bellone et al., 2016), indicating a potential role for BBB dysfunction and

neurobehavioral consequences. Although *Aqp4* was not identified in this study, *Aqp1* gene expression was induced following the combination of HLU+LDR at 4 months readaptation, suggesting similar BBB disturbances and possible shifts in brain water homeostasis (Song et al., 2016). In line with this, the induction of aquaporins may be linked to neuroinflammation (L. Li et al., 2011; Ikeshima-Kataoka 2016; Meli, Pirozzi, and Pelagalli 2018), while the upregulation of additional immune activation markers including *Cd74*, *H2-Ab1, C1ql2* and *Ifit-3* were also identified in combination HLU+LDR at 4 months readaptation, along with hypomethylated *Cfi and Mx1*, collectively suggesting chronic neuroinflammation may persist during brain readaptation post-combined HLU+LDR exposure.

In response to extracellular cues (i.e., edema), cells rearrange their structures and ion transport proteins to maintain equilibrium (Meli, Pirozzi, and Pelagalli 2018; Kotas and Medzhitov 2015; Medzhitov 2008). Furthermore, neuroplasticity is a homeostatic repair mechanism that involves the formation of new synapses and neural processes to compensate for their communication loss following brain infarct, injury or disease (Smart 2017). These new formations result in brain restructuring and function, which may be occurring during readaptation post-combined HLU+LDR exposures. Indeed, our results fall in line with responses that were observed within brains isolated hours post-LDR (10 cGy) exposure, whereby multiple synaptic ion transport genes were altered (Lowe et al., 2009). Herein, multiple cell structural genes, i.e., *Drc7*, *Dnah7b* and *Fmnl1* and ion transport genes, i.e., *Gria3*, *Grin2a*, *Calb2*, *Scn5a*, *Clcn5*, *Kcnj13*, *Kcnn1* and *Kcnh1*, were altered post-combination HLU+LDR conditions. Moreover, hypomethylation analyses indicated altered cytoskeletal organization, i.e., *Ndel1*, *Cald1* and *Lmod3* were observed in both LDR and HLU, suggesting cellular structural changes involved in neuroplasticity and readaptation processes are active, in order to restructure and/or maintain brain homeostasis.

Tight junctions, including claudins, connexins, and occludins contribute to BBB permeability and maintenance. A reduction in these molecules has been implicated in acute BBB disruption in neuroviral (Paul et al., 2017) and neuroinflammatory diseases (Yamazaki et al., 2019). In line with this, *Cdh12* was downregulated, whereas, *Cldn2* was induced in combination of HLU+LDR. Furthermore, *Cdhr4* was hypomethylated in both LDR alone and HLU+LDR at 4 months readaptation, while *Cdhr2* was hypomethylated in LDR and HLU alone, but not in combination HLU+LDR at 4 months readaptation, suggesting BBB restructuring, and further indicating the dynamic process of brain physiology during readaptation.

Adult neurogenesis is a process of neural and glia self-renewal and the generation of progenitor cells from radial glial cells (RGCs) within the subventricular and subgranular zones of the hippocampus (C. Zhao, Deng, and Gage 2008). RGCs are quiescent and are activated to initiate neurogenesis in response to a stimulus (Zhang and Jiao 2015), such as LDR (Mizumatsu et al., 2003; X. Z. Sun et al., 1997) or models of increased intracranial pressure (Kleindienst et al., 2005; D. Sun et al., 2009). Indeed, the majority of highly upregulated DEGs post-combination HLU+LDR were involved in HPA signaling, neuropeptide growth factors, and

neurogenesis, i.e., *Otp*, *Uncx*, *Ucn3*, *Avp*, *Pmch*, *Prox1*, *Calb2*, *Tbr-1*, and *Hcrt*. While the magnitude of gene expression does not always result in observable phenotypes, these notable DEGs may serve an important purpose in neuromaintenance during readaptation. Interestingly, *Crh*, a stress-induced hormone, was hypermethylated post-exposure to LDR and combination HLU+LDR at 4 months readaptation, implying transcriptional silencing of *Crh* may be engaged to suppress robust HPA activity during readaptation, which warrants further studies.

Readaptation results from combined exposures to HLU+LDR are in agreement with literature that spaceflight can alter neurofunction. Indeed, neurocognitive and behavioral effects via altered decision making, problem solving, visual perceptions, motor controls, and sleep–wake cycles have all been identified in flight (De la Torre 2014). While "space-brain" characteristics have suggested astronauts are at risk of developing neurological sequelae in flight (Jandial et al., 2018) and post-flight (Garrett-Bakelman et al., 2019; J. K. Lee et al., 2019). Due to the limitations of human neurological studies, mouse models provide an excellent alternative, with the caveat that fluid shifts in mice are marginal in comparison to their human counterparts. Therefore, follow-up studies utilizing larger mammals, i.e., rats, may provide more robust responses. Further, readaptation effects from radiation have been described to cause differential nervous system responses in sex, age, type of irradiation, duration of exposure and total doses (Anna Kovalchuk and Kolb 2017b; Smart 2017; Grammaticos, Giannoula, and Fountos 2013), while the readaptation effects of prolonged HLU on the neurohealth of different sexes and age groups are still unclear. Therefore, future studies approaching these additional demographics, both modeled in mice or rats, are required. Nonetheless, our readaptation study in mice exposed to combined HLU+LDR displayed a panel of neurological changes that were translational to humans, suggesting this is a suitable animal model to study spaceflight-related readaptation effects on the nervous system. Additionally, future studies should expand the types of radiation exposure animals experience in order to fully capture the spectrum of radiation that bombards astronauts during spaceflight. The space radiation environment beyond LEO contains several types of ionizing radiation. Galactic cosmic radiation (GCR) with relatively high linear energy transfer (LET) and proton radiation due to solar particle events (SPEs) contribute a significant portion of the radiation dose accumulated by astronaut crew members (Chancellor, Scott, and Sutton 2014). High-energy heavy ions (HZE radiation) can produce distinct patterns of energy deposition in cells and tissues (Cucinotta and Cacao 2019; Walker, Townsend, and Norbury 2013). However, secondary particles produced by the interaction of SPE protons and heavy-charged GCR particles with the spacecraft structure include γ-rays and x-rays (Chancellor, Scott, and Sutton 2014). These radiation sources can also deliver a significant fraction of the total mission dose and pose a hazard to astronauts' health (Chatterjee et al., 2019). For chronic irradiation exposure, we used flood sources consisting of plastic sheets embedded with the isotope 57Co that emits medium energy (122 and 136 keV) γ-rays. Additional research is needed to elucidate the CNS damage and readaptation response induced by the full spectrum of space radiation.

In summary, a multitude of neurological effects and persistent brain transcriptional alterations were observed at 4 months post-HLU+LDR in 21-day exposures. Changes included reduced transcriptional machinery, increased neurogenesis and neuropeptide production, and dysregulated cell structure and signaling genes. Thus, brain-related transcriptional changes are dynamic and plastic during readaptation from exposure to spaceflight-like conditions, and while these exposures may lead to long-term neurological consequences, the active processes of neuroplasticity and repair are engaged to maintain neural homeostasis.

2.5 Materials and Methods

2.5.1. Animals

Six-month-old, female C57BL/6J mice (Jackson Laboratory), as previously reported (Xiao Wen Mao et al., 2016), were utilized in this report. Upon arrival, they were acclimatized for 7 days in standard habitats at 20 °C with a 12 h:12 h light:dark cycle. Commercial pellet chow and hydrogel were available ad libitum. Animal health status, water and food intake were monitored daily. The study followed recommendations in the Guide for the Care and Use of Laboratory Animals (National Research Council et al., 2010) and was approved on April 30, 2014, by the Institutional Animal Care and Use Committee (IACUC) at Loma Linda University (Protocol number 8130028).

2.5.2. Experimental Conditions

Following acclimatization, animals were housed one per cage and assigned to one of four groups: (1) control (n = 3); (2) hindlimb unloaded (HLU) (n = 6); (3) low-dose irradiated (LDR) (n = 6); and (4) combination hindlimb unloaded and low-dose irradiated (HLU+LDR) (n = 4) for 21 days. Post-experimental exposure animals were group housed (3 per cage) for up to 4 months (4M).

2.5.3. Hindlimb Unloading and Whole-Body Irradiation

Hindlimb-unloaded animals (HLU) were treated as previously described (Morey-Holton et al., 2005), with animals maintained at a 35–40 head-down tilt with the hindlimbs elevated above the bottom of the cage. LDR gamma irradiation was performed by placing 57Co plates (185 MBq activity; GPI, Stoughton, WI) under the cages (1 plate per 2 cages). These cobalt plates deliver chronic low-dose/low-dose-rate γ-ray irradiation. A total dose of 0.04 Gy was delivered at a rate of 0.01 cGy/h over 21 days. HLU+LDR animals were both unloaded and irradiated as described here. Control animals received no treatment. All animals were handled similarly.

2.5.4. Euthanasia, Dissection, Tissue Storage, and Nucleic Acid Extraction

Mice were euthanized with 100% CO2 followed by immediate exsanguination by cardiac puncture. Immediately after euthanasia brains were removed and bisected along the midline and coronally within the half hemispheres, with the brainstem and cerebellum removed. The right caudal half hemisphere of the brain (containing mid- and hindbrain) from each mouse was placed in a sterile cryovial, snap frozen in liquid nitrogen and kept at −80 °C. AllPrep DNA/RNA/miRNA Universal Kit (Qiagen, Hilden, Germany) was used to extract RNA and DNA according to the manufacturer's instructions. Tissue homogenization was performed using the Tissue homogenizing CKMix (Bertin Instruments, Montigny-le-Bretonneux, France) on a Minilys homogenizer (Bertin Instruments). RNA and DNA concentrations were measured using a Qubit 3.0 Fluorometer (Thermo Fisher Scientific, Waltham, MA, USA) and stored at −80 °C.

2.5.5. RNA Sequencing

The Ovation Mouse RNA-Seq System 1–96 (NuGEN Technologies, Redwood City, CA, USA) was used per manufacturer's instructions to construct RNA-Seq libraries. A total of 100 ng of total RNA was used as input. First and second strands of cDNA were synthesized from total RNA spiked with 1 µL of 1:500 diluted ERCC ExFold RNA Spike-In Mix 1 (Life Technologies, Carlsbad, CA, USA) at the appropriate ratio. Following primer annealing and cDNA synthesis, the products were sheared using Covaris S220 Focused-ultrasonicator (Covaris Inc., Woburn, MA, USA) to obtain fragment sizes between 150–200 bp. This was followed by end-repair, adaptor index ligation and strand selection. Barcodes with unique indices out of 96 indices, were used per sample for multiplexing. Strand selection was performed using a custom InDA-C primer mixture SS5 Version5 for mice (NuGEN Technologies). Libraries were amplified by PCR with 17 cycles on a Mastercycler Pro (Eppendorf) and purified with RNAClean XP Agencourt beads (Beckman Coulter, Pasadena, CA, USA). These libraries were sequenced on a HiSeq 4000 (Illumina, Mira Loma, CA, USA) to generate 15–30 M 75-bp single end reads per sample. Raw data are available at NASA GeneLab (genelab.nasa.gov, accession GLDS-202) ("GLDS-202: Low Dose (0.04 Gy) Irradiation (LDR) and Hindlimb Unloading (HLU) Microgravity in Mice: Brain Transcriptomic and Epigenomic Data" n.d.).

2.5.6. Reduced Representation Bisulfite Sequencing (RRBS)

The RRBS library was constructed with 100 ng of gDNA as input using Ovation RRBS Methyl-Seq System (NuGEN Technologies) according to the manufacturer's protocol. Briefly, MspI enzyme, which cuts the DNA at CCGG sites, was used to digest gDNA into fragments. The fragments were directly subject to end blunting and phosphorylation followed by ligation to a methylated adapter with a single-base T overhang. The ligation products were finally repaired

in a thermal cycler under the program (60 °C for 10 min, 70 °C for 10 min, held at 4 °C). The product of the final repair reaction was used for bisulfite conversion using the EpiTect Fast DNA Bisulfite Kit (Qiagen). Bisulfite-converted DNA was then amplified on a Mastercycler Pro (Eppendorf, Hamburg, Germany) and bead-purified with Agencourt RNAClean XP Beads (Beckman Coulter). The RRBS libraries were sequenced on Illumina HiSeq 4000 at Loma Linda Center for Genomics to generate 15–55 M 70-bp single end reads per sample. Raw data are available at NASA GeneLab (genelab.nasa.gov, accession GLDS-202) ("GLDS-202: Low Dose (0.04 Gy) Irradiation (LDR) and Hindlimb Unloading (HLU) Microgravity in Mice: Brain Transcriptomic and Epigenomic Data" n.d.).

2.5.7. Differential Expression and Pathway Enrichment Analysis

Sequencing reads were first trimmed with Cutadapt (Martin 2011) using the Trim Galore! wrapper. Reads were mapped with STAR (Dobin et al., 2013), expression quantified with RSEM (B. Li and Dewey 2011) and tximport (Soneson, Love, and Robinson, n.d.) was used to import counts into DESeq2 (Love, Huber, and Anders 2014) for the detection of differentially expressed genes. Log2 fold-change and adjusted p-value cutoffs of 0.263 and 0.05 were used, respectively. Enriched GO terms were detected with WebGestalt (Liao et al., 2019) using the biological process database. Enriched pathways were detected with Reactome version 67 (Jassal et al., 2020) using an overrepresentation analysis and the full list of DEGs. Human pathways were found using Reactome's "Project to human" option. Heatmaps were generated using the webtool Morpheus, from the Broad Institute (https://software.broadinstitute.org/morpheus/) by performing hierarchical clustering using Euclidean distance. The volcano plots in Figure 2 were generated with the R package EnhancedVolcano. The UpSet diagram in Figure 4 was generated using Intervene (Khan and Mathelier 2017).

2.5.8. Differential Methylation Analysis

Sequencing reads were first trimmed with Cutadapt (Martin 2011) using the Trim Galore! wrapper using the "--RRBS" flag before processing with the "bismark" and "bismark_methylation_extractor" scripts from Bismark (Krueger and Andrews 2011). The differential methylation of gene bodies and promoters was performed with RnBeads version 1.10.8 and R version 3.4.1 (R Foundation, Boston, MA, USA) (Assenov et al., 2014). The combined ranking score was utilized to detect differential methylation. It combines absolute and relative effect sizes as well as p-values from statistical modeling into a single score.

2.6 Tables

Table 1. Differentially methylated promoters (DMPs) and differentially expressed genes (DEGs) at 4 months post-exposure conditions.

	DMP			DEG		
	Total DMP	**Hypomethylated DMP**	**Hypermethylated DMP**	**TotalDEG**	**Upregulated DEG**	**Downregulated DEG**
Radiation vs. Control	*137 promoters*	*118 promoters*	*19 promoters*	*Gm28900*	*Gm28900*	*Rps13*
				Gm43809	*Gm43809*	*Gm10357*
				Gm49345	*Gm49345*	
				Rps13		
				Gm10357		
Unloading vs. Control	*71 promoters*	*62 promoters*	*Gm35801*	*Ephb3*	*Ephb3*	*Gm12918*
			Gm13066	*Kcnc2*	*Kcnc2*	*Gm4149*
			Cnpy1	*Cdh11*	*Cdh11*	*Hist1h2bc*
			Cbfa2t2-ps1	*Gal3st3*	*Gal3st3*	*Htr1b*
			Gm15700	*Gm12918*		*Sh2d5*
			Gm27679	*Gm4149*		*Rps13*
			Hbq1a	*Hist1h2bc*		*Rpl36a-ps2*
			1700019N19Rik	*Htr1b*		
			Gm6024	*Sh2d5*		
				Rps13		
				Rpl36a-ps2		
Combined vs. Control	*170 promoters*	*119 promoters*	*51 promoters*	*270 genes*	*132 genes*	*138 genes*

Table 2. Gene list of overlapping differentially expressed genes (DEGs) and differentially methylated promoters (DMPs) within each post-exposure condition.

DEG/DEG Overlap	
Condition	**Gene**
Down HLU+LDR/Down HLU	*Gm12918*
Down HLU/Down LDR	*Rps13*
DMP/DMP Overlap	
Condition	**Methylated promoter of gene**
Hypo LDR/Hypo HLU	*Gm38049; Zfp648; 1700018C11Rik;*
	B230104I21Rik; Cald1; Lmod3
	Smco3; Angpt2; Gm28515
	Sln; Gm27159; Gm15271
	4933405E24Rik; Ndel1; Gm11373
	Gm5083; Cdhr2; Mir8097
	1700011A15Rik; Krt8; Mir8094
	Pcdh12; Tmem180
Hypo LDR/Hypo HLU+LDR	*Gm16564; Gm37267; Adamtsl2*
	Gm13413; Bbox1; Gm11805
	1700071G01Rik; Prr33; Cdhr4
	Slfn4; Krt19; Gm6401
	Gm15516; Mx2; Gm10268
	Gm25301; Armcx2; Gm15726
	Gm15247
Hypo HLU/Hypo HLU+LDR	*Gm29100; Nav1; Gm29282*
	Cfi; Mypopos; Adgb
	Gm26535; Abi3; Gm6969
	Lpar6
Hyper LDR/Hyper HLU+LDR	*Otos; Sys1; Klhl34*
	Crh; Pcdhb12
Hyper HLU/Hyper HLU+LDR	*Gm35801*
Hypo LDR/Hyper HLU+LDR	*Gm6987*
Hyper HLU/Hyper LDR	*1700019N19Rik; Cnpy1; Gm6024*
	Gm15700; Gm27679
Hypo LDR/Hypo HLU/Hypo HLU+LDR	*Trp53inp2; Gm22518; Dio1*
	Mir704; Gm805; Gm24998
	Gm15648
Hyper LDR/Hyper HLU/Hyper HLU+LDR	*Cbfa2t2-ps1*
DEG/DMP Overlap	
Condition	**Gene or Methylated promoter of gene**
Down HLU+LDR/Hyper HLU+LDR	*Gm7120*
Down HLU+LDR/Hypo HLU+LDR	*Ephx4*
Down HLU+LDR/Hypo HLU	*Blnk*
Up HLU+LDR/Hyper LDR	*Rgag4*

Table 3. Combination of LDR and HLU versus controls DEG and Reactome pathway IDs in mice that translated to human orthologs.

Pathway Name	Associated Mouse Genes	Mouse Pathway ID	Human Pathway ID
Neuronal System	*Gria3, Cacnb3, Hcn1, Camkk2, Kcnh1, Kcnma1, Kcnn1, Stx1a, Kcnv1, Kcnh5, Grin2a, Slc17a7, Dlgap1, Homer1, Snap25, Nrgn, Kcnh7, Slitrk1*	R-MMU-112316	R-HSA-112316
Potassium Channels	*Kcnn1, Kcnh5, Kcnma1, Hcn1, Kcnh1, Kcnv1, Kcnh7*	R-MMU-1296071	R-HSA-1296071
Voltage Gated Potassium Channels	*Kcnv1, Kcnh1, Kcnh5, Kcnh7*	R-MMU-1296072	R-HSA-1296072
cGMP effects	*Pde1a*	R-MMU-418457	R-HSA-418457
p75 NTR receptor-mediated signaling	*Rtn4r, Obscn, Lingo1, Kalrn*	R-MMU-193704	R-HSA-193704
Nitric oxide stimulates guanylate cyclase	*Pde1a*	R-MMU-392154	R-HSA-392154

2.7 Figures

Figure 1

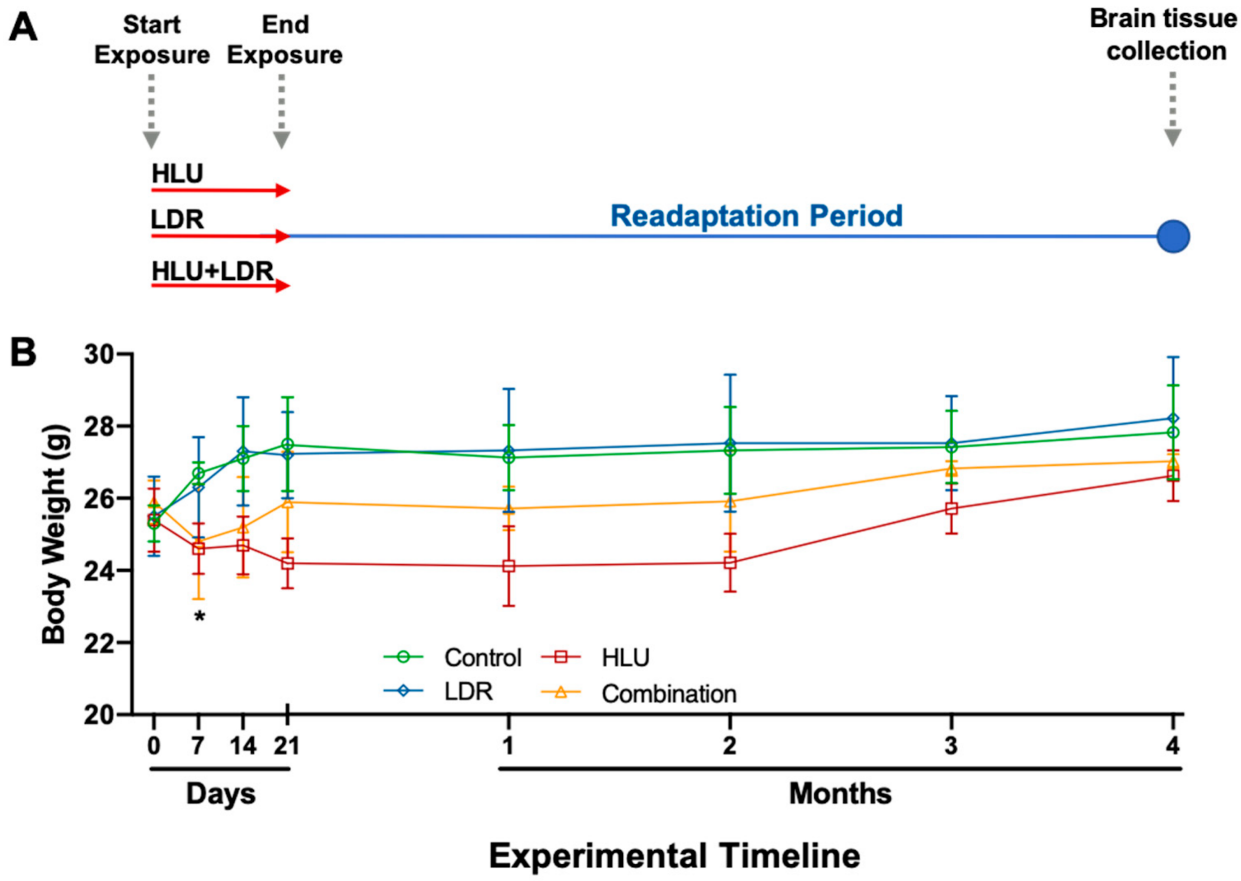


Figure 1. Experimental timeline and body weights during- and post-hindlimb unloading (HLU), low-dose irradiation (LDR) or combined HLU+LDR. **(A)** Whole brains were isolated from

C57BL/6J at 4 months post-experimental conditions of hindlimb unloading (HLU, 21 days), low-dose irradiation (LDR, 0.04 Gy for 21 days) or combined LDR and HLU timeline. **(B)** Body weights were monitored throughout the experiment (21 days) and during readaptation (4 months post-experimental exposure). Baseline weights were taken on day 0, prior to the initiation of HLU and LDR exposure. An unpaired *t*-test with Welch's correction compared experimental groups with controls. A * denotes significance of HLU only compared to controls at the 7-day timepoint. Error bars denote standard error of means. Dashed gray arrows highlight start and endpoints of experimental LDR and HLU treatments, and the final time point of the experiment when tissue was collected for transcriptomic analysis.

Figure 2

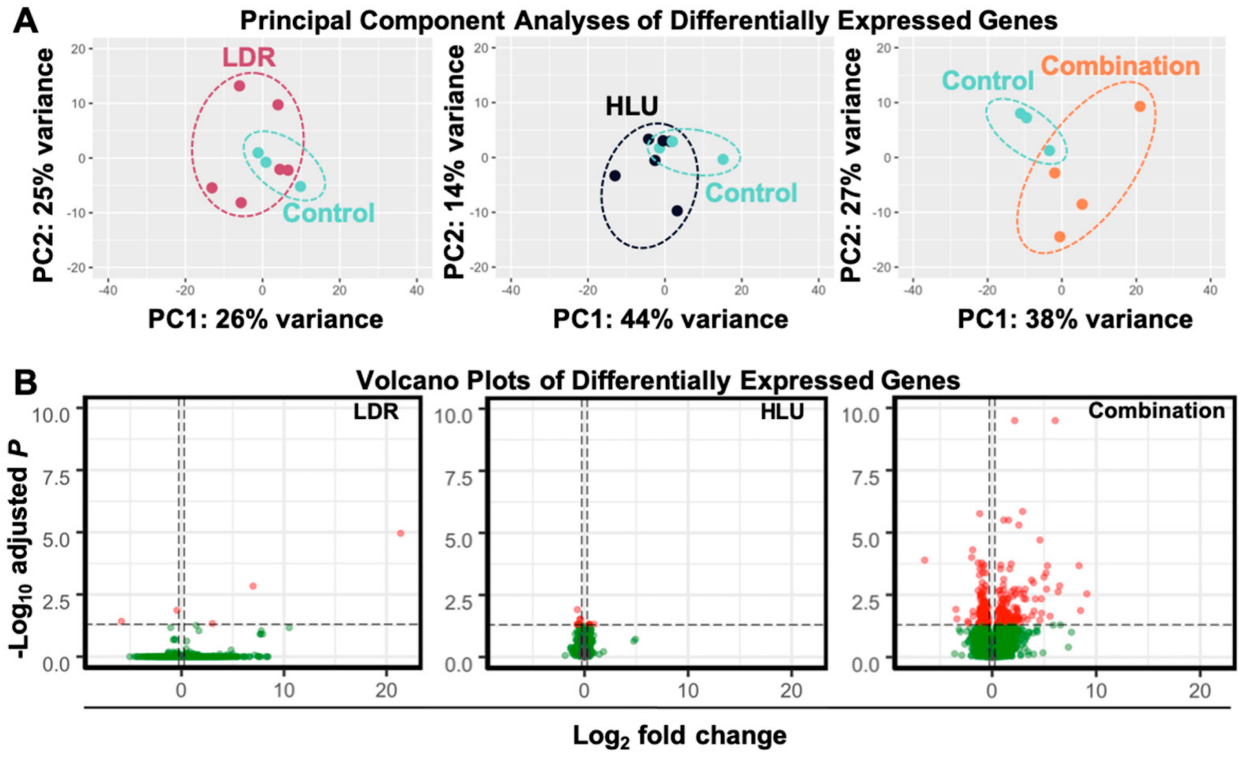
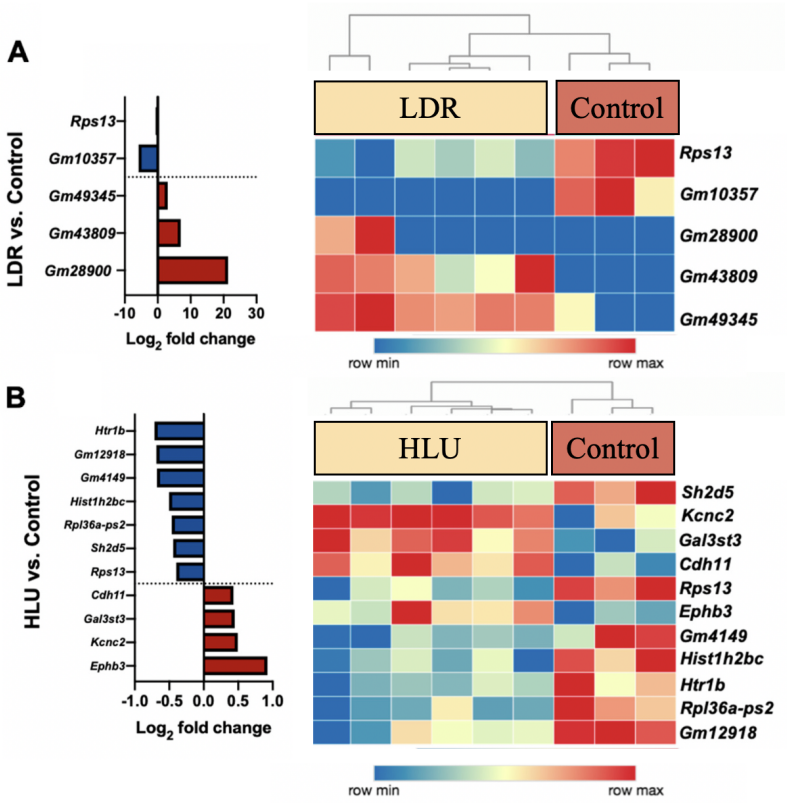


Figure 2. HLU in combination with LDR results in a heightened differentially expressed gene (DEG) profile in brains isolated at 4 months post-experimental conditions. DEG identified in brain isolated at 4 months post-LDR (0.04 Gy for 21 days), HLU (21 days) or combined HLU+LDR. **(A)** Principal component analysis (PCA) displays sample clustering within each experimental condition (n = 3–6). **(B)** Volcano plots indicate DEG profiles at 4 months post-experimental conditions. Vertical dashed lines indicate the Log2 fold-change cutoff used (0.263), while horizontal dashed lines indicate the adjusted p-value cutoff used (0.05). Red dots indicate differentially expressed genes that meet our adjusted p-value and Log2 fold-change cutoffs and compose our set of differentially expressed genes. Green dots indicate genes that do

not meet our adjusted *p*-value cutoff and, therefore, are not differentially expressed. Darker saturation of color represents overlap of genes.

Figure 3



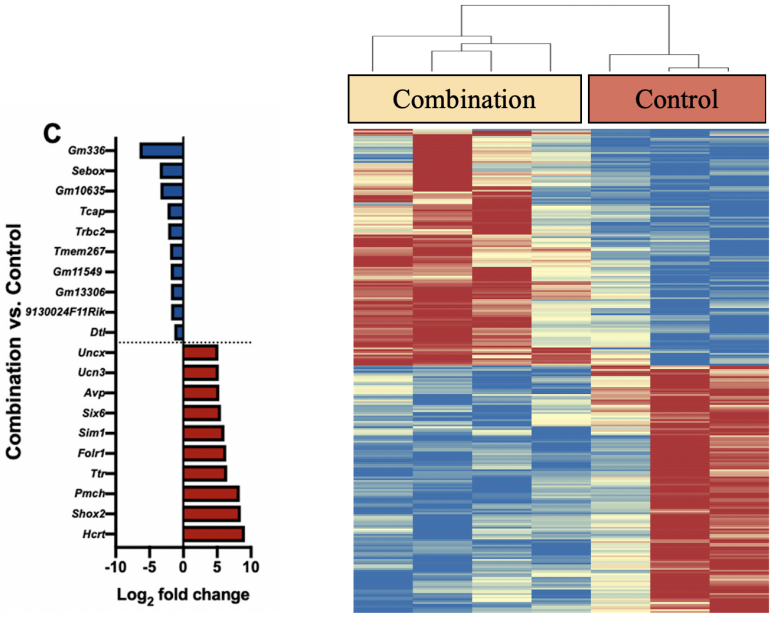


Figure 3. DEG profiles of brains isolated at 4 months post-HLU, LDR or combined HLU+LDR. **(A, left panel)** Graph depicts LDR versus control samples Log2 fold-change mean values (*n* = 3–6); **(A, right panel)** Heatmap displays individual DEG. **(B, left panel)** Graph depicts HLU versus control samples Log2 fold-change mean values (*n* = 3–6); **(B, right panel)** Heatmap displays individual DEG. **(C, left panel)** Graph depicts a subset of the largest Log2 fold-change gene set within combined HLU+LDR (combination) verses control samples Log2 fold-change mean values (*n* = 3–4); **(C, right panel)** Heatmap displays individual DEG.

Figure 4

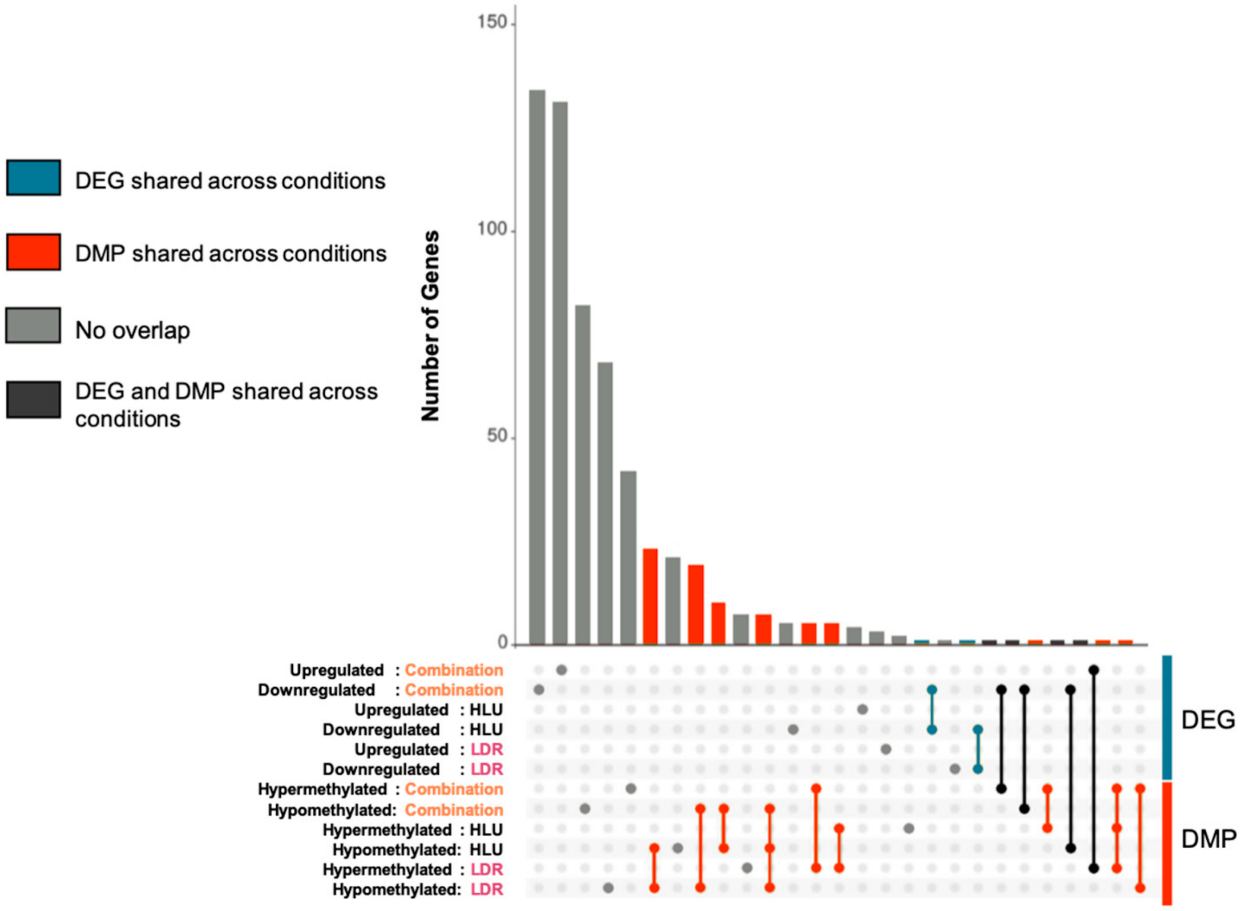


Figure 4. Minimal brain differential gene expression and promoter methylation overlap across different simulated spaceflight experimental conditions. Gene lists of DEGs and the genes corresponding to the differentially methylated promoters (DMPs) were created for each condition: LDR, HLU, or HLU+LDR combined versus controls. The dotted region depicts all conditions, and the connecting lines depict DEG/DMP overlap between experimental conditions. The vertical bars indicate the size of that overlap. A single, grey dot indicates the number of DEG/DMP that do not overlap with any other DEG/DMP. The red dots indicate DMP overlap across different conditions. The blue dots indicate DEG overlap across different conditions. The black dots indicate DEG and DMP overlap. "Upregulated/downregulated" denotes gene expression of DEG and "hypermethylated/hypomethylated" denotes methylation patterns of DMPs.

Figure 5

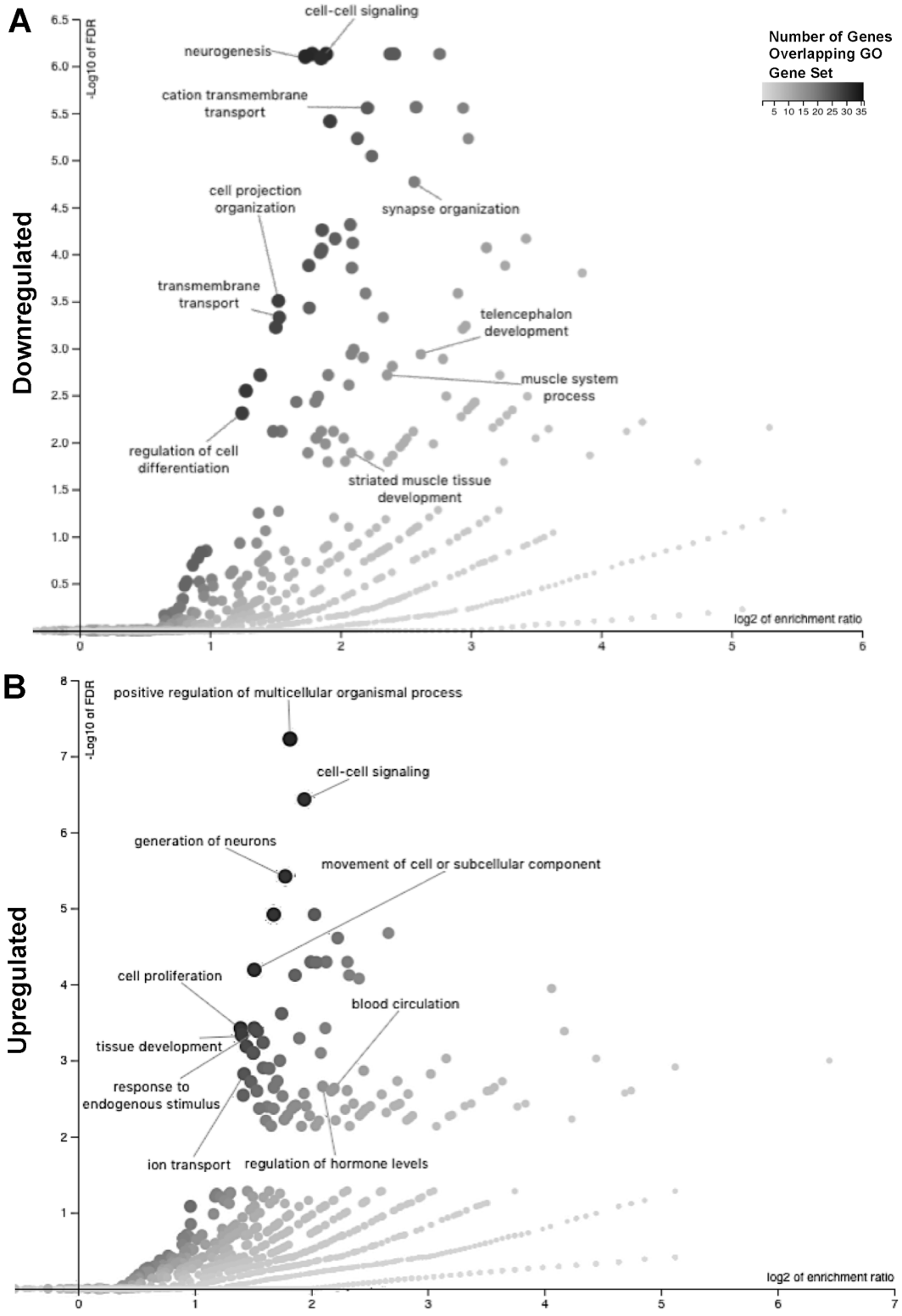
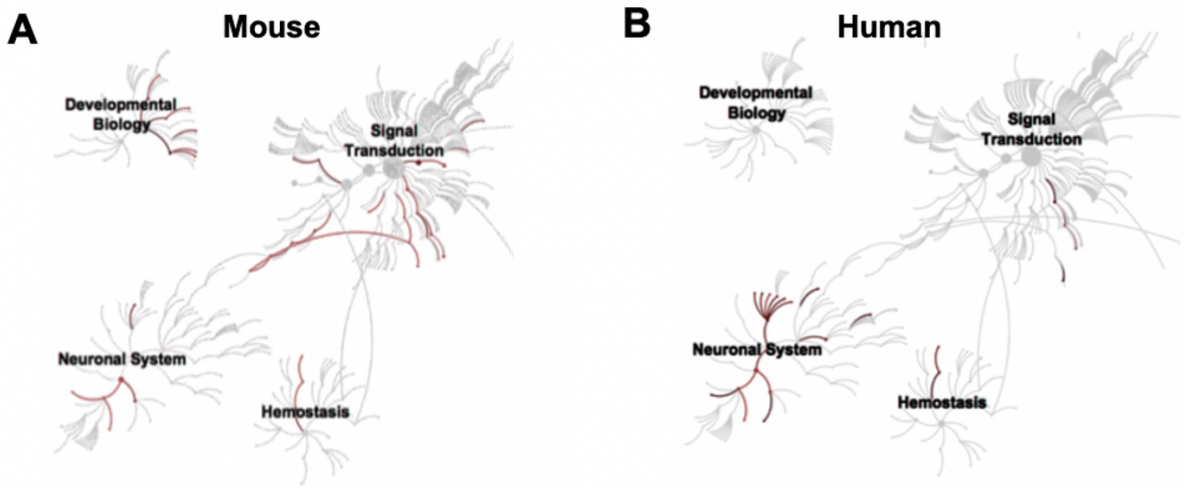

A
Downregulated
-Log10 of FDR
log2 of enrichment ratio
Number of Genes Overlapping GO Gene Set
5 10 15 20 25 30 35
cell-cell signaling
neurogenesis
cation transmembrane transport
cell projection organization
synapse organization
transmembrane transport
telencephalon development
muscle system process
regulation of cell differentiation
striated muscle tissue development
B
Upregulated
-Log10 of FDR
log2 of enrichment ratio
positive regulation of multicellular organismal process
cell-cell signaling
generation of neurons
movement of cell or subcellular component
cell proliferation
blood circulation
tissue development
response to endogenous stimulus
ion transport
regulation of hormone levels

Figure 5. Altered biological processes and pathways identified at 4 months post-combined HLU+LDR exposure. **(A,B)** WebGestalt graphs depict enriched gene ontology (GO) terms for downregulated **(A)** and upregulated **(B)** DEGs. The x-axis measures the enrichment ratio for each GO term on a Log2 scale. The y-axis measures -Log10 (false discovery rate, FDR). The size and color tone of the dot is proportional to the size of the category. The displayed categories are from WebGestalt's "Weighted set cover" option, which displays the GO categories that minimize redundancy.

2.8 Supplemental Figures

Sup Figure 1



Supplemental Figure 1. Abridged Reactome pathway analyses. (A) Mouse Reactome pathway analysis for DEG. **(B)** Human Reactome pathway analysis for DEG. Reactome pathways and subpathways are displayed hierarchically. Red lines indicate aspects that are enriched in DEG; Mouse DEG were converted to their human orthologs before pathway analysis using Reactome's "Project to human" option.

2.9 Author Contributions

Formal analysis, E.G.O., A.M.P., W.A.d.S., N.S., S.S., C.W. and J.M.G.; Funding acquisition, X.W.M.; Investigation, X.W.M.; Methodology, X.W.M.; Project administration,

J.M.G.; Validation, N.S.; Visualization, E.G.O., A.M.P., W.A.d.S. and N.S.; Writing—original draft, A.M.P.; Writing—review and editing, E.G.O., A.M.P., W.A.d.S., C.G.T.T., S.S.R., N.S., S.S., C.W., J.E.G.O., A.M.P., W.A.d.S.G. and X.W.M.

2.10 Data Availability

Raw data are available at NASA GeneLab (genelab.nasa.gov, accession GLDS-202).

2.11 References

1. Kovalchuk, A.; Kolb, B. Low dose radiation effects on the brain-from mechanisms and behavioral outcomes to mitigation strategies. *Cell Cycle* 2017, *16*, 1266–1270.
2. Lawley, J.S.; Petersen, L.G.; Howden, E.J.; Sarma, S.; Cornwell, W.K.; Zhang, R.; Whitworth, L.A.; Williams, M.A.; Levine, B.D. Effect of gravity and microgravity on intracranial pressure. *J. Physiol.* 2017, *595*, 2115–2127.
3. Chancellor, J.C.; Scott, G.B.I.; Sutton, J.P. Space Radiation: The Number One Risk to Astronaut Health beyond Low Earth Orbit. *Life* 2014, *4*, 491–510.
4. Seawright, J.W.; Samman, Y.; Sridharan, V.; Mao, X.W.; Cao, M.; Singh, P.; Melnyk, S.; Koturbash, I.; Nelson, G.A.; Hauer-Jensen, M.; et al. Effects of low-dose rate γ-irradiation combined with simulated microgravity on markers of oxidative stress, DNA methylation potential, and remodeling in the mouse heart. *PloS ONE* 2017, *12*, e0180594.
5. De la Torre, G.G. Cognitive neuroscience in space. *Life* 2014, *4*, 281–294.
6. Garrett-Bakelman, F.E.; Darshi, M.; Green, S.J.; Gur, R.C.; Lin, L.; Macias, B.R.; McKenna, M.J.; Meydan, C.; Mishra, T.; Nasrini, J.; et al. The NASA Twins Study: A multidimensional analysis of a year-long human spaceflight. *Science* 2019, *364*, eaau8650.
7. Lee, J.K.; Koppelmans, V.; Riascos, R.F.; Hasan, K.M.; Pasternak, O.; Mulavara, A.P.; Bloomberg, J.J.; Seidler, R.D. Spaceflight-Associated Brain White Matter Microstructural Changes and Intracranial Fluid Redistribution. *JAMA Neurol.* 2019, *76*, 412–419.
8. Ronca, A.E.; Moyer, E.L.; Talyansky, Y.; Lowe, M.; Padmanabhan, S.; Choi, S.; Gong, C.; Cadena, S.M.; Stodieck, L.; Globus, R.K. Behavior of mice aboard the International Space Station. *Sci. Rep.* 2019, *9*, 4717.
9. Bellone, J.A.; Gifford, P.S.; Nishiyama, N.C.; Hartman, R.E.; Mao, X.W. Long-term effects of simulated microgravity and/or chronic exposure to low-dose gamma radiation on behavior and blood-brain barrier integrity. *NPJ Microgravity* 2016, *2*, 16019.
10. Mao, X.W.; Nishiyama, N.C.; Pecaut, M.J.; Campbell-Beachler, M.; Gifford, P.; Haynes, K.E.; Becronis, C.; Gridley, D.S. Simulated Microgravity and Low-Dose/Low-Dose-Rate Radiation Induces Oxidative Damage in the Mouse Brain. *Radiat. Res.* 2016, *185*, 647–657.
11. Mao, X.W.; Nishiyama, N.C.; Campbell-Beachler, M.; Gifford, P.; Haynes, K.E.; Gridley, D.S.; Pecaut, M.J. Role of NADPH Oxidase as a Mediator of Oxidative Damage in Low-Dose Irradiated and Hindlimb-Unloaded Mice. *Radiat. Res.* 2017, *188*, 392–399.

12. Smart, D. Radiation Toxicity in the Central Nervous System: Mechanisms and Strategies for Injury Reduction. *Semin. Radiat. Oncol.* 2017, *27*, 332–339.
13. Cucinotta, F.A.; Cacao, E. Risks of Cognitive Detriments after Low Dose Heavy Ion and Proton Exposures. *Int. J. Radiat. Biol.* 2019, 1–36.
14. Mao, X.W. Low dose (0.04 Gy) irradiation (LDR) and hindlimb unloading (HLU) microgravity in mice: Brain transcriptomic and epigenomic data. In *NASA GeneLab*; 2018. Available online: https://genelab-data.ndc.nasa.gov/genelab/accession/GLDS-202 (accessed on 18 July 2019).
15. Chatterjee, S.; Pietrofesa, R.A.; Park, K.; Tao, J.-Q.; Carabe-Fernandez, A.; Berman, A.T.; Koumenis, C.; Sielecki, T.; Christofidou-Solomidou, M. LGM2605 Reduces Space Radiation-Induced NLRP3 Inflammasome Activation and Damage in In Vitro Lung Vascular Networks. *Int. J. Mol. Sci.* 2019, *20*, 176.
16. Khan, A.; Mathelier, A. Intervene: A tool for intersection and visualization of multiple gene or genomic region sets. *BMC Bioinformatics* 2017, *18*, 287.
17. Mizumatsu, S.; Monje, M.L.; Morhardt, D.R.; Rola, R.; Palmer, T.D.; Fike, J.R. Extreme sensitivity of adult neurogenesis to low doses of X-irradiation. *Cancer Res.* 2003, *63*, 4021–4027.
18. Najjar, S.; Pearlman, D.M.; Devinsky, O.; Najjar, A.; Zagzag, D. Neurovascular unit dysfunction with blood-brain barrier hyperpermeability contributes to major depressive disorder: A review of clinical and experimental evidence. *J. Neuroinflammation* 2013, *10*, 142.
19. Song, T.-T.; Bi, Y.-H.; Gao, Y.-Q.; Huang, R.; Hao, K.; Xu, G.; Tang, J.-W.; Ma, Z.-Q.; Kong, F.-P.; Coote, J.H.; et al. Systemic pro-inflammatory response facilitates the development of cerebral edema during short hypoxia. *J. Neuroinflammation* 2016, *13*, 63.
20. Li, L.; Zhang, H.; Varrin-Doyer, M.; Zamvil, S.S.; Verkman, A.S. Proinflammatory role of aquaporin-4 in autoimmune neuroinflammation. *FASEB J. Off. Publ. Fed. Am. Soc. Exp. Biol.* 2011, *25*, 1556–1566.
21. Ikeshima-Kataoka, H. Neuroimmunological Implications of AQP4 in Astrocytes. *Int. J. Mol. Sci.* 2016, *17*, 1306.
22. Meli, R.; Pirozzi, C.; Pelagalli, A. New Perspectives on the Potential Role of Aquaporins (AQPs) in the Physiology of Inflammation. *Front. Physiol.* 2018, *9*, 101.
23. Kotas, M.E.; Medzhitov, R. Homeostasis, inflammation, and disease susceptibility. *Cell* 2015, *160*, 816–827.
24. Medzhitov, R. Origin and physiological roles of inflammation. *Nature* 2008, *454*, 428–435.
25. Lowe, X.R.; Bhattacharya, S.; Marchetti, F.; Wyrobek, A.J. Early brain response to low-dose radiation exposure involves molecular networks and pathways associated with cognitive functions, advanced aging and Alzheimer's disease. *Radiat. Res.* 2009, *171*, 53–65.
26. Paul, A.M.; Acharya, D.; Duty, L.; Thompson, E.A.; Le, L.; Stokic, D.S.; Leis, A.A.; Bai, F. Osteopontin facilitates West Nile virus neuroinvasion via neutrophil "Trojan horse" transport. *Sci. Rep.* 2017, *7*, 4722.
27. Yamazaki, Y.; Shinohara, M.; Shinohara, M.; Yamazaki, A.; Murray, M.E.; Liesinger, A.M.; Heckman, M.G.; Lesser, E.R.; Parisi, J.E.; Petersen, R.C.; et al. Selective loss of cortical endothelial tight junction proteins during Alzheimer's disease progression. *Brain, J. Neurol.* 2019, *142*, 1077–1092.
28. Zhao, C.; Deng, W.; Gage, F.H. Mechanisms and functional implications of adult neurogenesis. *Cell* 2008, *132*, 645–660.
29. Zhang, J.; Jiao, J. Molecular Biomarkers for Embryonic and Adult Neural Stem Cell and Neurogenesis. *BioMed Res. Int.* 2015, *2015*, 727542.

30. Sun, X.Z.; Inouye, M.; Fukui, Y.; Hisano, S.; Sawada, K.; Muramatsu, H.; Muramatsu, T. An immunohistochemical study of radial glial cells in the mouse brain prenatally exposed to gamma-irradiation. *J. Neuropathol. Exp. Neurol.* 1997, *56*, 1339–1348.
31. Kleindienst, A.; McGinn, M.J.; Harvey, H.B.; Colello, R.J.; Hamm, R.J.; Bullock, M.R. Enhanced hippocampal neurogenesis by intraventricular S100B infusion is associated with improved cognitive recovery after traumatic brain injury. *J. Neurotrauma* 2005, *22*, 645–655.
32. Sun, D.; Bullock, M.R.; McGinn, M.J.; Zhou, Z.; Altememi, N.; Hagood, S.; Hamm, R.; Colello, R.J. Basic fibroblast growth factor-enhanced neurogenesis contributes to cognitive recovery in rats following traumatic brain injury. *Exp. Neurol.* 2009, *216*, 56–65.
33. Jandial, R.; Hoshide, R.; Waters, J.D.; Limoli, C.L. Space-brain: The negative effects of space exposure on the central nervous system. *Surg. Neurol. Int.* 2018, *9*, 9.
34. Grammaticos, P.; Giannoula, E.; Fountos, G.P. Acute radiation syndrome and chronic radiation syndrome. *Hell. J. Nucl. Med.* 2013, *16*, 56–59.
35. Walker, S.A.; Townsend, L.W.; Norbury, J.W. Heavy ion contributions to organ dose equivalent for the 1977 galactic cosmic ray spectrum. *Adv. Space Res.* 2013, *51*, 1792–1799.
36. *Guide for the Care and Use of Laboratory Animals*, 8th ed.; National Research Council (U.S.); Institute for Laboratory Animal Research (U.S.); National Academies Press (U.S.) (Eds.) National Academies Press: Washington, DC, USA, 2011; ISBN 978-0-309-15400-0.
37. Morey-Holton, E.; Globus, R.K.; Kaplansky, A.; Durnova, G. The hindlimb unloading rat model: Literature overview, technique update and comparison with space flight data. *Adv. Space Biol. Med.* 2005, *10*, 7–40.
38. Martin, M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal* 2011, *17*, 10.
39. Dobin, A.; Davis, C.A.; Schlesinger, F.; Drenkow, J.; Zaleski, C.; Jha, S.; Batut, P.; Chaisson, M.; Gingeras, T.R. STAR: Ultrafast universal RNA-seq aligner. *Bioinformatics* 2013, *29*, 15–21.
40. Li, B.; Dewey, C.N. RSEM: Accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics* 2011, *12*, 323.
41. Soneson, C.; Love, M.I.; Robinson, M.D. Differential analyses for RNA-seq: Transcript-level estimates improve gene-level inferences. *F1000Research* 2016, *4*, 1521.
42. Love, M.I.; Huber, W.; Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* 2014, *15*, 550.
43. Liao, Y.; Wang, J.; Jaehnig, E.J.; Shi, Z.; Zhang, B. WebGestalt 2019: Gene set analysis toolkit with revamped UIs and APIs. *Nucleic Acids Res.* 2019, *47*, W199–W205.
44. Fabregat, A.; Jupe, S.; Matthews, L.; Sidiropoulos, K.; Gillespie, M.; Garapati, P.; Haw, R.; Jassal, B.; Korninger, F.; May, B.; et al. The Reactome Pathway Knowledgebase. *Nucleic Acids Res.* 2018, *46*, D649–D655.
45. Krueger, F.; Andrews, S.R. Bismark: A flexible aligner and methylation caller for Bisulfite-Seq applications. *Bioinformatics* 2011, *27*, 1571–1572.
46. Assenov, Y.; Müller, F.; Lutsik, P.; Walter, J.; Lengauer, T.; Bock, C. Comprehensive analysis of DNA methylation data with RnBeads. *Nat. Methods* 2014, *11*, 1138–1140.

Chapter 3: RNA-Sequencing of the Murine Retina Post-Spaceflight Reveals Similarities to Retinitis Pigmentosa

Chapter 3 is adapted with minimal modification from:

Spaceflight influences gene expression, photoreceptor integrity, and oxidative stress-related damage in the murine retina. Eliah G. Overbey, Willian Abraham da Silveira, *et al*. Scientific Reports. 2019.

3.1 Abstract

Extended spaceflight has been shown to adversely affect astronaut visual acuity. The purpose of this study was to determine whether spaceflight alters gene expression profiles and induces oxidative damage in the retina. Ten week old adult C57BL/6 male mice were flown aboard the ISS for 35 days and returned to Earth alive. Ground control mice were maintained on Earth under identical environmental conditions. Within 38 (+/−4) hours after splashdown, mice ocular tissues were collected for analysis. RNA sequencing detected 600 differentially expressed genes (DEGs) in murine spaceflight retinas, which were enriched for genes related to visual perception, the phototransduction pathway, and numerous retina and photoreceptor phenotype categories. Twelve DEGs were associated with retinitis pigmentosa, characterized by dystrophy of the photoreceptor layer rods and cones. Differentially expressed transcription factors indicated changes in chromatin structure, offering clues to the observed phenotypic changes. Immunofluorescence assays showed degradation of cone photoreceptors and increased retinal oxidative stress. Total retinal, retinal pigment epithelium, and choroid layer thickness were significantly lower after spaceflight. These results indicate that retinal performance may decrease over extended periods of spaceflight and cause visual impairment.

Keywords: Transcriptomics, RNA sequencing, Gene expression, Experimental models of disease

3.2 Introduction

We are entering an era of renewed interest in space exploration. Although officially the space race ended decades ago, humanity is crossing the threshold into a second space age that includes new global partnerships and private companies. This new infrastructure and organization has brought the concept of long duration space missions, such as permanent moon settlements and manned missions to Mars, from a distant goal to a tangible reality (Dawson

2017). However, the space environment is fundamentally different from Earth. Microgravity and ionizing radiation compose a unique set of physiological stressors, the effects of which remain to be fully characterized (Xiao W. Mao et al., 2013; Overbey et al., 2019).

In humans, one of the most apparent physiological responses to the space environment is the redistribution of fluid throughout the body that occurs due to microgravity. This redistribution shifts fluid upward from the lower parts of the body (Nelson, Mulugeta, and Myers 2014). An increased intracranial pressure caused by this fluid shift has been attributed as a major cause of spaceflight-associated neuro-ocular syndrome (SANS) (A. G. Lee et al., 2018). 40% of astronauts have experienced SANS (Stenger et al., 2017), including the spaceflight subject of the NASA twins study (Garrett-Bakelman et al., 2019). The physiological impact of SANS encapsulates multiple components of the eye and includes optic disc edema, globe flattening (GF), choroidal and retinal folds, hyperopic refractive error shifts, and nerve fiber layer infarcts (i.e., cotton wool spots) (A. G. Lee et al., 2018; Mader et al., 2011). While the physiological characteristics of SANS are well documented, the mechanisms that drive SANS are still poorly understood.

To date, despite evidence showing ocular functioning impairment following exposure to the space environment, few studies have investigated molecular mechanisms involved in the spaceflight-associated changes in retinal structure and function. A previous study from our group characterized the effects of a 13-day exposure to the space environment aboard the Space Shuttle Atlantis (STS-135) in mice. This study found that spaceflight conditions induced oxidative damage that resulted in mitochondrial apoptosis in the mouse retina (Xiao W. Mao et al., 2013). Subsequent work with low-dose proton radiation and a simulated microgravity mice model showed a major role of simulated space radiation in the increase of oxidative stress, with synergistic effects when simulated microgravity was applied (X. W. Mao, Boerma, Rodriguez, Campbell-Beachler, Jones, Stanbouly, Sridharan, Nishiyama, et al., 2018; X. W. Mao, Boerma, Rodriguez, Campbell-Beachler, Jones, Stanbouly, Sridharan, Wroe, et al., 2018). In our current study, mice were exposed to the space environment for 35 days aboard the International Space Station (ISS) to determine whether the space environment induces oxidative damage on ocular structure and to characterize gene expression profiles of mouse retina exposed to spaceflight.

3.3 Results

3.3.1 The space environment causes changes in gene expression related to visual and RNA regulation pathways

The space environment caused a wide array of changes in gene expression, the most prominent were changes in ocular pathways and RNA processing. Twenty mice spent 35 days on the International Space Station as part of NASA's ninth Rodent Research experiment (RR-9). An additional twenty mice were housed on Earth in the same housing hardware used in flight for the

ground control group. RNA sequencing was performed on sixteen mice (n = 8/group). 600 differentially expressed genes (DEGs) were identified between the spaceflight and ground control groups using DESeq2 (Love, Huber, and Anders 2014) with an adjusted p-value threshold of 0.01. Hierarchical clustering was performed on the DEGs according to their expression values to determine which mice have similar DEG expression profiles. Mice clustered according to their inclusion in the ground control or spaceflight group. The spaceflight group had 286 upregulated genes and 314 downregulated genes compared to the ground control (Fig 1A).

An overrepresentation analysis (ORA) of gene ontology (GO), pathway, and phenotypic categories was performed on the DEGs using WebGestalt (Wang et al., 2017). From this analysis, 46 GO, 5 pathway, and 14 phenotypic categories were found with a false discovery rate (FDR) of less than 0.05 (Fig 1B–D). Included in the overrepresented categories are categories related to ocular function (GO categories: 'visual perception', 'response to light stimulus', 'sensory perception of light stimulus', 'retina development in camera-type eye'; Pathways: 'the phototransduction cascade' 'inactivation, recovery and regulation of the phototransduction cascade'; Phenotype: electrophysiology, morphology, and degeneration of the retina, rods, and cones), various RNA processing, splicing, and metabolism functions (GO categories: 'RNA processing', 'rRNA processing', 'mRNA processing', 'RNA splicing', via transesterification reactions', 'mRNA splicing, via spliceosome', 'rRNA metabolic process', 'ncRNA metabolic process', 'mRNA metabolic process', 'RNA transport'), and direct responses to the physical pressures of spaceflight (GO categories: 'response to abiotic stimulus', 'response to radiation', 'cellular response to stress').

Some genes were highly differentially expressed, but were not included in any of the GO category, pathway, or phenotype gene lists from the ORA (Sup Fig 1). *Drd4* was the most significantly differentially expressed gene and was upregulated in the spaceflight group (adjusted p-value: 4.31E-51; log2 fold-change: 0.812). *Drd4* is a dopamine receptor that control circadian rhythm in the mammalian retina (Jackson et al., 2011). Similarly, *Hist1h2bc* was not found in any categories of the ORA, but was upregulated in the spaceflight group (adjusted p-value: 1.66E-09; log2 fold-change: 0.526). *Hist1h2bc* has been previously shown to be upregulated in the aging retina (Banday et al., 2014). Together, these genes support previous studies suggesting that spaceflight disrupts circadian rhythms and is a potential model for aging (Banday et al., 2014; Demontis et al., 2017).

3.3.2 Genes associated with retinitis pigmentosa are differentially expressed in the space environment

We sought to determine whether any of the DEGs from the spaceflight samples were also differentially expressed in common retinal diseases. We compiled a list of disease-associated genes for the following retinal diseases: retinitis pigmentosa, diabetic retinopathy, age-related

macular degeneration, and retinal detachment. Human disease-associated gene lists were found using the DisGeNET database (Piñero et al., 2017). Genes in the DisGeNET database were filtered by their gene-disease association score (GDA). Only genes with a GDA greater than 0.2 were included in the analysis. This included disease-associated genes from expertly curated and animal model databases, but excluded disease-associated genes from inferences and text-mining databases. Disease-associated genes were converted to their mouse ortholog using the Mouse Genome Informatics database (Smith et al., 2018). Genes that did not pass the DESeq2 threshold for the number of mean counts (i.e., they had an adjusted p-value equal to 'NA'), were filtered out from the analysis.

Present in the DisGeNet database were 75 genes associated with the disease retinitis pigmentosa, 14 genes with diabetic retinopathy, 8 genes with age-related macular degeneration, and 5 genes with retinal detachment. Most of the disease-associated genes were unique to a single retinal disease, with the exception of *Abca4*, *Nmnat1*, *Casp3*, and *Crb1*, which were each shared across two diseases (Fig 2A). None of the disease-associated genes that were shared across multiple diseases were differentially expressed in spaceflight.

Of the 75 genes associated with retinitis pigmentosa, 12 were differentially expressed during spaceflight (*Sag, BC027072/Pcare, Guca1b, Rbp3, Ahi1, Guca1a, Prpf8, Rp1, Gucy2e, Cacna2d4, Pde6b, Dhdds*), the most of any of the four retinal diseases examined. Diabetic retinopathy had one disease-associated gene that was differentially expressed during spaceflight (*Hif1a*). Age-related macular degeneration and retinal detachment disease-associated genes were not differentially expressed during spaceflight (Fig 2B).

3.3.3 Differentially expressed transcription factors hint that spaceflight changes chromatin organization

DEGs reported in Fig 1A were filtered by their PANTHER protein class in order to find differentially expressed transcription factors (DETFs) (Mi et al., 2019). Out of the 600 DEGs, 29 are DETFs. Hierarchical clustering was performed on the DETFs according to their expression values. Mice clustered according to their inclusion in the ground control or spaceflight group (Fig 3A). Eight of the DETFs are necessary for development and maintenance of the retina and other structures of the eye: *Mafg*, *Isl1*, *Atf5*, *Kdm6b*, *Hmgb3*, *Prox1*, *Casz1* and *Hif1a (Anand and Lachke 2017; Bejarano-Escobar et al., 2015; Yasuda et al., 2014; Terada and Furukawa 2010; Mattar et al., 2018; Barben et al., 2018; Pérez de Sevilla Müller et al., 2017; Iida et al., 2015)*. Of special interest are *Cazs1* and *Hif1a*. *Cazs1* is a heterochromatin regulator that acts in a splice-variant specific manner (Mattar et al., 2018) and *Hif1a* inactivation has been shown to significantly mitigate photoreceptor degeneration in a chronic hypoxia-like stress model (Barben et al., 2018).

An overrepresentation analysis (ORA) of gene ontology (GO) terms was performed on the DETFs. 47 GO categories were found with an FDR of less than 0.05. The affinity propagation filter from WebGestalt was applied in order to find a reduced number of GO terms that are representative of all GO enrichment results. We found enrichment of GO categories related to DNA transcription ('positive regulation of transcription', DNA-templated, 'negative regulation of transcription, DNA-templated'), further affirming the role of these genes as transcription factors. Additionally, there were enriched GO categories relating to chromatin organization, including regulation of histone modification, histone lysine demethylation, negative regulation of histone H3K9 trimethylation (Fig 3B). The genes in these GO categories include *Kdm4a, Kdm4b, and Kdm6b*, which are known regulators of lysine methylation. *Kdm4a* and *Kdm4b* are regulators of H3K9me2/me3 (Labbé, Holowatyj, and Yang 2013), which is a regulator of constitutive heterochromatin and an indicator for the presence of constitutive heterochromatin. *Kdm6b* is a regulator of H3K27me2/me3 (Labbé, Holowatyj, and Yang 2013) and is primarily responsible for the silencing of gene expression.

3.3.4 The space environment decreases the thickness of retinal tissue and increases oxidative stress and cone photoreceptor damage

Micro-computed tomography (MicroCT) images were generated in order to characterize global ocular morphology and to measure the thickness of the retina and surrounding tissues for ground control and spaceflight mice (Fig 4A). Total retina, retinal pigment epithelium (RPE), and choroid layers of the eye decreased significantly in thickness ($p < 0.05$) during spaceflight (Fig 4B). Additionally, choroid deformation and folds were noticed in spaceflight mice (Fig 4C).

Immunofluorescence staining with peanut agglutinin (PNA), a specific marker for cone photoreceptors, was performed in order to view cone photoreceptors in the eye. Cone photoreceptors show signs of degradation during spaceflight (Fig 5A). Our quantitative assessment (Fig 5B) revealed a strong trend decrease in density of cones in the spaceflight mice (950 counts/mm2) compared to ground control group that had an average of 1199 counts/mm2 ($p = 0.06$).

The occurrence of lipid peroxidation was evaluated with antibody against 4-hydroxynonenal (4-HNE), which is an indicative marker of oxidative damage to the retina. Increased 4-HNE staining was seen in cone photoreceptors, retinal inner nuclear layer (INL), and ganglion cell layer (GCL) after spaceflight compared to ground controls (Fig 5C). As shown in Fig 5D, the fluorescent intensity profiles in the photoreceptor cones which reflects endogenous level of HNE, was also significantly increased ($p < 0.05$) in the spaceflight group compared to the ground controls. Fluorescent intensity from the HNE marker was also significantly increased in the GCL/INL of the spaceflight group for the retina ($p < 0.01$) (Fig 5E). Our RNA-seq experiment identified three DEGs - *Pde8a, Hif1a, Hgf* - that are part of the GO Biological

Process category for ‘negative regulation of oxidative stress-induced cell death’ (GO:1903202) (Fig 5F). *Pde8a* is a member of the phosphodiesterase family. Phosphodiesterase specific member inhibition has been previously shown to attenuate oxidative stress in diverse tissues (Masood et al., 2008; Zhuo et al., 2016; Koka et al., 2013; J. Li et al., 2016). *Hif1a* was shown to protect against oxidative stress directly at the mitochondria (H.-S. Li et al., 2019), but the effects of this gene were shown to be highly contextual in the presence of reactive oxygen species (Movafagh, Crook, and Vo 2015). *Hgf* activates an antioxidant signaling pathway that culminates in increased levels of nitric oxide and the activity of antioxidant enzymes regulated by this molecule (Guoguo et al., 2012; Romero-Puertas and Sandalio 2016). Taken together, the upregulation of *Pde8a* and the downregulation of the *Hif1a* and *Hgf* are linked to the development of increased oxidative stress in the retina.

3.4 Discussion

We have found a distinct gene expression signature in the retina of mice exposed to the spaceflight environment compared with the ground control. This signature is enriched for genes related to visual perception, the phototransduction pathway, and numerous retina and photoreceptor phenotype categories. Other studies have also shown changes of gene expression involved in cell structure, growth, migration and adhesion in the retinal cells exposed to simulated microgravity (J. E. Roberts et al., 2006; Corydon et al., 2016). Less anticipated was the overrepresentation of genes related to RNA processing, metabolism, and transport pathways for the first time described as relevant for spaceflight in mammals. A recent study in *Arabidopsis thaliana* has discovered mRNA isoforms created by alternative splicing that are unique to their spaceflight exposed samples (Beisel et al., 2019). Our results indicate that in addition to *Arabidopsis thaliana*, there may be splice isoform alterations in mammals during spaceflight as well that should be explored in future work and included in the risk assessment for long duration space missions.

Differential RNA splicing has also been implicated as a potential disease-causing mechanism for retinitis pigmentosa. Mutations in the gene *Rho*, which creates a protein that composes rod photoreceptors and is essential for vision in low-light conditions, can alter its splice isoforms. These alternate isoforms lead to the construction of abnormal RHO proteins and trigger the pathology of retinitis pigmentosa (Hernan et al., 2011). Additionally, proper alternative splicing of *Nxnl1* from rod photoreceptors and other ocular cell types has been associated with cone photoreceptor maintenance (Léveillard et al., 2004; Byrne et al., 2015). Disruption of these splice isoforms can reduce the density of cone photoreceptors and the thickness of the outer nuclear layer of the retina (Cronin et al., 2010). Overall, our results indicate that the space environment might create unique epigenetic events and differential RNA splicing, in addition to the altered gene expression profile reported. Future studies should address

which genes are alternatively spliced to understand potential changes in protein structure and find further associations with retinal diseases. Furthermore, mouse models for retinitis pigmentosa are available. Determining which mouse model is most similar to mice exposed to the space environment will help determine the appropriate model to perform Earth-analog experiments. Finding an appropriate Earth-based analog will accelerate the speed of research dedicated to finding a countermeasure for the retinal damage caused by spaceflight.

Several transcription factors were identified as necessary for development and maintenance of the retina and other structures of the eye (Anand and Lachke 2017). Although eight of the DETFs are known for their importance for the retinal development and maintenance in Earth conditions, 21 of the DETFs have no reports in the literature linking them to the eye tissue. From the DETF overrepresentation analysis, transcription factors were identified that are known to influence chromatin architecture. Changes in chromatin architecture indicate that the cause for changes in retinal structure observed during spaceflight is more complex than solely alterations in RNA processing. The overrepresentation analysis showed enrichment of transcriptional regulators for H3K9 demethylation, which is a marker for constitutive heterochromatin. Demethylation of H3K9 can result in the unraveling of this heterochromatin, which can then influence the accessibility of genes and regulatory elements in the genome (Lachner, O'Sullivan, and Jenuwein 2003). This can further drive changes in gene expression and either create abnormal and potentially harmful cellular behavior or act as a compensatory mechanism to mitigate damage the retinal tissue has experienced from spaceflight. The connections between chromatin reorganization as a disease-causing or disease-mitigating mechanism during spaceflight has not yet been explored. Determining the locations of chromatin reorganization could provide insight to these questions. Furthermore, the literature does not report any of the DETFs directly related to retinitis pigmentosa, which could indicate that these genes, although not yet described, could have a role in retinitis pigmentosa or that they are part of spaceflight specific expression network that can be explored as biomarkers and/or drug targets.

Compensatory mechanisms to mitigate cellular damage have been observed in response to oxidative stress. Our immunostaining for marker 4-HNE showed evidence of oxidative stress within the retina. Evidence of a response to this oxidative stress was also observed in the gene expression data. The gene *Bnip3* was differentially expressed, showing upregulation during spaceflight (adjusted p-value: 2.94E-03; log2 fold-change: 0.309). *Bnip3* is known for its role in removing misfolded proteins from the cell (Awan et al., 2014). Misfolded proteins can occur as a result of oxidative stress. Also upregulated in the spaceflight data are heat shock genes *Hsp90aa1*, *Hsp90b1*, and *Hspa4l* which can assist correct protein folding and are known to protect against oxidative stress (Kalmar and Greensmith 2009; Ciechanover and Kwon 2017). From this, we see evidence of a potential compensatory mechanism becoming activated in spaceflight in order to correct and mitigate cellular damage accumulating in retinal tissue. One

consideration for long duration space missions are the compensatory mechanisms that are activated in the space environment. For long duration missions, it becomes imperative to evaluate for how long these compensatory mechanisms can be sustained and how to extend or amplify its beneficial effects. Additionally, the nearest Earth-analog disease that we identified, retinitis pigmentosa, has known links to oxidative stress and mitochondrial dysfunction. Although no cure is currently known, supplementation with certain minerals and vitamins, especially vitamin A, has been demonstrated to delay the disease progression (Y. Zhao et al., 2019) and should be considered a possible countermeasure to spaceflight induced retinal damage.

We also found evidence that the circadian rhythm of the retina was altered in spaceflight. The circadian rhythm regulates behavioral and physiological processes over a 24-hour period, principally controlled by day/night patterns. Disruptions in the circadian rhythm can lead to irregular sleep patterns, mood, and metabolism function, as well as the development of cancer49. The differentially expressed gene with the smallest adjusted p-value, *Drd4* (adjusted p-value: 4.31E-51; log2 fold-change: 0.812), is a dopamine receptor that has been associated with the circadian rhythm. Specifically, *Drd4* downregulates the gene *Adcy1 (Jackson et al., 2011)*, which was also found to be differentially expressed in this study (adjusted p-value: 1.45E-09; log2 fold-change: −0.427). *Adcy1* produces cyclic AMP, a signalling mechanism in numerous retinal functions, such as retinomotor movements, disc shedding, certain types of retinal degeneration, and apoptosis of photoreceptors (Besharse, Dunis, and Burnside 1982; Weiss et al., 1995; Fassina et al., 1997). Circadian rhythm genes are of special interest in the space environment, where day/night light signals are irregular. For example, astronauts aboard the ISS will experience up to 16 sunsets and sunrises a day (Thirsk et al. 2009). However, the mice in our experiment were in constant 12-h light/dark cycle, which leads us to believe that events other than the light/dark cycle itself are driving the changes in the circadian rhythm.

Increased photoreceptor cone damage, reduced thickness of the retinal ONL, retinal INL, RPE, and choroid layers of the eye in the spaceflight group compared to ground control were reported in our study. Proton radiation-induced damage to the photoreceptor layer was also observed in our previous rodent study (X. W. Mao et al., 2009). It is important to note that changes in global ocular structure and retinal layer thickness could have clinical and/or pathological significance (Movafagh, Crook, and Vo 2015). The ONL contains the nuclei of the photoreceptor cells and the INL contains the cell bodies of bipolar, horizontal, amacrine, and Müller cells. Our published study from a Space Shuttle Mission (STS-135) showed a significant increase of apoptosis in the ONL and INL of the photoreceptor layer in spaceflight mice (Xiao W. Mao et al., 2013). Photoreceptor loss is responsible for irreversible blindness in many retinal diseases (MacLaren et al., 2006) and photoreceptors are more sensitive to environmental damage (LaVail, Anderson, and Hollyfield 1989). Retinal damage, retinal photoreceptor cell loss, and thinning of the inner retina precede subsequent capillary degeneration (Zheng et al., 2007; Ueda

et al., 2010). Our findings strongly implicate the space environment causes retinal injury and the potential development and progression of retinal degeneration.

The potential benefit of micro-CT in quantitative evaluation of ophthalmologic structure was encouraging. Our results showed degradation in ocular structures by spaceflight, particularly of the retina, RPE, and choroid layers of the eye, as evidenced by their decreased thickness. MicroCT provides an easy and non-destructive technique to image without any modification. Additionally, it shows the entire region of interest digitally, thereby increasing accessibility and reproducibility of the findings. Future work should use volumetric data to perform other analyses that take advantage of the micro-CT imaging capabilities, since it has been used successfully for studying many normal and pathologic tissues (Hann et al., 2011).

One area of concern is whether observed degradation of photoreceptors, retinal damage, and ocular structure alteration will persist or whether it will reverse once the organism is returned to Earth. Future studies should investigate the long-term impact of observed structural changes of the retina and retinal photoreceptors function that may lead to late retinal degeneration. Additional studies are needed to determine which aspects of spaceflight are responsible for photoreceptor damage and oxidative stress. Leading causes include radiation and possible alterations in intracranial pressure (ICP) or intraocular pressure (IOP) (Tymko, Boulet, and Donnelly 2017; Nelson et al., 2017) induced by microgravity. It is possible that both space radiation and microgravity work in conjunction with one another to cause a more severe pathological result than either would in isolation. Other factors that may be driving differential gene expression include increased stress an organism experiences due to the physiological pressures of spacecraft takeoff and landing and the psychological pressures of spaceflight. The body size of an organism also impacts the amount of interstitial fluid per gram of body weight, which could impact the magnitude of fluctuating pressures influencing the retina (Andreev-Andrievskiy et al., 2018). Studies using larger organisms could better model the degree of fluctuating pressures experienced by humans during spaceflight.

There are many potential avenues left to explore in order to understand the response of the mammalian retina to spaceflight. In this study we have characterized the physical response of the retina, the degradation of photoreceptors, and the presence of oxidative stress markers. We have also observed several genes with significant differential expression in the spaceflight condition were also differentially expressed in the disease retinitis pigmentosa. Additionally, we suspect the changes that we are observing during spaceflight are influenced by alternative splicing and chromatin reorganization. There remains much work to be done on understanding the mechanism of spaceflight-induced oxidative stress on retinal damage. Additional work will need to be done to separate disease-causing changes in gene expression from compensatory changes. We also believe these data will be useful for understanding other spaceflight-related health concerns, such as misregulation of the circadian rhythm and accelerated aging. A full list

of DEGs has been provided as a resource (Sup Table 1) toward further understanding of associations between spaceflight, retinal function, and overall mammalian health.

3.5 Methods

Flight and control conditions

SpaceX successfully launched the 12th Commercial Resupply Service (CRS-12) payload at the Kennedy Space Center (KSC) on a 35-day mission in August, 2017. 10-week-old male C57BL/6 mice (n = 20) (Jackson laboratories, Inc. Bar harbor, ME) were flown for NASA's ninth Rodent Research experiment (RR-9). The mice lived in NASA's Rodent Habitats (RH) aboard ISS for 35 days before returning to Earth via SpaceX's Dragon capsule. All mice were maintained at an ambient temperature of 26–28 °C with a 12-h light/dark cycle during the flight. GC mice were placed into the same housing hardware used in flight, and environmental parameters such as temperature and carbon dioxide (CO2) levels were matched as closely as possible based on telemetry data. GC mice were fed the same special NASA food bar diet as the space flown mice. All mice received the same *ad libitum* access to food and water. The study followed the recommendations in the Guide for the Care and Use of Laboratory Animals of the National Institutes of Health (NIH) and was approved by the Institutional Animal Care and Use Committee (IACUC) of Loma Linda University (LLU) and The National Aeronautics and Space Administration (NASA).

Post-flight evaluation of the mice

Upon return to the Earth, mice were transported to Loma Linda University (LLU) within 28 hours of splashdown. At LLU, animals were removed from the animal enclosure hardware and assessed for survival and health. It was reported that all the mice survived the 35-day space mission and were in good condition, i.e. no obvious deficiencies/abnormalities as described by the inspecting personnel.

Dissecting and preservation of mouse eyes after spaceflight

The mice were rapidly euthanized in 100% CO2 and eyes were collected within 38+/−4 hours of splashdown (n = 20/group). The retinas from rights were dissected and placed individually in sterile cryovials, snap frozen in liquid nitrogen and kept at −80 °C prior to use. The left eyes were fixed in 4% paraformaldehyde in phosphate buffered saline (PBS) for 24 hours and then rinsed with PBS for immunohistochemistry (IHC) assays. The right eyes from 2 experimental

groups: ground control and spaceflight (n = 8 mice per group) were dissected to obtain the retina. The retinas were flash-frozen and stored at −80 °C for genomic analysis.

Immunostaining assays and histology

Paraffin-embedded sections (Six μm) of left eye, roughly 100 μm apart, were used for analysis (n = 6/group). To evaluate oxidative damage in the retina, Immunostaining was performed on ocular sections using 4-HNE antibody specific for lipid peroxidation. Sections were incubated with the anti-4-HNE antibody (catalog no. HNE11-S, Alpha Diagnostic International Inc., San Antonio, TX) at 4 °C for 2 hours followed by a donkey anti-rabbit IgG fluorescence-conjugated secondary antibody (catalog no. A21206, Invitrogen Corp., Waltham, MA) for 2 hours at room temperature and counterstained with DAPI. For double labeling of HNE and PNA, sections were incubated with fluorescein isothiocyanate (FITC)-conjugated Peanut agglutinin (1:100 in 1% BSA) for an hour at room temperature for labeling the photoreceptor cones. Sections were then incubated overnight (18-21 hours) at 4 °C with primary antibody against rabbit anti-4-HNE (Alpha Diagnostic Intl. Inc. San Antonio, TX). After washing three times in PBS, sections were further treated with secondary antibody Alexa Fluor 568 goat anti-rabbit IgG (Life Technologies, Eugene, Oregon). PNA was obtained from (Vector Laboratories, Burlingame, CA). The cell nuclei were counterstained with DAPI and were mounted and coverslipped with Vectashield Hard-Set Mounting Medium (Vector Laboratories). To characterize layers of retinal structure, a series of 6 μm sections were also stained with hematoxylin and eosin (H&E). Images were captured with a BZ-X710 All-in-One inverted fluorescence microscope. In each sample, a total of 5 fields of images were sampled systematically at 20X magnification spanning the entire retina sections.

Quantification of immunostaining

For quantitative analysis, the number of PNA- positive cells were counted in three sections of tissue from each animal using Image J counting plugin 1.41 software (National Institutes of Health, Bethesda, MD; http://rsbweb.nih.gov/ij/). The density profiles were expressed as the mean number of PNA cells/mm2, and counts were averaged across each group. To determine HNE immunoreactivity, fluorescence intensity was measured on 5 randomly selected fields on each section and calculated using Image J software. The data were extracted and averaged within the group. A detailed method was described in Mao *et al*. (Xiao W. Mao et al., 2013).

Sample preparation for MicroCT scanning

A subset of eye samples (n = 6/group) were fixed in 4% formaldehyde in PBS. After fixation, the mice eyes were dehydrated in ethanol. To prevent an abrupt shrinkage of the fixed

sample, a graded series of ethanoic solutions were used. Afterwards, the mice eyes were stained in 10% phosphomolybdic acid solution (PMA) for 6 days. Samples were washed in absolute ethanol and then placed each eye in the same rotation in individual 2 mL plastic containers filled with absolute ethanol for scanning.

Micro CT scanning and analysis

Samples were scanned using a high resolution micro-computed tomography system (SkyScan 1272 desktop micro-CT system, Bruker, Kontich, Belgium), with an accelerating source voltage of 50 keV, a source current of 80 mA with an integration time of 90 min. During the scanning process, the samples were rotated at 180 degrees, with an imaging voxel size of 4.5 um, frame averaging of 4, rotation step of 0.4 and no filter. These images were saved for the reconstruction of the 3D object. The thickness of the retina and surrounding tissues were measured on MicroCT images by descriptive analysis. Measurements of the lens, retina, RPE layer, choroid and sclera were performed in the middle slice of the sagittal view of each sample using the optical nerve as reference. Three measurements were recorded of each structure to finally obtain an average.

RNA/DNA extraction

Retina tissues (n = 8/group) were homogenized with Precellys CKM beads (Hilden, Germany) in 600 µl of RLT plus buffer. Total RNA and DNA were extracted using the QIAGEN AllPrep DNA/RNA/miRNA Universal Kit according to the manufacturer's protocol. After isolation, RNA and DNA were frozen and kept at −80 °C until further use. Isolated RNA samples were quantified and purity checked by NanoDrop spectrophotometry (Thermo Fisher Scientific, Chino, CA). DNA and RNA were further quantified using the Qubit dsDNA High Sensitivity Kit and RNA Broad Range Kit, respectively (Life Technologies, Carlsbad, CA). RNA quality was evaluated using the Agilent 2200 TapeStation and RNA ScreenTape with 1ul of randomly selected RNA denatured at 72 °C for three minutes per manufacturer's instructions (Santa Clara, CA). RNA integrity numbers (RIN) of tested RNA samples had values between 7.7 and 8.9.

RNA-Seq library construction

The Ovation® Mouse RNA-Seq System 1-16 (NuGEN Technologies, #0348) was used per manufacturer's instructions to construct stranded RNA-seq libraries. 100 ng of total RNA was used as input. First and second strands of cDNA were synthesized from total RNA (100 ng) spiked with 1 µl of 1:500 diluted ERCC ExFold RNA Spike-In Mix 2 (Life Technologies, Carlsbad, CA) at the appropriate ratio. Following primer annealing and cDNA synthesis, the products were sheared using Covaris S220 Focused-ultrasonicator (Covaris Inc., Woburn, MA).

130 µl of each sample was sheared according to the manufacturer's instructions. The parameters were set as follows: 10% duty factor, peak power 175 and 200 cycles per burst at 4 °C for 200 seconds to obtain fragment sizes between 150–200 bp. This was followed by end-repair, adaptor index ligation and strand selection. Strand selection was performed by using custom InDA-C primer mixture SS5 Version5 for mice with cytoplasmic and mitochondrial ribosomal RNA depletion. Finally, libraries were amplified using 17 cycles (Mastercycler® pro, Eppendorf, Hamburg, Germany), and purified with RNAClean XP Agencourt beads.

Library quantification and quality control (QC) test

The final amplified libraries were purified using Agencourt RNAClean XP beads (Beckman Coulter, Indianapolis, IN) and quantified using Qubit dsDNA HS Kit on Qubit 3.0 Fluorometer (Life Technologies, Carlsbad, CA). Quality and peak size was determined using D1000 ScreenTape on the Agilent 2200 TapeStation (Agilent Technologies, Santa Clara, CA).

Sequencing

Final libraries were each diluted to 4 nM and further quantitated to ensure high accuracy quantification for consistent pooling of barcoded libraries and maximization of the number of clusters in the Illumina flow cell. Libraries of different indices were pooled for sequencing together in equimolar amounts. Pooled libraries were quantified by Qubit prior to sequencing. For sequencing, clusters were generated on cBot with HiSeq3000/4000 PE cluster kit (Illumina, Inc., San Diego, CA) from 5 ul of 2 nM final library pool (200 pM at cBot loading). Sequencing was performed on an Illumina HiSeq4000 (Loma Linda University Center for Genomics) using 75 cycles SBS reagents (Illumina, Inc., San Diego, CA). Single-indexed and single-end reads with 75 nucleotides length (1 × 75 bp) were generated. Fastq files generated from the sequence run were demultiplexed (Loma Linda University Center for Genomics), quality checked (quality score >38), trimmed from adapters with Trimmomatic (version 0.01) and mapped against mouse reference genome (NCBI_build_37.2) with TOPHAT2 (version 2). Overall mapping rate was ~85%. Expressed transcripts were counted with Multi HTSeq (version 0.02).

Computational pipeline/RNA-Seq analysis tools

Differential Expression analysis was carried out using raw counts from input to the R package DEseq2 Version 1.22.2 using R version 3.5.1. A gene was considered differentially expressed (DE) if was below a false discovery rate (FDR) adjusted p-value of 0.1 (Benjamini and Hochberg 1995). Heatmaps were generated using the webtool Morpheus, from the Broad Institute (https://software.broadinstitute.org/morpheus/). Hierarchical clustering was performed

on rows and columns using one minus Pearson correlation. DEGs that are transcription factors were annotated using the Panther database version 14.1 (Mi et al., 2019).

GO, gene network analysis, and phenotype graphs were created by performing an overrepresentation enrichment analysis using WebGestalt (Wang et al., 2017). GO analysis was carried out using the biological process database. Gene network analysis was carried out using the Reactome database (Fabregat et al., 2018). Phenotype analysis was carried out using mammalian phenotype ontology. FDRs were calculated using the Benjamini-Hochberg procedure.

Human disease associated genes were found for the diseases retinitis pigmentosa, age-related macular degeneration, diabetic retinopathy, and retinal detachment using the database DisGeNET version 6.0 with a gene-disease association (GDA) score greater than 0.2 (Piñero et al., 2017). Human genes were converted to their mouse orthologs using the Mouse Genome Informatics database (Smith et al., 2018). The genes KIZ, PCARE, RP2, and RP9 did not have their orthologs listed in the Mouse Genome Informatics database and were manually annotated using the GeneCards database (v4.10.0 Build 4) (Rebhan 1997).

3.6 Figures

Figure 1

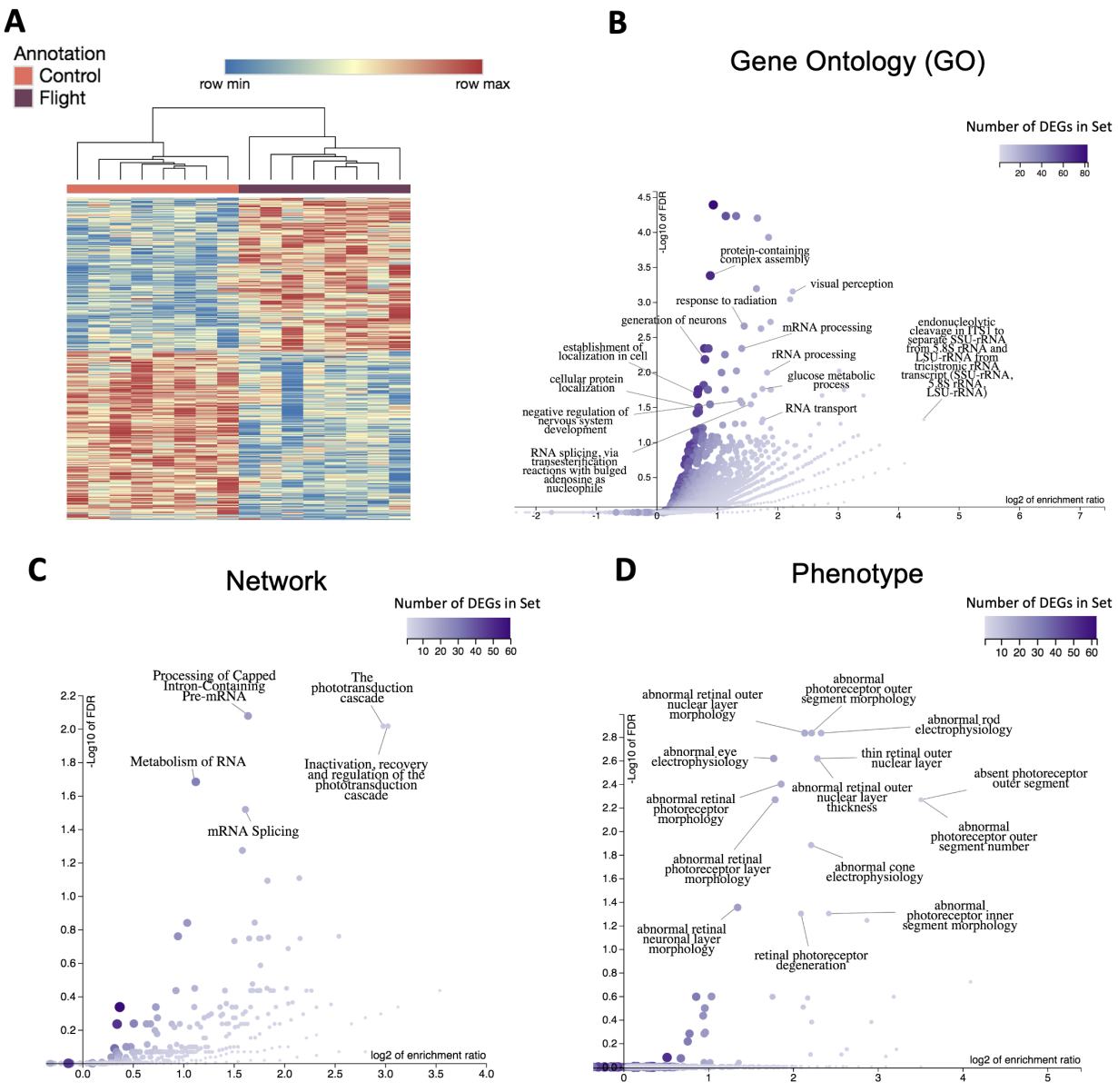
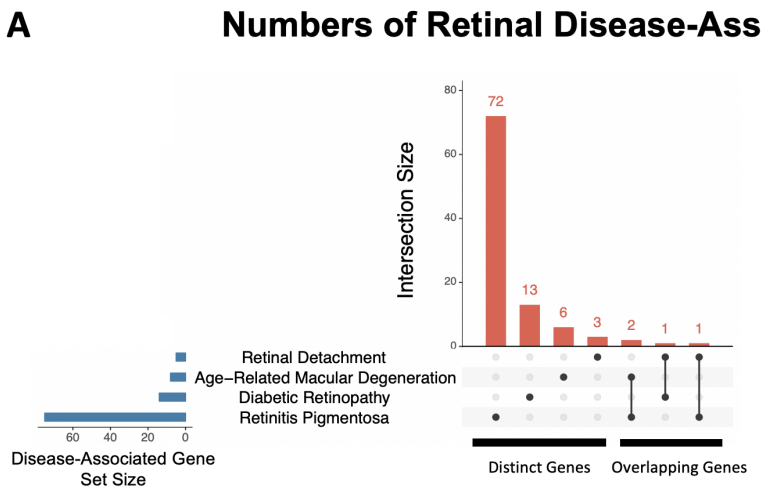


Figure 1. DEG clustering and functions between spaceflight and control mice. (A) Hierarchical clustering of the 600 DEGs between spaceflight and control mice using an adjusted p-value threshold of 0.1. The spaceflight group had 286 upregulated genes and 314 downregulated genes compared to the ground control; **(B)** Enriched gene ontology (GO) biological process categories for DEGs. The affinity propagation option from WebGestalt was applied to the select representative display categories; **(C)** Enriched networks among DEGs from the Reactome database; **(D)** Enriched phenotypes impacted by the DEGs from the Mammalian Phenotype Ontology; **(B-D)** Overrepresented categories were found relating to ocular function (GO categories: 'visual perception', 'response to light stimulus', 'sensory perception of light stimulus', 'retina development in camera-type eye'; Pathways: 'the phototransduction cascade'

'inactivation, recovery and regulation of the phototransduction cascade'; Phenotype: electrophysiology, morphology, and degeneration of the retina, rods, and cones), various RNA processing, splicing, and metabolism functions (GO categories: 'RNA processing', 'rRNA processing', 'mRNA processing', 'RNA splicing', via transesterification reactions', 'mRNA splicing, via spliceosome', 'rRNA metabolic process', 'ncRNA metabolic process', 'mRNA metabolic process', 'RNA transport'), and direct responses to the physical pressures of spaceflight (GO categories: 'response to abiotic stimulus', 'response to radiation', 'cellular response to stress').

Figure 2



Genes Shared Across Multiple Disease Association Lists	
Retinitis Pigmentosa and Age-Related Macular Degeneration	*Abca4* *Nmnat1*
Diabetic Retinopathy and Retinal Detachment	*Casp3*
Retinitis Pigmentosa and Retinal Detachment	*Crb1*

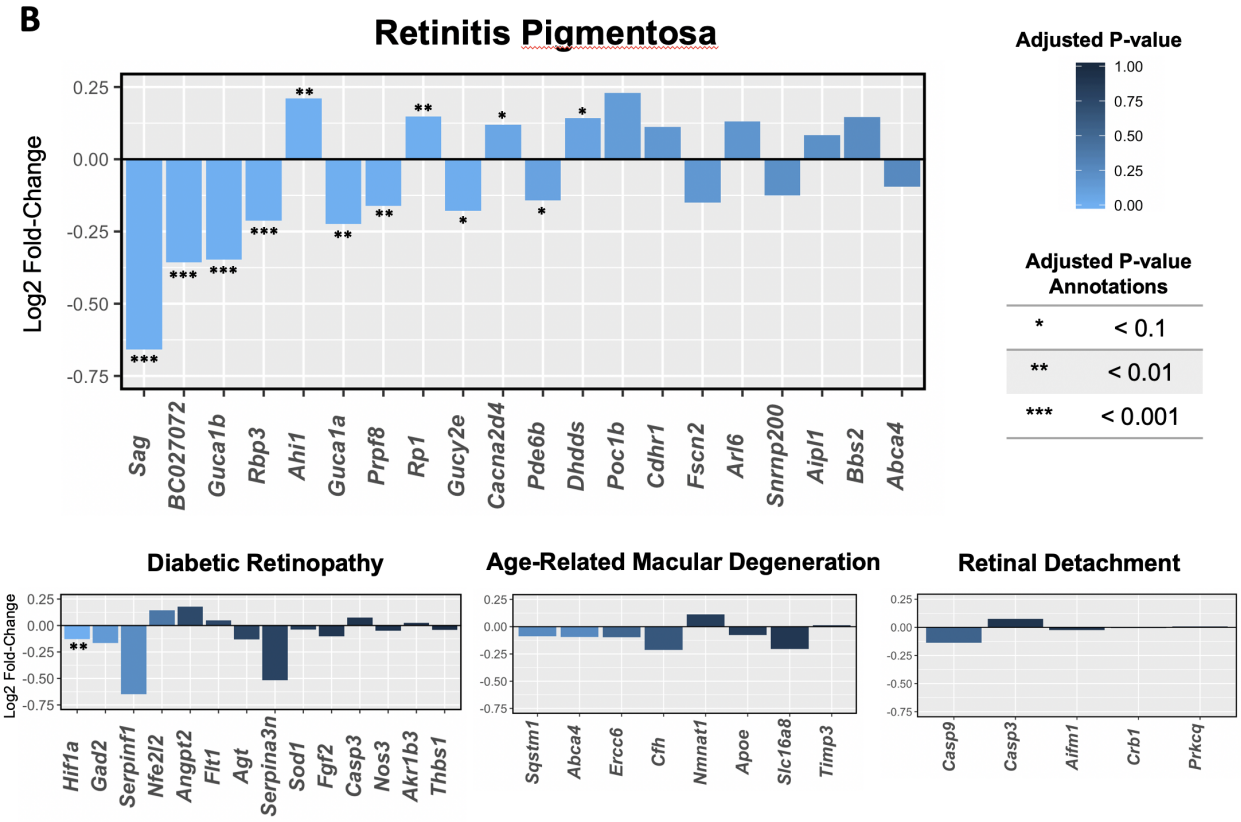


Figure 2. Retinal disease-associated gene expression. (A) UpSet diagram showing the number of genes in each disease set from DisGeNet. Genes that are distinct to each set are in the first four columns. Genes shared among sets are in the last three columns; **(B)** Log2 fold-change values of disease-associated genes. Genes are plotted from left to right in order of adjusted p-value. Bars are colored by the magnitude of the adjusted p-value. Adjusted p-values are further annotated based on their order of magnitude. The plot for retinitis pigmentosa displays the top 20 most significant genes based on adjusted p-value. All other diseases are displaying all of their associated genes.

Figure 3

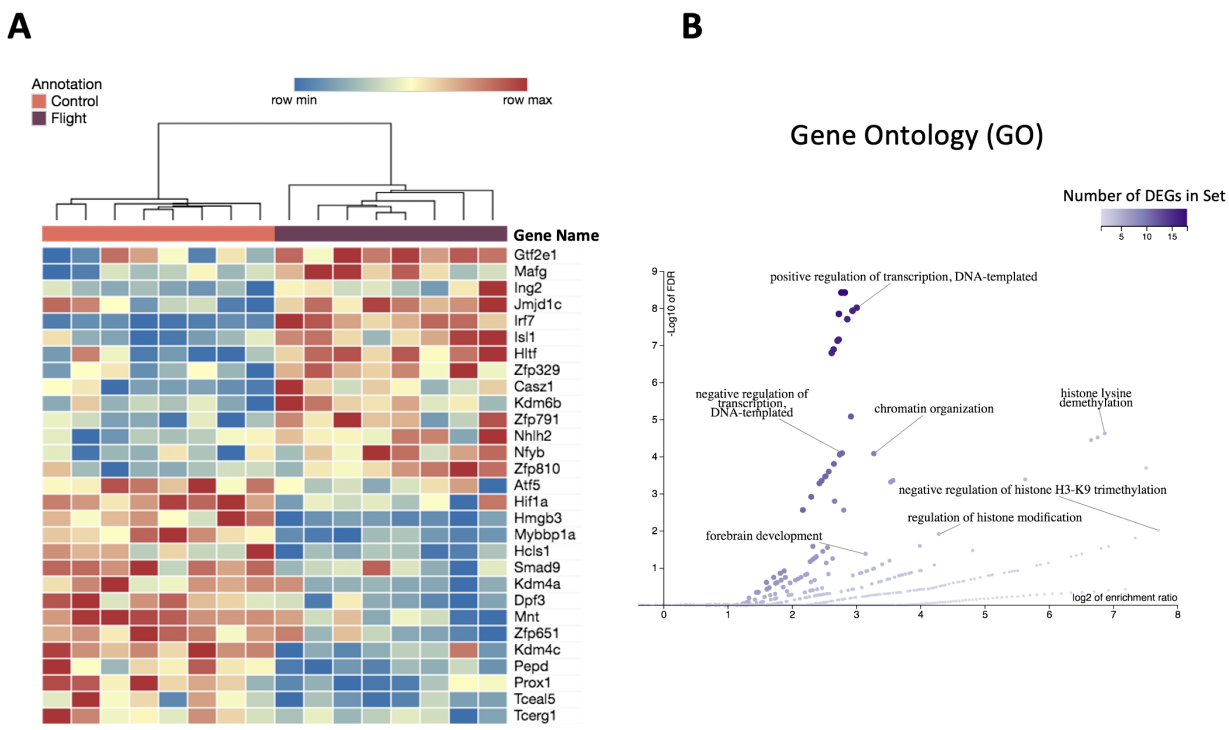
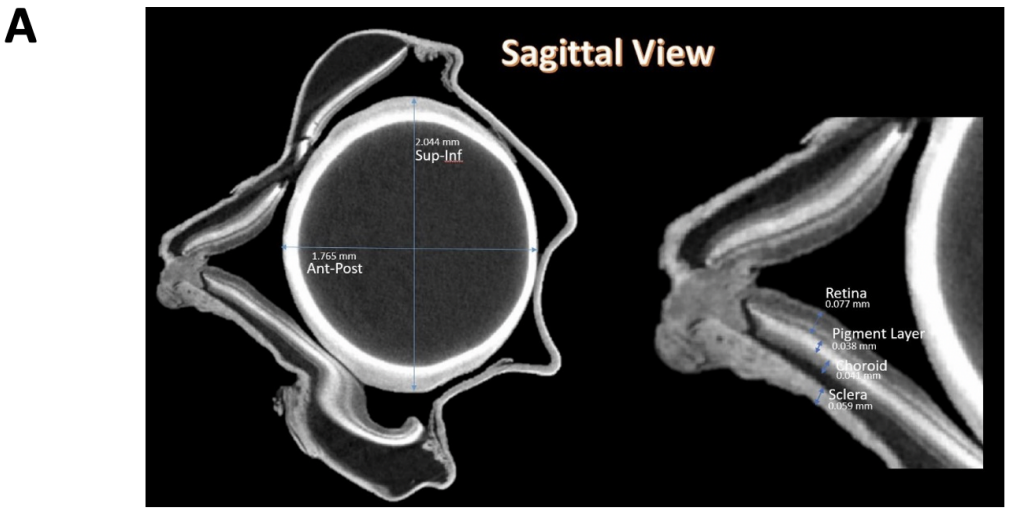


Figure 3. Transcription factor clustering and functions between spaceflight and control mice. (A) Hierarchical clustering of the 29 DETFs between spaceflight and ground control mice; **(B)** Enriched gene ontology (GO) biological process categories for DETFs.

Figure 4



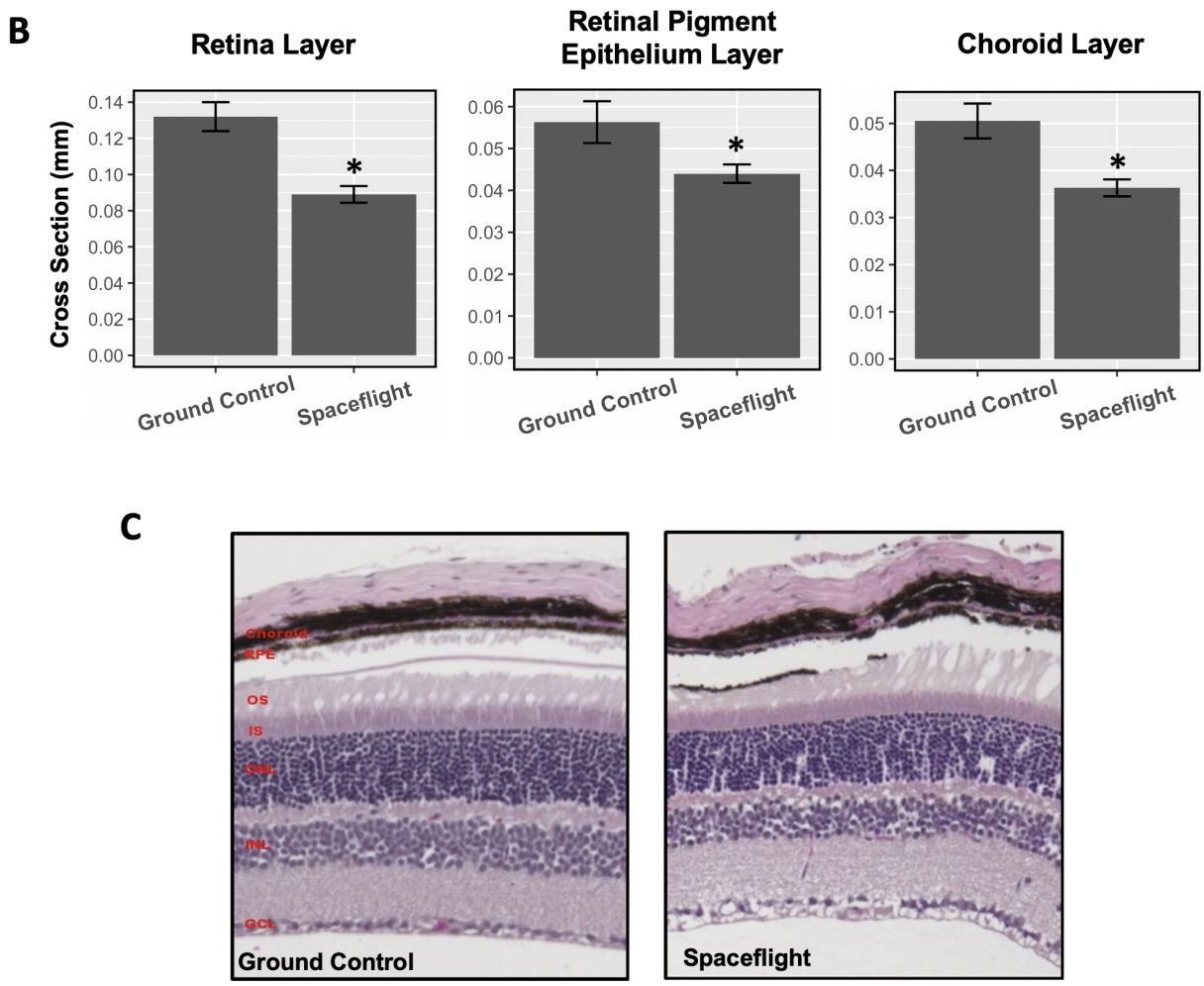


Figure 4. Spaceflight decreases the thickness of multiple layers of the eye. (A) Sagittal view of of a ground control mouse. Layers of the eye on the right side of the image are annotated, from top-to-bottom, retina (0.077mm), retina pigment layer (RPE, 0.038mm), choroid (0.041mm), sclera (0.059mm); **(B)** Average thickness of the retinal layer, RPE layer, and the choroid layer measured by MicroCT in the spaceflight and control groups. Counts were averaged across five retinas per group. Values were represented as mean thickness + standard error (SEM). SEM of the mean is marked with error bars. Significantly lower in cross section thickness in the spaceflight group compared to the ground control group is denoted '*' ($p<0.05$). **(C)** Cross sections of the retina from control and spaceflight mice. GCL: ganglion cell layer; INL: inner nuclear layer; ONL: outer nuclear layer; IS: inner segment; OS: outer segment, RPE: pigment epithelium layer.

Figure 5

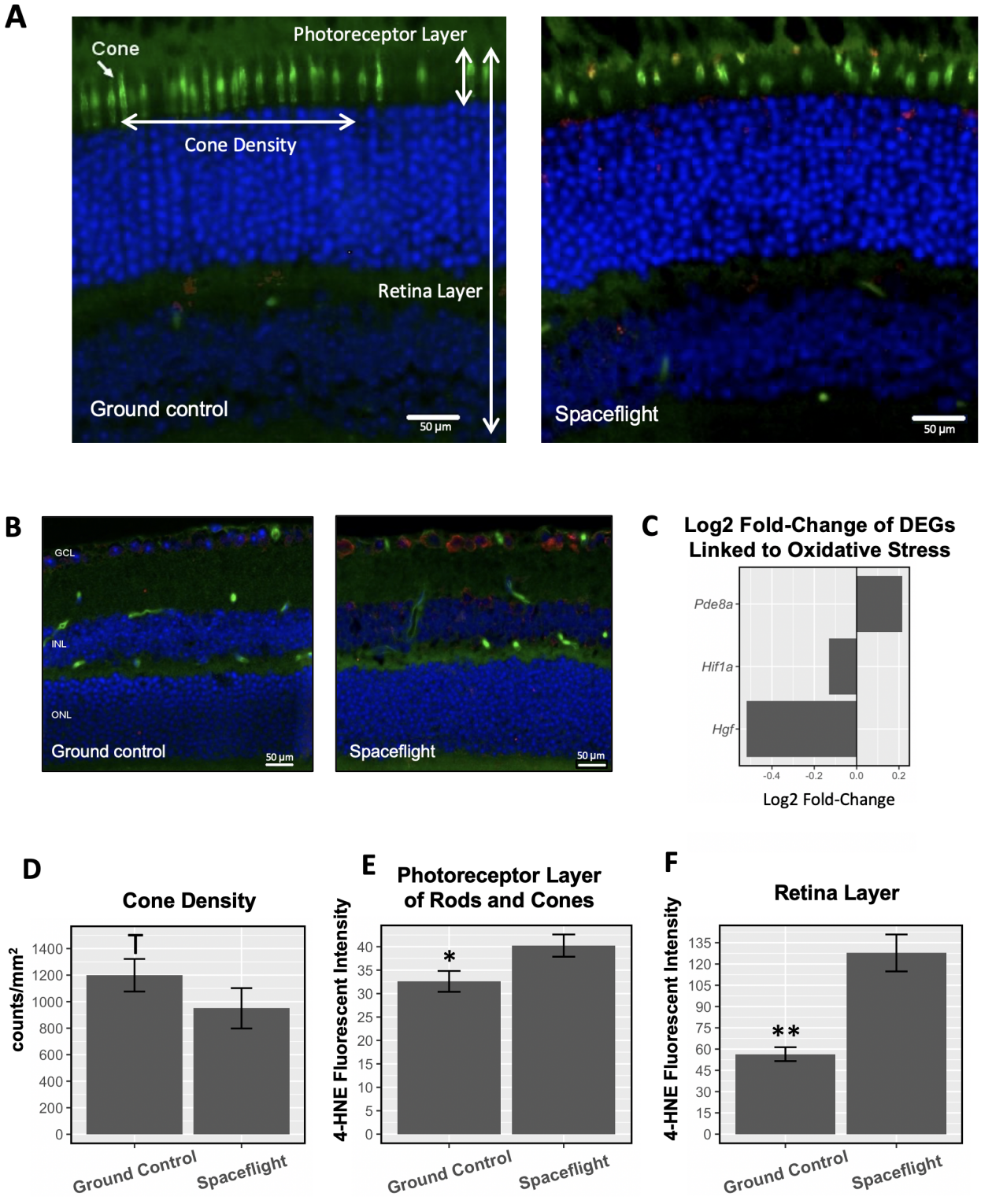
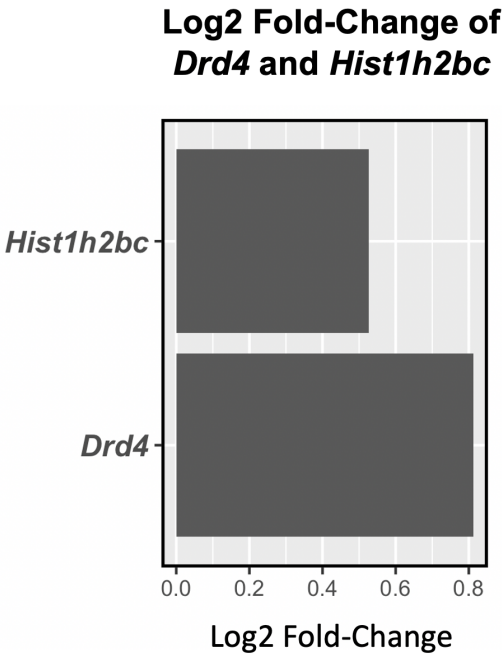


Figure 5. Spaceflight causes photoreceptor degradation and oxidative stress. (A) Immunofluorescence staining for PNA, a marker for cone photoreceptors (green), and HNE, a

marker for oxidative stress (red), in the photoreceptor layer. The nuclei were counterstained with DAPI (blue); Scale bar =50 mm. (B) Immunofluorescence staining for HNE (red) in the retina layer. Scale bar = 50 μm; (C) Log2 fold-change of DEGs under the GO category "negative regulation of oxidative stress-induced cell death"; (D) Cell density for PNA–positive cone photoreceptors; (E) Fluorescent intensity of the HNE marker in the photoreceptor layer of rods (arrow) and cones (arrow); (F) Fluorescent intensity of the HNE marker across the retina; (D-F) Counts were averaged across five retinas per group. Values are represented as mean density ± SEM. Significance values of the spaceflight group compared to ground controls are denoted with '*' ($p<0.05$), '**' ($p<0.01$), and 'T' (strong trend decrease compared ground controls; $p=0.06$).

3.7 Supplemental Figures

Sup Fig 1



Supplementary Figure 1. Log2 fold-change of *Drd4* and *Hist1h2bc*. *Drd4* has an FDR of 4.31E-51 and is related to circadian rhythms. *Hist1h2bc* has an FDR of 1.66E-09 and is related to the aging retina.

3.8 Author Contributions

X.M. designed research. X.M., N.N. and. S.S. carried out the main laboratory experiments. E.G.O., W.A.S., G.H. and X.M. performed data analysis, data interpretation and wrote the main manuscript. M.D.D., M.P., C.W. J.S.W., D.C.Z. and G.R.T. reviewed the manuscript.

3.9 Data Availability

The datasets generated during and/or analysed during the current study are available in the GEO repository, under the accession identifier GSE131954.

3.10 References

1. Dawson Linda. *The Politics and Perils of Space Exploration.* Cham: Springer International Publishing; 2016. The New Space Race; pp. 1–24.

2. Mao XW, et al. Spaceflight environment induces mitochondrial oxidative damage in ocular tissue. *Radiat. Res.* 2013;**180**:340–350. doi: 10.1667/RR3309.1.

3. Overbey EG, et al. Mice Exposed to Combined Chronic Low-Dose Irradiation and Modeled Microgravity Develop Long-Term Neurological Sequelae. *International Journal of Molecular Sciences.* 2019;**20**(17):4094. doi: 10.3390/ijms20174094.

4. Nelson ES, Mulugeta L, Myers JG. Microgravity-induced fluid shift and ophthalmic changes. *Life.* 2014;**4**:621–665. doi: 10.3390/life4040621.

5. Lee AG, Mader TH, Robert Gibson C, Brunstetter TJ, Tarver WJ. Space flight-associated neuro-ocular syndrome (SANS) *Eye.* 2018;**32**:1164–1167. doi: 10.1038/s41433-018-0070-y.

6. Stenger, M. B. *et al*. Evidence Report: Risk of Spaceflight Associated Neuro-ocular Syndrome (SANS) (2017).

7. Garrett-Bakelman, F. E. *et al*. The NASA Twins Study: A multidimensional analysis of a year-long human spaceflight. *Science***364** (2019).

8. Mader TH, et al. Optic disc edema, globe flattening, choroidal folds, and hyperopic shifts observed in astronauts after long-duration space flight. *Ophthalmology.* 2011;**118**:2058–2069. doi: 10.1016/j.ophtha.2011.06.021.

9. Mao X. W., Boerma M., Rodriguez D., Campbell-Beachler M., Jones T., Stanbouly S., Sridharan V., Nishiyama N. C., Wroe A., Nelson G. A. Combined Effects of Low-Dose Proton

Radiation and Simulated Microgravity on the Mouse Retina and the Hematopoietic System. *Radiation Research.* 2018;**192**(3):241. doi: 10.1667/RR15219.1.

10. Mao XW, et al. Acute Effect of Low-Dose Space Radiation on Mouse Retina and Retinal Endothelial Cells. *Radiat. Res.* 2018;**190**:45–52. doi: 10.1667/RR14977.1.

11. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* 2014;**15**:550. doi: 10.1186/s13059-014-0550-8.

12. Wang J, Vasaikar S, Shi Z, Greer M, Zhang B. WebGestalt 2017: a more comprehensive, powerful, flexible and interactive gene set enrichment analysis toolkit. *Nucleic Acids Res.* 2017;**45**:W130–W137. doi: 10.1093/nar/gkx356.

13. Jackson CR, Chaurasia SS, Hwang CK, Iuvone PM. Dopamine D4 receptor activation controls circadian timing of the adenylyl cyclase 1/cyclic AMP signaling system in mouse retina. *Eur. J. Neurosci.* 2011;**34**:57–64. doi: 10.1111/j.1460-9568.2011.07734.x.

14. Banday AR, et al. Replication-dependent histone genes are actively transcribed in differentiating and aging retinal neurons. *Cell Cycle.* 2014;**13**:2526–2541. doi: 10.4161/15384101.2015.941757.

15. Guo J-H, et al. Keeping the right time in space: importance of circadian clock and sleep for physiology and performance of astronauts. *Mil Med Res.* 2014;**1**:23. doi: 10.1186/2054-9369-1-23.

16. Demontis GC, et al. Human Pathophysiological Adaptations to the Space Environment. *Front. Physiol.* 2017;**8**:547. doi: 10.3389/fphys.2017.00547.

17. Piñero J, et al. DisGeNET: a comprehensive platform integrating information on human disease-associated genes and variants. *Nucleic Acids Res.* 2017;**45**:D833–D839. doi: 10.1093/nar/gkw943.

18. Smith CL, et al. Mouse Genome Database (MGD)-2018: knowledgebase for the laboratory mouse. *Nucleic Acids Research.* 2018;**46**:D836–D842. doi: 10.1093/nar/gkx1006.

19. Mi H, Muruganujan A, Ebert D, Huang X, Thomas PD. PANTHER version 14: more genomes, a new PANTHER GO-slim and improvements in enrichment analysis tools. *Nucleic Acids Res.* 2019;**47**:D419–D426. doi: 10.1093/nar/gky1038.

20. Anand D, Lachke SA. Systems biology of lens development: A paradigm for disease gene discovery in the eye. *Exp. Eye Res.* 2017;**156**:22–33. doi: 10.1016/j.exer.2016.03.010.

21. Bejarano-Escobar R, et al. Expression and function of the LIM-homeodomain transcription factor Islet-1 in the developing and mature vertebrate retina. *Exp. Eye Res.* 2015;**138**:22–31. doi: 10.1016/j.exer.2015.06.021.

22. Yasuda M, Tanaka Y, Ryu M, Tsuda S, Nakazawa T. RNA sequence reveals mouse retinal transcriptome changes early after axonal injury. *PLoS One.* 2014;**9**:e93258. doi: 10.1371/journal.pone.0093258.

23. Terada K, Furukawa T. Sumoylation controls retinal progenitor proliferation by repressing cell cycle exit in Xenopus laevis. *Dev. Biol.* 2010;**347**:180–194. doi: 10.1016/j.ydbio.2010.08.023.

24. Mattar P, Stevanovic M, Nad I, Cayouette M. Casz1 controls higher-order nuclear organization in rod photoreceptors. *Proc. Natl. Acad. Sci. U. S. A.* 2018;**115**:E7987–E7996. doi: 10.1073/pnas.1803069115.

25. Barben M, et al. Hif1a inactivation rescues photoreceptor degeneration induced by a chronic hypoxia-like stress. *Cell Death Differ.* 2018;**25**:2071–2085. doi: 10.1038/s41418-018-0094-7.

26. Pérez de Sevilla Müller L, Azar SS, de Los Santos J, Brecha NC. Prox1 Is a Marker for AII Amacrine Cells in the Mouse Retina. *Front. Neuroanat.* 2017;**11**:39. doi: 10.3389/fnana.2017.00039.

27. Iida A, et al. Roles of histone H3K27 trimethylase Ezh2 in retinal proliferation and differentiation. *Developmental Neurobiology.* 2015;**75**:947–960. doi: 10.1002/dneu.22261.

28. Labbé RM, Holowatyj A, Yang Z-Q. Histone lysine demethylase (KDM) subfamily 4: structures, functions and therapeutic potential. *Am. J. Transl. Res.* 2013;**6**:1–15.

29. Masood A, Nadeem A, Mustafa SJ, O'Donnell JM. Reversal of oxidative stress-induced anxiety by inhibition of phosphodiesterase-2 in mice. *J. Pharmacol. Exp. Ther.* 2008;**326**:369–379. doi: 10.1124/jpet.108.137208.

30. Zhuo Y, et al. Inhibition of phosphodiesterase-4 reverses the cognitive dysfunction and oxidative stress induced by Aβ25–35 in rats. *Metab. Brain Dis.* 2016;**31**:779–791. doi: 10.1007/s11011-016-9814-1.

31. Koka S, Das A, Salloum FN, Kukreja RC. Phosphodiesterase-5 inhibitor tadalafil attenuates oxidative stress and protects against myocardial ischemia/reperfusion injury in type 2 diabetic mice. *Free Radic. Biol. Med.* 2013;**60**:80–88. doi: 10.1016/j.freeradbiomed.2013.01.031.

32. Li J, et al. Protective effects of BAY 73-6691, a selective inhibitor of phosphodiesterase 9, on amyloid-β peptides-induced oxidative stress in *in-vivo* and *in-vitro* models of Alzheimer's disease. *Brain Research.* 2016;**1642**:327–335. doi: 10.1016/j.brainres.2016.04.011.

33. Li, H.-S. *et al*. HIF-1α protects against oxidative stress by directly targeting mitochondria. *Redox Biol* 101109, 10.1016/j.redox.2019.101109 (2019).

34. Movafagh S, Crook S, Vo K. Regulation of hypoxia-inducible factor-1a by reactive oxygen species: new developments in an old debate. *J. Cell. Biochem.* 2015;**116**:696–703. doi: 10.1002/jcb.25074.

35. Guoguo S, et al. HGF-mediated inhibition of oxidative stress by 8-nitro-cGMP in high glucose-treated rat mesangial cells. *Free Radic. Res.* 2012;**46**:1238–1248. doi: 10.3109/10715762.2012.701292.

36. Romero-Puertas MC, Sandalio LM. Nitric Oxide Level Is Self-Regulating and Also Regulates Its ROS Partners. *Front. Plant Sci.* 2016;**7**:316. doi: 10.3389/fpls.2016.00316.

37. Roberts JE, et al. Simulated microgravity induced damage in human retinal pigment epithelial cells. *Mol. Vis.* 2006;**12**:633–638.

38. Corydon TJ, et al. Reduced Expression of Cytoskeletal and Extracellular Matrix Genes in Human Adult Retinal Pigment Epithelium Cells Exposed to Simulated Microgravity. *Cell. Physiol. Biochem.* 2016;**40**:1–17. doi: 10.1159/000452520.

39. Beisel NS, Noble J, Barbazuk WB, Paul A-L, Ferl RJ. Spaceflight-induced alternative splicing during seedling development in Arabidopsis thaliana. *NPJ Microgravity.* 2019;**5**:9. doi: 10.1038/s41526-019-0070-7.

40. Hernan I, et al. Cellular expression and siRNA-mediated interference of rhodopsin cis-acting splicing mutants associated with autosomal dominant retinitis pigmentosa. *Invest. Ophthalmol. Vis. Sci.* 2011;**52**:3723–3729. doi: 10.1167/iovs.10-6933.

41. Léveillard T, et al. Identification and characterization of rod-derived cone viability factor. *Nat. Genet.* 2004;**36**:755–759. doi: 10.1038/ng1386.

42. Byrne LC, et al. Viral-mediated RdCVF and RdCVFL expression protects cone and rod photoreceptors in retinal degeneration. *J. Clin. Invest.* 2015;**125**:105–116. doi: 10.1172/JCI65654.

43. Cronin T, et al. The disruption of the rod-derived cone viability gene leads to photoreceptor dysfunction and susceptibility to oxidative stress. *Cell Death Differ.* 2010;**17**:1199–1210. doi: 10.1038/cdd.2010.2.

44. Lachner M, O'Sullivan RJ, Jenuwein T. An epigenetic road map for histone lysine methylation. *J. Cell Sci.* 2003;**116**:2117–2124. doi: 10.1242/jcs.00493.

45. Awan MUF, et al. Neuroprotective role of BNIP3 under oxidative stress through autophagy in neuroblastoma cells. *Mol. Biol. Rep.* 2014;**41**:5729–5734. doi: 10.1007/s11033-014-3444-7.

46. Kalmar B, Greensmith L. Induction of heat shock proteins for protection against oxidative stress. *Adv. Drug Deliv. Rev.* 2009;**61**:310–318. doi: 10.1016/j.addr.2009.02.003.

47. Ciechanover A, Kwon YT. Protein Quality Control by Molecular Chaperones in Neurodegeneration. *Front. Neurosci.* 2017;**11**:185. doi: 10.3389/fnins.2017.00185.

48. Zhao Y, et al. Vitamins and Mineral Supplements for Retinitis Pigmentosa. *J. Ophthalmol.* 2019;**2019**:8524607.

49. Zhang Luoying, Ptáček Louis J., Fu Ying-Hui. *Progress in Molecular Biology and Translational Science.* 2013. Diversity of Human Clock Genotypes and Consequences; pp. 51–81.

50. Besharse JC. Effects of cyclic adenosine 3′,5′-monophosphate on photoreceptor disc shedding and retinomotor movement. Inhibition of rod shedding and stimulation of cone elongation. *The Journal of General Physiology.* 1982;**79**:775–790. doi: 10.1085/jgp.79.5.775.

51. Weiss, E. R., Hao, Y., Dickerson, C. D. & Osawa, S. Altered cAMP levels in retinas from transgenic mice expressing a rhodopsin mutant. *Biochemical and* (1995).

52. Fassina, G., Aluigi, M. G. & Gentleman, S. The cAMP analog 8-Cl-cAMP inhibits growth and induces differentiation and apoptosis in retinoblastoma cells. *J. Cancer* (1997).

53. Thirsk R, Kuipers A, Mukai C, Williams D. The space-flight environment: the International Space Station and beyond. *CMAJ.* 2009;**180**:1216–1220. doi: 10.1503/cmaj.081125.

54. Mao XW, et al. Radioprotective effect of a metalloporphyrin compound in rat eye model. *Curr. Eye Res.* 2009;**34**:62–72. doi: 10.1080/02713680802546948.

55. MacLaren RE, et al. Retinal repair by transplantation of photoreceptor precursors. *Nature.* 2006;**444**:203–207. doi: 10.1038/nature05161.

56. LaVail, M. M., Anderson, R. E. & Hollyfield, J. G. *Inherited and environmentally induced retinal degenerations*: *proceedings of the International Symposium on Retinal Degenerations*, *held in San Francisco, California, September 2 and 3*, *1988*. (Liss, 1989).

57. Zheng L, Gong B, Hatala DA, Kern TS. Retinal ischemia and reperfusion causes capillary degeneration: similarities to diabetes. *Invest. Ophthalmol. Vis. Sci.* 2007;**48**:361–367. doi: 10.1167/iovs.06-0510.

58. Ueda K, et al. Retinal blood vessels are damaged in a rat model of NMDA-induced retinal degeneration. *Neurosci. Lett.* 2010;**485**:55–59. doi: 10.1016/j.neulet.2010.08.061.

59. Hann CR, Bentley MD, Vercnocke A, Ritman EL, Fautsch MP. Imaging the aqueous humor outflow pathway in human eyes by three-dimensional micro-computed tomography (3D micro-CT) *Exp. Eye Res.* 2011;**92**:104–111. doi: 10.1016/j.exer.2010.12.010.

60. Tymko MM, Boulet LM, Donnelly J. Intracranial pressure in outer space: preparing for the mission to Mars. *J. Physiol.* 2017;**595**:4587–4588. doi: 10.1113/JP274315.

61. Nelson ES, et al. The impact of ocular hemodynamics and intracranial pressure on intraocular pressure during acute gravitational changes. *J. Appl. Physiol.* 2017;**123**:352–363. doi: 10.1152/japplphysiol.00102.2017.

62. Andreev-Andrievskiy, A. A., Popova, A. S., Lagereva, E. A. & Vinogradova, O. L. Fluid shift versus body size: changes of hematological parameters and body fluid volume in hindlimb-unloaded mice, rats and rabbits. *J. Exp. Biol*. **221** (2018).

63. Benjamini Y, Hochberg Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple. *Testing. J. R. Stat. Soc. Series B Stat. Methodol.* 1995;**57**:289–300.

64. Fabregat Antonio, Jupe Steven, Matthews Lisa, Sidiropoulos Konstantinos, Gillespie Marc, Garapati Phani, Haw Robin, Jassal Bijay, Korninger Florian, May Bruce, Milacic Marija, Roca Corina Duenas, Rothfels Karen, Sevilla Cristoffer, Shamovsky Veronica, Shorser Solomon, Varusai Thawfeek, Viteri Guilherme, Weiser Joel, Wu Guanming, Stein Lincoln, Hermjakob Henning, D'Eustachio Peter. The Reactome Pathway Knowledgebase. *Nucleic Acids Research.* 2017;**46**(D1):D649–D655. doi: 10.1093/nar/gkx1132.

65. Rebhan M. GeneCards: integrating information about genes, proteins and diseases. *Trends in Genetics.* 1997;**13**:163. doi: 10.1016/S0168-9525(97)01103-7.

Chapter 4: A Standardized Pipeline for RNA-seq Data Analysis

Chapter 4 is adapted with minimal modification from:

NASA GeneLab RNA-Seq Consensus Pipeline: Standardized Processing of Short-Read RNA-Seq Data. Eliah G. Overbey, Amanda M. Saravia-Butler, *et al*. Under review at iScience.

4.1 Abstract

With the development of transcriptomic technologies, we are able to quantify precise changes in gene expression profiles from astronauts and other organisms exposed to spaceflight. Members of NASA GeneLab and GeneLab-associated analysis working groups (AWGs) have developed a consensus pipeline for analyzing short-read RNA-sequencing data from spaceflight-associated experiments. The pipeline includes quality control, read trimming, mapping, and gene quantification steps, culminating in the detection of differentially expressed genes. This data analysis pipeline and the results of its execution using data submitted to GeneLab are now all publicly available through the GeneLab database. We present here the full details and rationale for the construction of this pipeline in order to promote transparency, reproducibility and reusability of pipeline data, to provide a template for data processing of future spaceflight-relevant datasets, and to encourage cross-analysis of data from other databases with the data available in GeneLab.

4.2 Introduction

Opportunities to perform biological studies in space are rare due to high costs and a limited number of funding sources, rocket launches, and spaceflight crew hours for experimental procedures. Additionally, spaceflight research is decentralized and distributed across numerous laboratories in the United States and abroad. As a result, studies performed in different laboratories often utilize different organisms, strains, cell lines, and experimental procedures.

Adding to this complexity are variance in spaceflight factors and/or confounders within each study, such as degree of radiation exposure, experiment duration, CO_2 concentration, light cycle, and water availability, all of which can have effects on an organism's health and gene expression profiles during spaceflight (Rutter et al., n.d.). In order to optimize the integration of data from this diverse array of spaceflight experiments, it is paramount that variations in data processing are minimized.

There is presently no consensus on how best to analyze RNA-seq data and the impact of analysis tool selection on results is an active field of research. Indeed, selections of trimming parameters (Williams et al., 2016), read aligner (Yang et al., 2015), quantification tool (Teng et al., 2016), and differential expression detection algorithm (Costa-Silva, Domingues, and Lopes 2017) all affect results. Because of such challenges, groups like ENCODE and MINSEQE have developed standardized analysis pipelines for better comparison of RNA-seq datasets (ENCODE Project Consortium et al., 2020; "FGED: MINSEQE" n.d.).

The NASA GeneLab database (https://genelab-data.ndc.nasa.gov/genelab/projects) was created as a central repository for spaceflight-related omics-data. The repository includes data from experiments that profile transcription (RNA-seq, microarray), DNA/RNA methylation, protein expression, metabolite pools, and metagenomes. The most prevalent data type in this repository is RNA-seq from organisms exposed to spaceflight conditions. As of August 2020, the NASA GeneLab database has over eighty datasets with RNA-sequencing data (Sup Table S1). These datasets include *Homo sapiens* (human), *Mus musculus* (mouse), *Drosophila melanogaster* (fruit fly), *Arabidopsis thaliana* (model higher plant), *Oryzias latipes* (Japanese rice fish), *Helix lucorum* (land snail), *Brassica rapa* (Fast Plant®), *Eruca vesicaria* (arugula/edible plant), *Euprymna scolopes* (Hawaiian bobtail squid), *Ceratopteris richardii* (aquatic fern), and the bacterium, *Bacillus subtilis* from experiments performed during true spaceflight on various orbital platforms such as the Space Shuttle and International Space Station (ISS), as well as spaceflight-analog studies, such as hindlimb unloading and bed rest studies (Berrios et al., n.d.).

NASA's GeneLab and Ames Life Sciences Data Archive (ALSDA) projects have put forward an ambitious strategy focused on integrating data, metadata, and biospecimens to fully

utilize the 40+ years of archived NASA Life Sciences data (Scott et al., 2020). One of the first steps in this effort is the ability to analyze how experimental factors common to multiple datasets impact molecular signaling. Such meta analysis can only occur if metadata, data, and processed data are harmonized. As part of this strategy, GeneLab engaged with the scientific community and held its first Analysis Working Group (AWG) workshop in 2018. Spaceflight researchers from universities and organizations across the United States and abroad met to begin the creation of a standardized, consensus data-processing pipeline for one of the most common types of spaceflight datasets: transcription profiling via RNA-sequencing. Scientists at this workshop met to discuss the merits of various bioinformatic software tools for processing RNA-sequencing data, and ultimately agreed on a single pipeline of these tools.

The main driver for developing the consensus pipeline was to present consistently processed data to the public, therefore making space-relevant multi-omics data more accessible and reusable. The overall goals were: 1) To get more consistently processed data to the public; 2) To provide output data from every step of the consensus pipeline so users can download and use these "intermediate" data; 3) To support easier and more consistent analysis of space-relevant data by users including those in the NASA AWGs ; and 4) To allow easier cross-analysis of experiments to identify effects that result from the spaceflight environment, independent of confounding factors. In addition, many of these data in the GeneLab database have not been previously analyzed, as their generation was relatively recent. Therefore, providing new and processed datasets to the public allows biologists and others to more easily interpret these data, and contributes significantly to our collective knowledge of the effects of spaceflight on terrestrial organisms.

Here we present the RNA-seq consensus pipeline (RCP) developed by the GeneLab AWG along with the rationale behind the tool settings and options selected. The RCP includes three distinct steps: data pre-processing, data processing, and differential gene expression computation/annotation (Fig 1A]) These steps use tools for quality control (FastQC, MultiQC) (Andrews and Others 2010; Ewels et al., 2016), read trimming (TrimGalore) (Krueger 2019), mapping (STAR) (Dobin et al., 2013), quantification (RSEM) (B. Li and Dewey 2011), and differential gene expression calculation/annotation (DESeq2) (Love, Huber, and Anders 2014)

(Fig 1B). The RCP has been integrated into the GeneLab database and files produced by the RCP for each RNA-seq dataset hosted in GeneLab are and will continue to be publicly available for download.

4.3 Results

4.3.1 Data Pre-processing: Quality Control and Trimming

There are three distinct steps to the RCP, the first of which is data preprocessing (Fig 2A). The pipeline begins with quality control (QC) of raw FASTQ files from a short-read Illumina sequencer using the FastQC software (Andrews and Others 2010) (Fig 2B). FastQC is one of the most widely used QC programs for short-read sequencing data. It provides information which can be used to assess sample and sequencing quality, including base statistics, per base sequencing quality, per sequence quality scores, per base sequence content, per base GC content, per sequence GC content, per base N content, sequence length distributions, sequence duplication levels, overrepresented sequences and k-mer content.

The FastQC program is run on each individual sample file. However, reviewing the FastQC results for each sample file can be tedious and time consuming. Experiments typically have many sample files (biological and/or technical replicates) for multiple experimental conditions (spaceflight, ground control, etc). For this reason, we also use the MultiQC package (Ewels et al., 2016) (Fig 2C) to create a summary statistics report that includes the same quality control result categories from FastQC across all experiment samples.

After performing quality control on the raw FASTQ data, reads are trimmed using TrimGalore (Krueger 2019) to remove sequencing adapters that would disrupt read mapping during the data processing pipeline step (Fig 2D). Quality trimming is not performed as this has been shown to decrease the accuracy of quantification results (Williams et al., 2016). TrimGalore is a wrapper program that uses the cutadapt program (Martin 2011) for read trimming. TrimGalore was selected for the RCP due to its simplified command line interface, thorough output of trimming metrics, and ability to automatically detect adapters. In this step, bases that are part of a sequencing adapter are removed from each read and reads that become too short will

subsequently be removed. After trimming, the quality control programs, FastQC and MultiQC, are again run on the trimmed FASTQ files for viewing the quality control metrics of the reads that will be used for data processing. Once the data has been preprocessed, the sequenced reads are ready for mapping and quantification.

4.3.2 Data Processing: Read Mapping and Sample Quantification

In the data processing step (Fig 1; Step 2A), the trimmed reads are first aligned to the reference genome (Fig 3A) with STAR, a splice-aware aligner (Dobin et al., 2013). STAR must be run in two steps. The first step is to create indexed genome files (Fig 3B). These files are used to assist read mapping and only need to be generated once for each reference genome file. This step requires reference FASTA and GTF files (Sup Table S2). Some datasets include the External RNA Control Consortium (ERCC) spike-in control - a pool of 96 synthetic RNAs with various lengths and GC content covering a 2^{20} concentration range (Jiang et al., 2011). If ERCC spike-ins were included, the spike-in FASTA and GTF files are appended to the reference FASTA and GTF files, respectively. The second step of STAR mapping is to use the indexed reference genome and the trimmed reads from the preprocessing step in order to map the reads to the genome and the transcriptome (Fig 3C). STAR will also produce genome mapped data, which can optionally be used to find reads that map outside of annotated reference transcripts. STAR mapping output data are in BAM format, which has a separate entry for each mapped read and states which transcript each read mapped to. In order to improve the detection and quantification of splice sites, STAR is run in "two-pass mode". Here, splice sites are detected in the initial mapping to the reference and used to build a new reference that includes these splice sites. Reads are then re-mapped to this dynamically generated reference to improve the quantification of splice isoforms (Dobin et al., 2013).

The second part of processing is quantifying the number of reads mapped to each annotated transcript and gene (Fig 1A; Step 2B). For this task, the RCP uses RSEM (B. Li and Dewey 2011). The main reasons for using RSEM are its ability to account for reads that map to multiple transcripts and distinguish gene isoforms. In short-read sequencing experiments it is likely that some number of reads will map to multiple regions in the genome. RSEM computes

maximum likelihood abundance estimates to split the read count across multiple genes. Similar to STAR, RSEM is run in two distinct phases. The first phase uses the reference genome and GTF files (with or without ERCC as appropriate) (Sup Table S2) to prepare indexed genome files (Fig 4B]) The second phase uses the indexed files and the mapped reads from STAR to assign counts to each gene (Fig 4C). There are two output files generated for each sample: counts assigned to genes and counts assigned to isoforms. Gene counts are used to calculate differential gene expression. Isoform counts are also generated as an option to look at differential isoform expression, but are not used during differential gene expression calculation in the RCP. Once the RSEM count files are generated, the data are used to compute differentially expressed genes. A list of the reference genomes used in the GeneLab pipeline is available in Supplemental Table 2 (Sup Table 2). These reference genomes were the most recent releases at the time each STAR and RSEM indexed references were created. While it is possible to run STAR mapping through the RSEM toolkit, we elected not to do this because the alignment parameters used in this case are from ENCODE's STAR-RSEM pipeline and are not customizable. Thus we would have been precluded from using the precise mapping parameters agreed to by the GeneLab AWG.

We elected to adopt a mapping-based approach rather than rapidly quantifying the reads via a k-mer-based counting algorithm, pseudo-aligners, or a quasi-mapping method that utilizes RNA-seq inference procedures such as Kallisto (Bray et al., 2016) or Salmon (Patro et al., 2017) despite their speed advantages. This is because alignment-free quantification tools do not accurately quantify low-abundant and small RNAs especially when biological variation is present (Wu et al., 2018). Furthermore, alignment of reads allows for additional analyses beyond transcript and gene quantification such as measurement of gene body coverage and detection of novel transcripts.

There are several alignment-based mapping tools available and each has advantages and disadvantages. An alignment tool that is sensitive to splice-isoforms is critical to accurately identify how expression of splice-isoforms is affected by the spaceflight environment. DNA-specific aligners such as BWA (H. Li and Durbin 2009) and Bowtie (Langmead et al., 2009) cannot handle intron-sized gaps and thus an RNA-seq specific aligner is needed (Baruzzo et al., 2017). In addition to splice-awareness, when selecting an aligner the following criteria

were also considered: ability to input both single- and paired-end reads, handle strand-specific data, applicability to a variety of different model organisms with both low- and high-complexity genomic regions, efficient runtime and memory usage, ability to identify chimeric reads, high sensitivity, low rate of false discovery, and ability to output both genome and transcriptome alignments. Several studies have been conducted to compare the wide variety of available RNA-seq specific alignment tools, and of these, the STAR aligner consistently performs better than, or on par with the tools tested for the indicated criteria (Baruzzo et al., 2017; Schaarschmidt et al., 2020; Raplee, Evsikov, and Marín de Evsikova 2019).

4.3.3 Experimental Results: Differential Gene Expression Calculations and Addition of Gene Annotations

Once reads have been mapped and quantified, differential expression analysis is performed using the DESeq2 R package (Figs 1, 5A; Step 3). Unlike the previous steps, a custom R script(Fig 5B) is used to run DESeq2, to create both unnormalized and normalized counts tables, and to generate a differential gene expression (DGE) output table containing normalized counts for each sample, DGE results, and gene annotations. The GeneLab DGE R script also creates computer-readable tables that are used by the GeneLab visualization portal to generate various plots so users can easily view and begin interpreting the processed data. These scripts are provided in the NASA GeneLab_Data_Processing Github repository (https://github.com/nasa/GeneLab_Data_Processing). In the following sections we describe each step of these sections in order.

The GeneLab DGE R script requires three inputs: the quantified counts data from the previous (RSEM) step, sample metadata from the Investigation, Study, and Assay (ISA) tables in the ISA.zip file (provided in the GeneLab repository with each dataset) (Sansone et al., 2012; Rocca-Serra et al., 2010), and the organisms.csv file (Sup Table 3), which is used to specify the organism used in the study and relevant gene annotations to load. Since samples from some GeneLab RNA-seq datasets contain ERCC spike-in and others do not, there are two versions of the GeneLab DGE R script, one for datasets with ERCC spike-in and one for those without. Prior to running either script, paths to directories containing the input data and the output data location

must be defined. Each script starts by defining the organism used in the study, which should be consistent with the name in the organisms.csv file so that it matches the abbreviations used in the PANTHER database (Mi, Muruganujan, and Thomas 2013; Thomas 2003) for each organism. Next, the metadata from the ISA.zip file are imported and formatted for use with the DESeq2 package. During metadata formatting, groups for comparison are defined based on experimental factors and a sample table is created to specify the group to which each sample belongs. Next, a contrasts matrix is generated, which specifies the groups that will be compared during DGE analysis; each group is compared with every other group in a pairwise manner in both directions (i.e. spaceflight vs. ground and ground vs. spaceflight). This approach provides the user with the results for all possible group comparisons, allowing each user to select the most relevant comparisons for their particular scientific questions. After metadata formatting, the RSEM gene count data files from each sample are listed and re-ordered (to match the order the samples appear in the metadata), then imported with the R package, tximport (Soneson, Love, and Robinson, n.d.), and sample names are assigned. Prior to running DESeq2, a value of 1 is added to genes with lengths of zero, which is necessary to make a DESeqDataSet object. A DESeqDataSet object is then created using the formatted metadata and the count data that was imported with tximport.

For datasets that contain samples with ERCC spike-in, we use the GeneLab_DGE_wERCC.R script. To reduce the possibility of skewing the data during DESeq2 normalization (McIntyre et al., 2011; Risso et al., 2011; Conesa et al., 2016; Law et al., 2016), all genes that have a sum of less than 10 counts across all samples are removed. The cutoff value of 10 is a best-practice recommended by the DESeq2 tutorial on Bioconductor. These filtered data are then prepared for normalization and DGE analysis with DESeq2. Since there is no consensus for whether or not ERCC-normalization improves the accuracy of the results (Risso et al., 2014), the GeneLab project and its AWG members decided to perform the DGE analysis both with and without ERCC-normalization (for datasets with samples containing ERCC spike-in).

To enable DESeq2 analysis with and without considering ERCC reads, the DESeqDataSet object is used to create a DESeqDataSet object containing only ERCC reads. Since all samples must contain ERCC spike-in for ERCC-normalization, the DESeqDataSet

object containing only ERCC reads is used to identify and remove any samples that do not contain ERCC reads. Next, a DESeqDataSet object containing only non-ERCC reads is created by removing rows containing ERCC reads. These data are then used for DESeq2 analysis.

For DESeq2 analysis with ERCC-normalization, the size factor object of the non-ERCC data is replaced with ERCC size factors for re-scaling in the first DESeq2 step. For DESeq2 analysis without ERCC-normalization, the DESeq2 default algorithm is applied to the DESeqDataSet object containing only non-ERCC reads. The unnormalized and DESeq2-normalized counts data as well as the sample table are then outputted as CSV files. The 'Unnormalized_Counts.csv', 'Normalized_Counts.csv', and 'ERCC_Normalized_Counts.csv' files for each RNA-seq dataset are available in the GeneLab Data Repository; the 'SampleTable.csv' file is used internally for verifying and validating the processed data prior to publication.

There are two types of hypothesis tests that can be run with DESeq2, the likelihood ratio test (LRT), which is similar to an analysis of variance (ANOVA) calculation in linear regression and allows for comparison across all groups, and the Wald test, in which the estimated standard error of a log2 fold change is used to compare differences between two groups. The DGE step of the RCP performs both of these analyses. After normalization, the DESeq2 likelihood ratio test design is applied to the normalized data (both ERCC- and nonERCC-normalized data) to generate the F statistic p-value, which is similar to an ANOVA p-value and reveals genes that are changed in any number of combinations of all factors defined in the experiment.

To prepare for building a gene/pathway annotation database, the STRINGdb (Szklarczyk et al., 2019) and PANTHER.db (Thomas 2003) libraries are loaded and the organisms.csv file is read and used to indicate the Bioconductor AnnotationData Package needed (Huber et al., 2015; Gentleman et al., 2004). The current gene annotation database for the organism specified at the beginning of the R script is then loaded.

Next, DGE tables containing normalized counts for each sample, pairwise DGE results, and current gene annotations as well as computer-readable DGE tables (that will be used for visualization) are created first with nonERCC-normalized data and then with ERCC-normalized data. For pairwise DGE analysis, first normalized count data are used to create two output tables,

one that is used to create the human-readable DGE output table provided to users with processed data for each dataset, and the respective computer-readable DGE output table that contains additional columns and is used to visualize the data. Next, normalized count data are iterated through Wald Tests to generate pairwise comparisons of all groups based on the contrasts matrix that was generated during metadata formatting. The pairwise DGE analysis results are then added as columns to both DGE output tables.

Then an annotation database is built by first defining the "keytype", which indicates the primary type of annotation used (for most GeneLab datasets this is ENSEMBL). The keytype is then used to map to annotations in the organism-specific Bioconductor AnnotationData Package, and the following annotation columns are added to the annotation database: SYMBOL, GENENAME, ENSEMBL (if not the primary), REFSEQ, and ENTREZID. STRING and GOSLIM annotation columns are also added to the annotation database using the STRINGdb and PANTHER.db R packages, respectively. All of the aforementioned annotation columns are added to the annotation database to enable users to perform downstream analyses without having to map gene IDs themselves. Once the annotation database is complete, additional calculations are performed on the normalized count data before assembling the final DGE output tables.

Means and standard deviations of normalized count data for each gene across all samples, and for samples within each respective group, are calculated and added as columns to the DGE output tables. A column containing the F statistic p-value, calculated previously, is also added to the DGE output tables. The following columns are added only to the computer-readable DGE output table (used for visualization): a column to indicate whether each gene (or pathway) is up- or down-regulated for each pairwise comparison, a column to indicate genes that are differentially expressed using a p-value cutoff of ≤ 0.1 and another column using a p-value cutoff of ≤ 0.05, a column indicating the log2 of the p-value for each pairwise comparison and another column indicating the log2 of the adjusted p-value, both of which are used to create Volcano plots. After all columns are added to the DGE tables, both the human- and computer-readable DGE tables are combined with the current annotation database to create the complete human- and computer-readable DGE tables. Principal component analysis (PCA) is also performed on the normalized count data and used to create PCA plots for the GeneLab data visualization

portal. DGE analysis of datasets without ERCC spike-in is performed exactly the same way as the nonERCC-normalized approach described above, except that no ERCC reads have to be removed from the DESeqDataSet object prior to DESeq analysis.

Both the GeneLab_DGE_wERCC.R and the GeneLab_DGE_noERCC.R scripts produce the following output files: Unnormalized_Counts.csv (*), Normalized_Counts.csv (*), SampleTable.csv (#), contrasts.csv (*), differential_expression.csv (*), visualization_output_table.csv (**), visualization_PCA_table.csv (**) (Fig 5B). The GeneLab_DGE_wERCC.R script will also produce the following additional output files: ERCC_rawCounts_unfiltered.csv (#), ERCC_rawCounts_filtered.csv (#), ERCCnorm_contrasts.csv (*), ERCC_Normalized_Counts.csv (*), ERCCnorm_differential_expression.csv (*), visualization_output_table_ERCCnorm.csv (**), visualization_PCA_table_ERCCnorm.csv (**) (Fig 5B).

4.4 Discussion

The differentially expressed genes calculated by the RCP can be further explored with a variety of tools designed for higher-order analysis. For example, there are tools which can look for enriched pathways, gene ontology terms, or protein and/or metabolite networks. Popular software tools among the GeneLab working group members include WebGestalt (Liao et al., 2019), STRING (Szklarczyk et al., 2019), GSEA (Subramanian et al., 2005), PIANO (Väremo, Nielsen, and Nookaew 2013), Reactome (Szklarczyk et al. 2019), and ToppFun (Chen et al. 2009). There is no universal consensus on which tools are the most useful for higher-order analysis (Nguyen et al., 2019). RCP users are encouraged to try multiple tools in order to analyze their data from a variety of perspectives.

The RCP has been designed to handle sequencing experiments that either lack or include the ERCC RNA spike-in mix - a set of 96 polyadenylated RNAs that can be used during differential gene expression calculation to normalize read counts across samples (Munro et al., 2014). However, the use of normalization according to ERCC spike-ins remains controversial among AWG members, and Munro *et al.,* suggested its usage only for determining limit of

detection of ratio (LODR), expression ratio variability and measurement bias (Munro et al., 2014). For this reason, ERCC normalization remains optional in the GeneLab pipeline and both kinds of DGE outputs are provided in the GeneLab database.

A high number of biological replicates can increase certainty in the differentially expressed genes determined by the RCP. However conducting experiments in spaceflight often limits the number of biological replicates that a researcher can include. Therefore it is important to note that at least three biological replicates are required for the pipeline, specifically for DESeq2, to perform its statistical methods. However, at least six replicates are suggested in order to minimize the false discovery rate (FDR) (Schurch et al., 2016). Finally, RNA-seq datasets hosted on GeneLab that do not contain biological replicates are only processed up until unnormalized (raw) counts are obtained, the step right before DESeq2 is used for DGE calculation.

More advanced RCP users might have additional data inquiries that fall beyond the scope of this pipeline. For this reason, there are two parts of the pipeline that include additional output that are not used in our differential gene expression computation. The first is in the output from STAR, mapping output is also provided in genomic coordinates. This is useful for obtaining reads that are mapped outside of the reference transcriptome. For example, this may be used to find novel genes, transcripts, or exons that have not yet been annotated by consortiums. The second part of the pipeline with alternative output files is RSEM. This also provides transcript-level counts which can be used to investigate differential isoform expression. Moreover, intermediate files are provided as outputs to allow users to use components of the pipeline that they find useful.

The GeneLab database also includes other types of transcriptomic data. As discussed in this article, the RCP is not used for microarray data which are fundamentally different, and the AWG is still debating the best approach for cross-dataset comparisons between microarrays. GeneLab also accepts data from long read experiments, such as those produced by Pacific Biosciences' (PacBio) single-molecule real-time (SMRT) sequencing (R. J. Roberts, Carneiro, and Schatz 2013) and Oxford Nanopore Technologies' (ONT) nanopore sequencing (Jain et al., 2016). Long-read data would be processed with similar steps to the RCP, but will require tools

specifically designed for the intricacies of long-read data, such as reads that contain multiple splice junctions and reads which currently have a higher base-calling error-rate. Currently, long-reads are typically used for DNA sequencing and were recently highlighted on board of the ISS using ONT for de novo assembly of the *E. coli* genome from raw reads (Castro-Wallace et al., 2017). However, even though throughput and accuracy remain far inferior to short-reads, long-reads offer some advantages for RNAseq as well, with less ambiguity for genes and isoforms detection, much faster mapping, potential identification of genes not yet known from reference genomes and eventually less bias in DGE.

To conclude, the RCP is specifically designed for RNA-seq data from short-read sequencers and has been developed in order to encourage and facilitate analysis of spaceflight multi-omic data. The creation of the RCP by a large community of scientists (GeneLab AWG: https://genelab.nasa.gov/awg) and the sharing of pipeline details in a peer-reviewed article provide analysis transparency and enable data reproducibility.

4.5 Methods

The tools used in the consensus pipeline are documented in Supplemental Table 4: Pipeline Tools and Links (Sup Table 4). Due to NASA security requirements, all software is updated monthly with security patching. Therefore, tool versions used to process each RNA-seq dataset hosted on the GeneLab Data Repository are provided in the RNA-seq protocol section. Specific commands, options, and flags for each tool used in the RCP are reported in the figures of the main text. Note that some packages listed here are dependencies of the packages used in the RCP. More information about such dependencies can be found in the tool documentation.

This pipeline has been run on short-read RNA-seq data in the GeneLab database (https://genelab-data.ndc.nasa.gov/genelab/projects) and is applied to new submissions to the database. Any updates to the software used in the pipeline will be noted in the Github repository GeneLab_Data_Processing (https://github.com/nasa/GeneLab_Data_Processing).

4.6 Figures

Figure 1

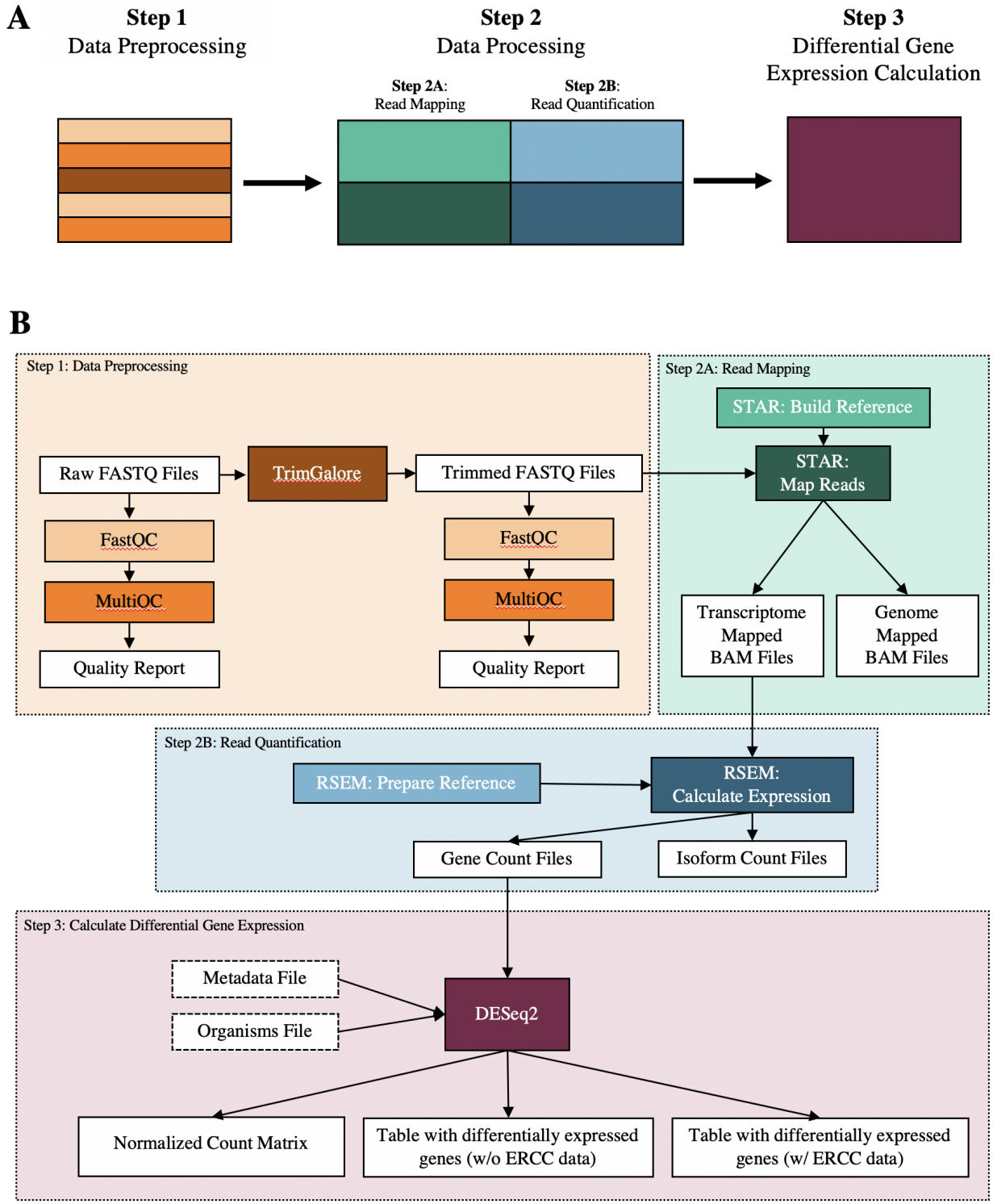
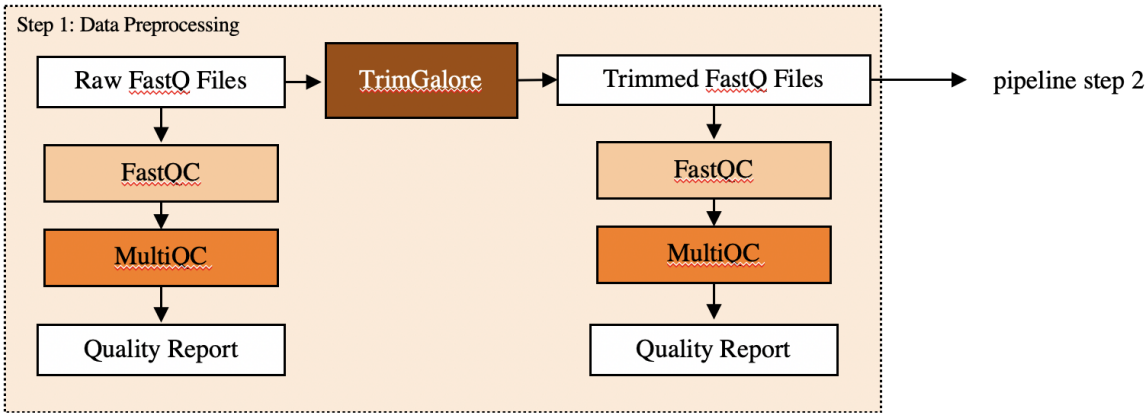


Figure 1: GeneLab RNA-seq Consensus Pipeline (RCP). (A) The three broad steps of the RCP. The RCP handles: 1) Data preprocessing to trim sequencing adapters and to provide quality control metrics; 2) Data processing to map reads to the reference genome and quantify the number of read counts per gene; and 3) Differential gene expression calculation, which will provide a list of differentially expressed genes that can be sorted by adjusted p-value and log fold-change. **(B)** The full RCP annotated with tools, input files, and output files.

Figure 2

A



B

FastQC	
Parameters	fastqc -o /path/to/output/directory \ -t number_of_threads \ /path/to/input/files
Input data files	fastq.gz
Output data files	fastqc.html (FastQC report) fastqc.zip (FastQC raw data)

C

MultiQC	
Parameters	multiqc -o /path/to/output/directory \ /path/to/fastqc/output/files
Input data files	fastqc.html (FastQC report) fastqc.zip (FastQC raw data)
Output data files	multiqc_report.html (multiqc report) multiqc_data (directory containing multiqc raw data)

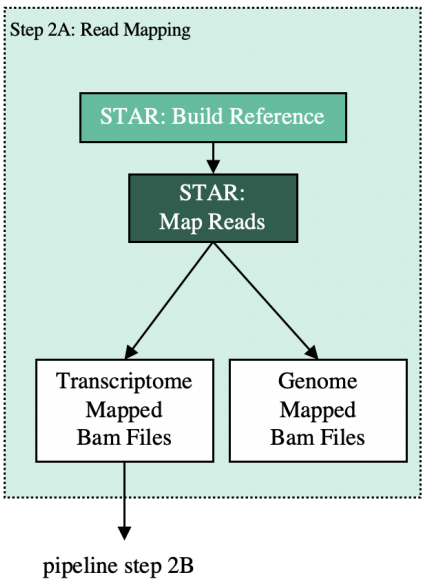
D

TrimGalore	
Parameters	trim_galore --gzip \ --path_to_cutadapt /path/to/cutadapt \ --phred33 \ --illumina \ # if adapters are not illumina, replace with adapters used --output_dir /path/to/TrimGalore/output/directory \ --paired \ # only for PE studies /path/to/forward/reads /path/to/reverse/reads # if SE, replace the last line with only /path/to/forward/reads
Input data files	*fastq.gz (raw reads)
Output data files	*fastq.gz (trimmed reads) *trimming_report.txt (trimming report)

Figure 2: Data preprocessing (pipeline step 1): Quality control and trimming. (A) Data Preprocessing pipeline. FastQ files from Illumina base-calling software are quality checked using FastQC and MultiQC. Data is then trimmed using TrimGalore and are re-checked for quality; **(B)** Flags used for FastQC program; **(C)** Flags used for MultiQC program; **(D)** Flags used for TrimGalore program; trimmed reads (*fastq.gz) are then used as input data for FastQC (B) followed by MultiQC (C) to generate trimmed read quality metrics. Tool versions used to process each dataset are included in the RNA-seq processing protocol in the GLDS Repository.

Figure 3

A



B

STAR	Build Reference	
Parameters	STAR --runThreadN NumberOfThreads \ # Number of available cores on server node --runMode genomeGenerate \ --limitGenomeGenerateRAM 35000000000 \ # min needed for mouse ref --genomeDir /path/to/STAR/genome/directory \ --genomeFastaFiles /path/to/genome/fasta/file \ --sjdbGTFfile /path/to/annotation/gtf/file \ --sjdbOverhang ReadLength-1	
Input data files	.fasta (genome sequences) .gtf (genome annotation) Note that for analyzing samples with ERCC spike-in, the genome fasta and gtf files should have the ERCC fasta and gtf files appended them.	
Output data files	The following will be generated in the /STAR/genome/directory: • chrLength.txt • chrNameLength.txt • chrName.txt • chrStart.txt • exonGeTrInfo.tab • exonInfo.tab	• geneInfo.tab • Genome • genomeParameters.txt • SA • SAindex • sjdbInfo.txt • sjdbList.fromGTF.out.tab • sjdbList.out.tab • transcriptInfo.tab

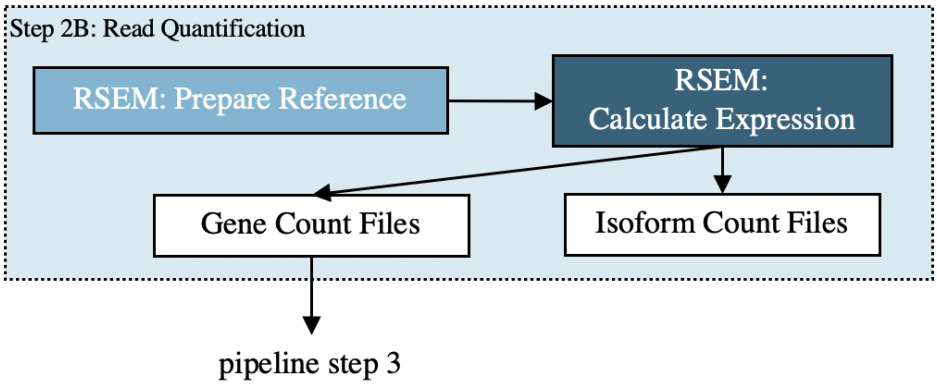
C

STAR	Map Reads	
Parameters	STAR --twopassMode Basic \ --limitBAMsortRAM available_memory_in_bytes \ --genomeDir /path/to/STAR/genome/directory \ --outSAMunmapped Within \ --outFilterType BySJout \ --outSAMattributes NH HI AS NM MD MC \ --outFilterMultimapNmax 20 \ --outFilterMismatchNmax 999 \ --outFilterMismatchNoverReadLmax 0.04 \ --alignIntronMin 20 \ --alignIntronMax 1000000 \ --alignMatesGapMax 1000000 \ # only needed for PE studies --alignSJoverhangMin 8 \ --alignSJDBoverhangMin 1 \ --sjdbScore 1 \ --readFilesCommand zcat \ --runThreadN NumberOfThreads \ --outSAMtype BAM SortedByCoordinate \ --quantMode TranscriptomeSAM \ --outSAMheaderHD @HD VN:1.4 SO:coordinate \ --outFileNamePrefix /path/to/STAR-output/directory/<sample_name> \ --readFilesIn /path/to/trimmed_forward_reads \ /path/to/trimmed_reverse_reads # only needed for PE studies	
Input data files	STAR index directory *fastq.gz (trimmed reads)	
Output data files	Files *Aligned.sortedByCoord.out.bam (sorted mapping to genome) *Aligned.toTranscriptome.out.bam (sorted mapping to transcriptome) *Log.final.out (reads mapped, etc) *Log.out *Log.progress.out *SJ.out.tab	Directories *_STARgenome sjdbInfo.txt sjdbList.out.tab *_STARpass1 Log.final.out SJ.out.tab *_STARtmp directory containing subdirectories that are empty – this was the location for temp files that were automatically removed after successful completion

Figure 3: Data processing (pipeline step 2A): Read mapping. (A) Data processing pipeline. Trimmed reads are mapped to their reference genome and transcriptome with STAR. Gene counts are then quantified with RSEM; **(B)** Flags used for generating the indexed STAR reference files; **(C)** Flags used for mapping reads with STAR. Tool versions used to process each dataset are included in the RNA-seq processing protocol in the GLDS Repository.

Figure 4

A



B

RSEM	Prepare Reference
Parameters	rsem-prepare-reference --gtf /path/to/annotation/gtf/file \ /path/to/genome/fasta/file \ /path/to/RSEM/genome/directory/RSEM_ref_prefix
Input data files	.fasta (genome sequence) .gtf (genome annotation) Note that for analyzing samples with ERCC spike-in, the genome FASTA and GTF files should have the ERCC FASTA and GTF files appended to them.
Output data files	The following will be generated in the /RSEM/genome/directory RSEM_ref_prefix.chrlist RSEM_ref_prefix.grp RSEM_ref_prefix.idx.fa RSEM_ref_prefix.n2g.idx.fa RSEM_ref_prefix.seq RSEM_ref_prefix.ti RSEM_ref_prefix.transcripts.fa

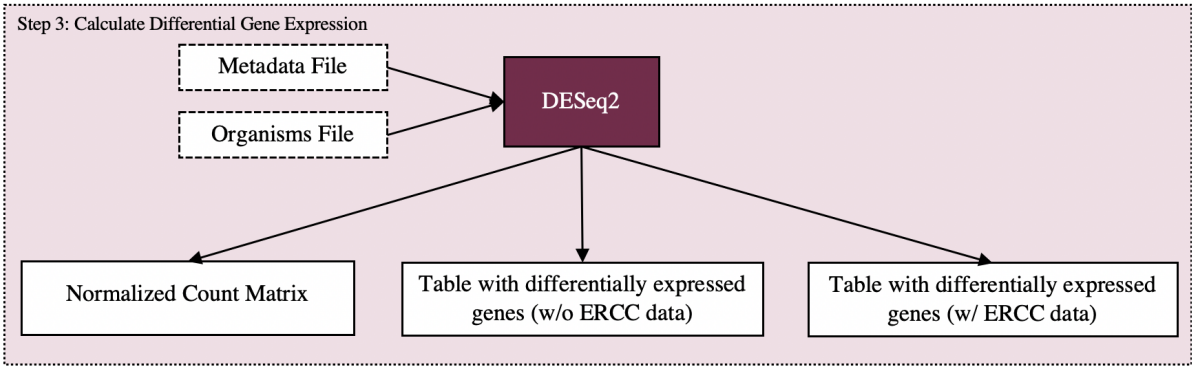
C

RSEM	Calculate Expression
Parameters	rsem-calculate-expression --num-threads NumberOfThreads \ --alignments \ --bam \ --paired-end \ # only for PE studies --seed 12345 \ --estimate-rspd \ --no-bam-output \ --strandedness reverse \ # For Illumina TruSeq stranded protocols, reads are derived from the reverse strand /path/to/*Aligned.toTranscriptome.out.bam \ /path/to/RSEM/genome/directory/RSEM_ref_prefix \ /path/to/RSEM/output/directory
Input data files	*Aligned.toTranscriptome.out.bam RSEM indexed reference
Output data files	*genes.results (counts per gene) *isoforms.results (counts per isoform) *stat (directory containing the following stats files) *cnt *model *theta

Figure 4: Data processing (pipeline step 2B): Gene quantification. (A) Data processing pipeline. Mapping results from STAR are quantified by RSEM; **(B)** Parameters for RSEM indexed reference files generation; **(C)** Parameters for quantifying gene and isoform counts with RSEM. Tool versions used to process each dataset are included in the RNA-seq processing protocol in the GLDS Repository.

Figure 5

A



B

Output Type	Output File Name (GeneLab_DGE_noERCC.R and GeneLab_DGE_wERCC.R)	Output File Name (GeneLab_DGE_wERCC.R only)
Available with RNA-seq processed data in the GLDS Repository (*)	Unnormalized_Counts.csv Normalized_Counts.csv contrasts.csv differential_expression.csv	ERCCnorm_contrasts.csv ERCC_Normalized_Counts.csv ERCCnorm_differential_expression.csv
Used to generate interactive plots from RNA-seq processed data in the GLDS visualization portal (**)	visualization_output_table.csv visualization_PCA_table.csv	visualization_output_table_ERCCnorm.csv visualization_PCA_table_ERCCnorm.csv
Used for internal QC and/or V&V (#)	SampleTable.csv	ERCC_rawCounts_unfiltered.csv ERCC_rawCounts_filtered.csv

Figure 5: Differential gene expression calculation (pipeline step 3). (A) Data processing pipeline. The R program DESeq2 is run in order to determine which genes are differentially expressed between experimental conditions using gene count files from RSEM. **(B)** Output files generated. The table columns distinguish which script produces each output. The columns distinguish how those output files are used.

4.7 Supplemental Tables and Scripts

Supplemental Tables S1-S4 and Supplemental Scripts S1-2 are in Appendix A: Chapter 4 Supplemental Tables and Scripts

4.8 Author Contributions

All authors developed the initial analysis scheme at the 2018 GeneLab AWG workshop. EGO, ZZ, KSR, HF, WAdS, RB and JMG refined this into a draft pipeline. EGO and AMSB wrote and

validated the final processing scripts. EGO and AMSB wrote the original manuscript draft and generated figures. All authors reviewed and edited the final draft.

4.9 References

Anand, Deepti, and Salil A. Lachke. 2017. "Systems Biology of Lens Development: A Paradigm for Disease Gene Discovery in the Eye." *Experimental Eye Research*. https://doi.org/10.1016/j.exer.2016.03.010.

Andreev-Andrievskiy, Alexander A., Anfisa S. Popova, Evgeniia A. Lagereva, and Olga L. Vinogradova. 2018. "Fluid Shift versus Body Size: Changes of Hematological Parameters and Body Fluid Volume in Hindlimb-Unloaded Mice, Rats and Rabbits." *The Journal of Experimental Biology* 221 (Pt 17). https://doi.org/10.1242/jeb.182832.

Andrews, Simon, and Others. 2010. "FastQC: A Quality Control Tool for High Throughput Sequence Data." Babraham Bioinformatics, Babraham Institute, Cambridge, United Kingdom.

Assenov, Yassen, Fabian Müller, Pavlo Lutsik, Jörn Walter, Thomas Lengauer, and Christoph Bock. 2014. "Comprehensive Analysis of DNA Methylation Data with RnBeads." *Nature Methods*. https://doi.org/10.1038/nmeth.3115.

Awan, M. Umer Farooq, Murtaza Hasan, Javed Iqbal, Runhong Lei, Wang Fu Lee, Ma Hong, Hong Qing, and Yulin Deng. 2014. "Neuroprotective Role of BNIP3 under Oxidative Stress through Autophagy in Neuroblastoma Cells." *Molecular Biology Reports* 41 (9): 5729–34.

Banday, Abdul Rouf, Marybeth Baumgartner, Sahar Al Seesi, Devi Krishna Priya Karunakaran, Aditya Venkatesh, Sean Congdon, Christopher Lemoine, et al. 2014. "Replication-Dependent Histone Genes Are Actively Transcribed in Differentiating and Aging Retinal Neurons." *Cell Cycle* 13 (16): 2526–41.

Barben, Maya, Divya Ail, Federica Storti, Katrin Klee, Christian Schori, Marijana Samardzija, Stylianos Michalakis, et al. 2018. "Hif1a Inactivation Rescues Photoreceptor Degeneration Induced by a Chronic Hypoxia-like Stress." *Cell Death and Differentiation* 25 (12): 2071–85.

Baruzzo, Giacomo, Katharina E. Hayer, Eun Ji Kim, Barbara Di Camillo, Garret A. FitzGerald, and Gregory R. Grant. 2017. "Simulation-Based Comprehensive Benchmarking of RNA-Seq Aligners." *Nature Methods* 14 (2): 135–39.

Beisel, Nicole S., Jerald Noble, W. Brad Barbazuk, Anna-Lisa Paul, and Robert J. Ferl. 2019. "Spaceflight-Induced Alternative Splicing during Seedling Development in Arabidopsis Thaliana." *NPJ Microgravity* 5 (April): 9.

Bejarano-Escobar, Ruth, Guadalupe Álvarez-Hernán, Ruth Morona, Agustín González, Gervasio Martín-Partido, and Javier Francisco-Morcillo. 2015. "Expression and Function of the LIM-Homeodomain Transcription Factor Islet-1 in the Developing and Mature Vertebrate Retina." *Experimental Eye Research* 138 (September): 22–31.

Bellone, John A., Peter S. Gifford, Nina C. Nishiyama, Richard E. Hartman, and Xiao Wen Mao. 2016. "Long-Term Effects of Simulated Microgravity And/or Chronic Exposure to Low-Dose Gamma Radiation on Behavior and Blood–brain Barrier Integrity." *Npj Microgravity* 2 (1): 16019.

Benjamini, Yoav, and Yosef Hochberg. 1995. "Controlling the False Discovery Rate: A Practical

and Powerful Approach to Multiple Testing." *Journal of the Royal Statistical Society. Series B, Statistical Methodology* 57 (1): 289–300.
Berrios, Daniel, Jonathan Galazka, Kirill Gorev, Samrawit Gebre, and Sylvain Costes. n.d. "Interfaces for the Exploration of Space Omics Data." *Nucleic Acids Research*.
Besharse, J. C., D. A. Dunis, and B. Burnside. 1982. "Effects of Cyclic Adenosine 3',5'-Monophosphate on Photoreceptor Disc Shedding and Retinomotor Movement. Inhibition of Rod Shedding and Stimulation of Cone Elongation." *The Journal of General Physiology* 79 (5): 775–90.
Bray, Nicolas L., Harold Pimentel, Páll Melsted, and Lior Pachter. 2016. "Near-Optimal Probabilistic RNA-Seq Quantification." *Nature Biotechnology* 34 (5): 525–27.
Byrne, Leah C., Deniz Dalkara, Gabriel Luna, Steven K. Fisher, Emmanuelle Clérin, Jose-Alain Sahel, Thierry Léveillard, and John G. Flannery. 2015. "Viral-Mediated RdCVF and RdCVFL Expression Protects Cone and Rod Photoreceptors in Retinal Degeneration." *The Journal of Clinical Investigation* 125 (1): 105–16.
Castro-Wallace, Sarah L., Charles Y. Chiu, Kristen K. John, Sarah E. Stahl, Kathleen H. Rubins, Alexa B. R. McIntyre, Jason P. Dworkin, et al. 2017. "Nanopore DNA Sequencing and Genome Assembly on the International Space Station." *Scientific Reports* 7 (1): 18022.
Chancellor, Jeffery C., Graham B. I. Scott, and Jeffrey P. Sutton. 2014. "Space Radiation: The Number One Risk to Astronaut Health beyond Low Earth Orbit." *Life* 4 (3): 491–510.
Chatterjee, Shampa, Ralph A. Pietrofesa, Kyewon Park, Jian-Qin Tao, Alejandro Carabe-Fernandez, Abigail T. Berman, Constantinos Koumenis, Thais Sielecki, and Melpo Christofidou-Solomidou. 2019. "LGM2605 Reduces Space Radiation-Induced NLRP3 Inflammasome Activation and Damage in In Vitro Lung Vascular Networks." *International Journal of Molecular Sciences* 20 (1). https://doi.org/10.3390/ijms20010176.
Chen, Jing, Eric E. Bardes, Bruce J. Aronow, and Anil G. Jegga. 2009. "ToppGene Suite for Gene List Enrichment Analysis and Candidate Gene Prioritization." *Nucleic Acids Research* 37 (Web Server issue): W305–11.
Ciechanover, Aaron, and Yong Tae Kwon. 2017. "Protein Quality Control by Molecular Chaperones in Neurodegeneration." *Frontiers in Neuroscience* 11 (April): 185.
Conesa, Ana, Pedro Madrigal, Sonia Tarazona, David Gomez-Cabrero, Alejandra Cervera, Andrew McPherson, Michał Wojciech Szcześniak, et al. 2016. "A Survey of Best Practices for RNA-Seq Data Analysis." *Genome Biology* 17 (January): 13.
Corydon, Thomas J., Vivek Mann, Lasse Slumstrup, Sascha Kopp, Jayashree Sahana, Anne Louise Askou, Nils E. Magnusson, et al. 2016. "Reduced Expression of Cytoskeletal and Extracellular Matrix Genes in Human Adult Retinal Pigment Epithelium Cells Exposed to Simulated Microgravity." *Cellular Physiology and Biochemistry: International Journal of Experimental Cellular Physiology, Biochemistry, and Pharmacology* 40 (1-2): 1–17.
Costa-Silva, Juliana, Douglas Domingues, and Fabricio Martins Lopes. 2017. "RNA-Seq Differential Expression Analysis: An Extended Review and a Software Tool." *PloS One* 12 (12): e0190152.
Cronin, T., W. Raffelsberger, I. Lee-Rivera, C. Jaillard, M-L Niepon, B. Kinzel, E. Clérin, et al. 2010. "The Disruption of the Rod-Derived Cone Viability Gene Leads to Photoreceptor Dysfunction and Susceptibility to Oxidative Stress." *Cell Death and Differentiation* 17 (7): 1199–1210.
Cucinotta, Francis A., and Eliedonna Cacao. 2019. "Risks of Cognitive Detriments after Low

Dose Heavy Ion and Proton Exposures." *International Journal of Radiation Biology* 95 (7): 985–98.
Dawson, Linda. 2017. "The New Space Race." *The Politics and Perils of Space Exploration*. https://doi.org/10.1007/978-3-319-38813-7_1.
De la Torre, Gabriel G. 2014. "Cognitive Neuroscience in Space." *Life* 4 (3): 281–94.
Demontis, Gian C., Marco M. Germani, Enrico G. Caiani, Ivana Barravecchia, Claudio Passino, and Debora Angeloni. 2017. "Human Pathophysiological Adaptations to the Space Environment." *Frontiers in Physiology* 8 (August): 547.
Dobin, Alexander, Carrie A. Davis, Felix Schlesinger, Jorg Drenkow, Chris Zaleski, Sonali Jha, Philippe Batut, Mark Chaisson, and Thomas R. Gingeras. 2013. "STAR: Ultrafast Universal RNA-Seq Aligner." *Bioinformatics* 29 (1): 15–21.
ENCODE Project Consortium, Michael P. Snyder, Thomas R. Gingeras, Jill E. Moore, Zhiping Weng, Mark B. Gerstein, Bing Ren, et al. 2020. "Perspectives on ENCODE." *Nature* 583 (7818): 693–98.
Ewels, Philip, Måns Magnusson, Sverker Lundin, and Max Käller. 2016. "MultiQC: Summarize Analysis Results for Multiple Tools and Samples in a Single Report." *Bioinformatics* 32 (19): 3047–48.
Fabregat, Antonio, Steven Jupe, Lisa Matthews, Konstantinos Sidiropoulos, Marc Gillespie, Phani Garapati, Robin Haw, et al. 2018. "The Reactome Pathway Knowledgebase." *Nucleic Acids Research* 46 (D1): D649–55.
Fassina, Gianfranco, Maria Grazia Aluigi, Susan Gentleman, Paul Wong, Tania Cai, Adriana Albini, and Douglas M. Noonan. 1997. "The cAMP Analog 8-Cl-cAMP Inhibits Growth and Induces Differentiation and Apoptosis in Retinoblastoma Cells." *International Journal of Cancer*. https://doi.org/3.0.co;2-#">10.1002/(sici)1097-0215(19970917)72:6<1088::aid-ijc25>3.0.co;2-#.
"FGED: MINSEQE." n.d. Accessed September 4, 2020. http://fged.org/projects/minseqe/.
Garrett-Bakelman, Francine E., Manjula Darshi, Stefan J. Green, Ruben C. Gur, Ling Lin, Brandon R. Macias, Miles J. McKenna, et al. 2019. "The NASA Twins Study: A Multidimensional Analysis of a Year-Long Human Spaceflight." *Science* 364 (6436). https://doi.org/10.1126/science.aau8650.
Gentleman, Robert C., Vincent J. Carey, Douglas M. Bates, Ben Bolstad, Marcel Dettling, Sandrine Dudoit, Byron Ellis, et al. 2004. "Bioconductor: Open Software Development for Computational Biology and Bioinformatics." *Genome Biology* 5 (10): R80.
"GLDS-202: Low Dose (0.04 Gy) Irradiation (LDR) and Hindlimb Unloading (HLU) Microgravity in Mice: Brain Transcriptomic and Epigenomic Data." n.d. NASA GeneLab Data Systems. Accessed September 24, 2020. https://genelab-data.ndc.nasa.gov/genelab/accession/GLDS-202/.
Grammaticos, Philip, Evanthia Giannoula, and George P. Fountos. 2013. "Acute Radiation Syndrome and Chronic Radiation Syndrome." *Hellenic Journal of Nuclear Medicine* 16 (1): 56–59.
Guoguo, Shang, Takaaki Akaike, Jiang Tao, Chen Qi, Zhang Nong, and Li Hui. 2012. "HGF-Mediated Inhibition of Oxidative Stress by 8-Nitro-cGMP in High Glucose-Treated Rat Mesangial Cells." *Free Radical Research* 46 (10): 1238–48.
Hann, Cheryl R., Michael D. Bentley, Andrew Vercnocke, Erik L. Ritman, and Michael P.

Fautsch. 2011. “Imaging the Aqueous Humor Outflow Pathway in Human Eyes by Three-Dimensional Micro-Computed Tomography (3D Micro-CT).” *Experimental Eye Research*. https://doi.org/10.1016/j.exer.2010.12.010.

Hernan, Imma, María José Gamundi, Ester Planas, Emma Borràs, Miquel Maseras, and Miguel Carballo. 2011. “Cellular Expression and siRNA-Mediated Interference of Rhodopsin Cis-Acting Splicing Mutants Associated with Autosomal Dominant Retinitis Pigmentosa.” *Investigative Ophthalmology & Visual Science* 52 (6): 3723–29.

Huber, Wolfgang, Vincent J. Carey, Robert Gentleman, Simon Anders, Marc Carlson, Benilton S. Carvalho, Hector Corrada Bravo, et al. 2015. “Orchestrating High-Throughput Genomic Analysis with Bioconductor.” *Nature Methods* 12 (2): 115–21.

Iida, Atsumi, Toshiro Iwagawa, Yukihiro Baba, Shinya Satoh, Yujin Mochizuki, Hiromitsu Nakauchi, Takahisa Furukawa, Haruhiko Koseki, Akira Murakami, and Sumiko Watanabe. 2015. “Roles of Histone H3K27 Trimethylase Ezh2 in Retinal Proliferation and Differentiation.” *Developmental Neurobiology*. https://doi.org/10.1002/dneu.22261.

Ikeshima-Kataoka, Hiroko. 2016. “Neuroimmunological Implications of AQP4 in Astrocytes.” *International Journal of Molecular Sciences* 17 (8). https://doi.org/10.3390/ijms17081306.

Jackson, Chad R., Shyam S. Chaurasia, Christopher K. Hwang, and P. Michael Iuvone. 2011. “Dopamine D4 Receptor Activation Controls Circadian Timing of the Adenylyl Cyclase 1/cyclic AMP Signaling System in Mouse Retina.” *European Journal of Neuroscience*. https://doi.org/10.1111/j.1460-9568.2011.07734.x.

Jain, Miten, Hugh E. Olsen, Benedict Paten, and Mark Akeson. 2016. “Erratum to: The Oxford Nanopore MinION: Delivery of Nanopore Sequencing to the Genomics Community.” *Genome Biology* 17 (1): 256.

Jandial, Rahul, Reid Hoshide, J. Dawn Waters, and Charles L. Limoli. 2018. “Space-Brain: The Negative Effects of Space Exposure on the Central Nervous System.” *Surgical Neurology International* 9 (January): 9.

Jassal, Bijay, Lisa Matthews, Guilherme Viteri, Chuqiao Gong, Pascual Lorente, Antonio Fabregat, Konstantinos Sidiropoulos, et al. 2020. “The Reactome Pathway Knowledgebase.” *Nucleic Acids Research* 48 (D1): D498–503.

Jiang, L., F. Schlesinger, C. A. Davis, Y. Zhang, R. Li, M. Salit, T. R. Gingeras, and B. Oliver. 2011. “Synthetic Spike-in Standards for RNA-Seq Experiments.” *Genome Research*. https://doi.org/10.1101/gr.121095.111.

Kalmar, Bernadett, and Linda Greensmith. 2009. “Induction of Heat Shock Proteins for Protection against Oxidative Stress.” *Advanced Drug Delivery Reviews* 61 (4): 310–18.

Khan, Aziz, and Anthony Mathelier. 2017. “Intervene: A Tool for Intersection and Visualization of Multiple Gene or Genomic Region Sets.” *BMC Bioinformatics* 18 (1): 287.

Kleindienst, Andrea, Melissa J. McGinn, Harlan B. Harvey, Raymond J. Colello, Robert J. Hamm, and M. Ross Bullock. 2005. “Enhanced Hippocampal Neurogenesis by Intraventricular S100B Infusion Is Associated with Improved Cognitive Recovery after Traumatic Brain Injury.” *Journal of Neurotrauma*. https://doi.org/10.1089/neu.2005.22.645.

Koka, Saisudha, Anindita Das, Fadi N. Salloum, and Rakesh C. Kukreja. 2013. “Phosphodiesterase-5 Inhibitor Tadalafil Attenuates Oxidative Stress and Protects against Myocardial Ischemia/reperfusion Injury in Type 2 Diabetic Mice.” *Free Radical Biology & Medicine* 60 (July): 80–88.

Kotas, Maya E., and Ruslan Medzhitov. 2015. “Homeostasis, Inflammation, and Disease

Susceptibility." *Cell* 160 (5): 816–27.
Kovalchuk, A., and B. Kolb. 2017. "Low Dose Radiation Effects on the Brain–from Mechanisms and Behavioral Outcomes to Mitigation Strategies." *Cell Cycle* . https://www.tandfonline.com/doi/abs/10.1080/15384101.2017.1320003.
Kovalchuk, Anna, and Bryan Kolb. 2017a. "Low Dose Radiation Effects on the Brain--from Mechanisms and Behavioral Outcomes to Mitigation Strategies." *Cell Cycle* 16 (13): 1266–70.
———. 2017b. "Low Dose Radiation Effects on the Brain – from Mechanisms and Behavioral Outcomes to Mitigation Strategies." *Cell Cycle*. https://doi.org/10.1080/15384101.2017.1320003.
Krueger, Felix. 2019. *Trim Galore: A Wrapper around Cutadapt and FastQC to Consistently Apply Adapter and Quality Trimming to FastQ Files, with Extra Functionality for RRBS Data* (version Version 0.6.5). https://github.com/FelixKrueger/TrimGalore.
Krueger, Felix, and Simon R. Andrews. 2011. "Bismark: A Flexible Aligner and Methylation Caller for Bisulfite-Seq Applications." *Bioinformatics* 27 (11): 1571–72.
Labbé, Roselyne M., Andreana Holowatyj, and Zeng-Quan Yang. 2013. "Histone Lysine Demethylase (KDM) Subfamily 4: Structures, Functions and Therapeutic Potential." *American Journal of Translational Research* 6 (1): 1–15.
Lachner, Monika, Roderick J. O'Sullivan, and Thomas Jenuwein. 2003. "An Epigenetic Road Map for Histone Lysine Methylation." *Journal of Cell Science* 116 (Pt 11): 2117–24.
Langmead, Ben, Cole Trapnell, Mihai Pop, and Steven L. Salzberg. 2009. "Ultrafast and Memory-Efficient Alignment of Short DNA Sequences to the Human Genome." *Genome Biology* 10 (3): R25.
LaVail, Matthew M., Robert Eugene Anderson, and Joe G. Hollyfield. 1989. *Inherited and Environmentally Induced Retinal Degenerations: Proceedings of the International Symposium on Retinal Degenerations, Held in San Francisco, California, September 2 and 3, 1988*. Liss.
Law, Charity W., Monther Alhamdoosh, Shian Su, Xueyi Dong, Luyi Tian, Gordon K. Smyth, and Matthew E. Ritchie. 2016. "RNA-Seq Analysis Is Easy as 1-2-3 with Limma, Glimma and edgeR." *F1000Research* 5 (June). https://doi.org/10.12688/f1000research.9005.3.
Lawley, Justin S., Lonnie G. Petersen, Erin J. Howden, Satyam Sarma, William K. Cornwell, Rong Zhang, Louis A. Whitworth, Michael A. Williams, and Benjamin D. Levine. 2017. "Effect of Gravity and Microgravity on Intracranial Pressure." *The Journal of Physiology* 595 (6): 2115–27.
Lee, Andrew G., Thomas H. Mader, C. Robert Gibson, Tyson J. Brunstetter, and William J. Tarver. 2018. "Space Flight-Associated Neuro-Ocular Syndrome (SANS)." *Eye*. https://doi.org/10.1038/s41433-018-0070-y.
Lee, Jessica K., Vincent Koppelmans, Roy F. Riascos, Khader M. Hasan, Ofer Pasternak, Ajitkumar P. Mulavara, Jacob J. Bloomberg, and Rachael D. Seidler. 2019. "Spaceflight-Associated Brain White Matter Microstructural Changes and Intracranial Fluid Redistribution." *JAMA Neurology* 76 (4): 412–19.
Léveillard, Thierry, Saddek Mohand-Saïd, Olivier Lorentz, David Hicks, Anne-Claire Fintz, Emmanuelle Clérin, Manuel Simonutti, et al. 2004. "Identification and Characterization of Rod-Derived Cone Viability Factor." *Nature Genetics* 36 (7): 755–59.
Liao, Yuxing, Jing Wang, Eric J. Jaehnig, Zhiao Shi, and Bing Zhang. 2019. "WebGestalt 2019:

Gene Set Analysis Toolkit with Revamped UIs and APIs." *Nucleic Acids Research* 47 (W1): W199–205.
Li, Bo, and Colin N. Dewey. 2011. "RSEM: Accurate Transcript Quantification from RNA-Seq Data with or without a Reference Genome." *BMC Bioinformatics* 12 (August): 323.
Li, H., and R. Durbin. 2009. "Fast and Accurate Short Read Alignment with Burrows-Wheeler Transform." *Bioinformatics*. https://doi.org/10.1093/bioinformatics/btp324.
Li, Hong-Sheng, Yan-Ni Zhou, Lu Li, Sheng-Fu Li, Dan Long, Xue-Lu Chen, Jia-Bi Zhang, Li Feng, and You-Ping Li. 2019. "HIF-1α Protects against Oxidative Stress by Directly Targeting Mitochondria." *Redox Biology*, January, 101109.
Li, Jian, Chun-Na Liu, Ning Wei, Xi-Dong Li, Yuan-Yuan Liu, Rui Yang, and Yu-Jie Jia. 2016. "Protective Effects of BAY 73-6691, a Selective Inhibitor of Phosphodiesterase 9, on Amyloid-β Peptides-Induced Oxidative Stress in in-Vivo and in-Vitro Models of Alzheimer's Disease." *Brain Research*. https://doi.org/10.1016/j.brainres.2016.04.011.
Li, Lihua, Hua Zhang, Michel Varrin-Doyer, Scott S. Zamvil, and A. S. Verkman. 2011. "Proinflammatory Role of aquaporin‐4 in Autoimmune Neuroinflammation." *FASEB Journal: Official Publication of the Federation of American Societies for Experimental Biology* 25 (5): 1556–66.
Love, Michael I., Wolfgang Huber, and Simon Anders. 2014. "Moderated Estimation of Fold Change and Dispersion for RNA-Seq Data with DESeq2." *Genome Biology* 15 (12): 550.
Lowe, Xiu R., Sanchita Bhattacharya, Francesco Marchetti, and Andrew J. Wyrobek. 2009. "Early Brain Response to Low-Dose Radiation Exposure Involves Molecular Networks and Pathways Associated with Cognitive Functions, Advanced Aging and Alzheimer's Disease." *Radiation Research* 171 (1): 53–65.
MacLaren, R. E., R. A. Pearson, A. MacNeil, R. H. Douglas, T. E. Salt, M. Akimoto, A. Swaroop, J. C. Sowden, and R. R. Ali. 2006. "Retinal Repair by Transplantation of Photoreceptor Precursors." *Nature* 444 (7116): 203–7.
Mader, Thomas H., C. Robert Gibson, Anastas F. Pass, Larry A. Kramer, Andrew G. Lee, Jennifer Fogarty, William J. Tarver, et al. 2011. "Optic Disc Edema, Globe Flattening, Choroidal Folds, and Hyperopic Shifts Observed in Astronauts after Long-Duration Space Flight." *Ophthalmology* 118 (10): 2058–69.
Mao, Xiao Wen, Nina C. Nishiyama, Mary Campbell-Beachler, Peter Gifford, Kristine E. Haynes, Daila S. Gridley, and Michael J. Pecaut. 2017. "Role of NADPH Oxidase as a Mediator of Oxidative Damage in Low-Dose Irradiated and Hindlimb-Unloaded Mice." *Radiation Research* 188 (4): 392–99.
Mao, Xiao Wen, Nina C. Nishiyama, Michael J. Pecaut, Mary Campbell-Beachler, Peter Gifford, Kristine E. Haynes, Caroline Becronis, and Daila S. Gridley. 2016. "Simulated Microgravity and Low-Dose/Low-Dose-Rate Radiation Induces Oxidative Damage in the Mouse Brain." *Radiation Research* 185 (6): 647–57.
Mao, Xiao W., Michael J. Pecaut, Louis S. Stodieck, Virginia L. Ferguson, Ted A. Bateman, Mary Bouxsein, Tamako A. Jones, et al. 2013. "Spaceflight Environment Induces Mitochondrial Oxidative Damage in Ocular Tissue." *Radiation Research* 180 (4): 340–50.
Mao, X. W., M. Boerma, D. Rodriguez, M. Campbell-Beachler, T. Jones, S. Stanbouly, V. Sridharan, N. C. Nishiyama, A. Wroe, and G. A. Nelson. 2018. "Combined Effects of Low-Dose Proton Radiation and Simulated Microgravity on the Mouse Retina and the Hematopoietic System." *Radiation Research*, November.

https://doi.org/10.1667/RR15219.1.
Mao, X. W., M. Boerma, D. Rodriguez, M. Campbell-Beachler, T. Jones, S. Stanbouly, V. Sridharan, A. Wroe, and G. A. Nelson. 2018. "Acute Effect of Low-Dose Space Radiation on Mouse Retina and Retinal Endothelial Cells." *Radiation Research* 190 (1): 45–52.
Mao, X. W., J. D. Crapo, T. Mekonnen, N. Lindsey, P. Martinez, D. S. Gridley, and J. M. Slater. 2009. "Radioprotective Effect of a Metalloporphyrin Compound in Rat Eye Model." *Current Eye Research* 34 (1): 62–72.
Martin, Marcel. 2011. "Cutadapt Removes Adapter Sequences from High-Throughput Sequencing Reads." *EMBnet.journal* 17 (1): 10–12.
Masood, Anbrin, Ahmed Nadeem, S. Jamal Mustafa, and James M. O'Donnell. 2008. "Reversal of Oxidative Stress-Induced Anxiety by Inhibition of Phosphodiesterase-2 in Mice." *The Journal of Pharmacology and Experimental Therapeutics* 326 (2): 369–79.
Mattar, Pierre, Milanka Stevanovic, Ivana Nad, and Michel Cayouette. 2018. "Casz1 Controls Higher-Order Nuclear Organization in Rod Photoreceptors." *Proceedings of the National Academy of Sciences of the United States of America* 115 (34): E7987–96.
McIntyre, Lauren M., Kenneth K. Lopiano, Alison M. Morse, Victor Amin, Ann L. Oberg, Linda J. Young, and Sergey V. Nuzhdin. 2011. "RNA-Seq: Technical Variability and Sampling." *BMC Genomics* 12 (June): 293.
Medzhitov, Ruslan. 2008. "Origin and Physiological Roles of Inflammation." *Nature* 454 (7203): 428–35.
Meli, Rosaria, Claudio Pirozzi, and Alessandra Pelagalli. 2018. "New Perspectives on the Potential Role of Aquaporins (AQPs) in the Physiology of Inflammation." *Frontiers in Physiology* 9 (February): 101.
Mi, Huaiyu, Anushya Muruganujan, Dustin Ebert, Xiaosong Huang, and Paul D. Thomas. 2019. "PANTHER Version 14: More Genomes, a New PANTHER GO-Slim and Improvements in Enrichment Analysis Tools." *Nucleic Acids Research* 47 (D1): D419–26.
Mi, Huaiyu, Anushya Muruganujan, and Paul D. Thomas. 2013. "PANTHER in 2013: Modeling the Evolution of Gene Function, and Other Gene Attributes, in the Context of Phylogenetic Trees." *Nucleic Acids Research* 41 (Database issue): D377–86.
Mizumatsu, Shinichiro, Michelle L. Monje, Duncan R. Morhardt, Radoslaw Rola, Theo D. Palmer, and John R. Fike. 2003. "Extreme Sensitivity of Adult Neurogenesis to Low Doses of X-Irradiation." *Cancer Research* 63 (14): 4021–27.
Morey-Holton, Emily, Ruth K. Globus, Alexander Kaplansky, and Galina Durnova. 2005. "The Hindlimb Unloading Rat Model: Literature Overview, Technique Update and Comparison with Space Flight Data." *Experimentation With Animal Models In Space*. https://doi.org/10.1016/s1569-2574(05)10002-1.
Movafagh, Shahrzad, Sean Crook, and Kim Vo. 2015. "Regulation of Hypoxia-Inducible Factor-1a by Reactive Oxygen Species: New Developments in an Old Debate." *Journal of Cellular Biochemistry* 116 (5): 696–703.
Munro, Sarah A., Steven P. Lund, P. Scott Pine, Hans Binder, Djork-Arné Clevert, Ana Conesa, Joaquin Dopazo, et al. 2014. "Assessing Technical Performance in Differential Gene Expression Experiments with External Spike-in RNA Control Ratio Mixtures." *Nature Communications* 5 (September): 5125.
Najjar, Souhel, Daniel M. Pearlman, Orrin Devinsky, Amanda Najjar, and David Zagzag. 2013. "Neurovascular Unit Dysfunction with Blood-Brain Barrier Hyperpermeability Contributes

to Major Depressive Disorder: A Review of Clinical and Experimental Evidence." *Journal of Neuroinflammation* 10 (December): 142.
National Research Council, Division on Earth and Life Studies, Institute for Laboratory Animal Research, and Committee for the Update of the Guide for the Care and Use of Laboratory Animals. 2010. *Guide for the Care and Use of Laboratory Animals: Eighth Edition*. National Academies Press.
Nelson, Emily S., Lealem Mulugeta, Andrew Feola, Julia Raykin, Jerry G. Myers, Brian C. Samuels, and C. Ross Ethier. 2017. "The Impact of Ocular Hemodynamics and Intracranial Pressure on Intraocular Pressure during Acute Gravitational Changes." *Journal of Applied Physiology* 123 (2): 352–63.
Nelson, Emily S., Lealem Mulugeta, and Jerry G. Myers. 2014. "Microgravity-Induced Fluid Shift and Ophthalmic Changes." *Life* 4 (4): 621–65.
Nguyen, Tuan-Minh, Adib Shafi, Tin Nguyen, and Sorin Draghici. 2019. "Identifying Significantly Impacted Pathways: A Comprehensive Review and Assessment." *Genome Biology* 20 (1): 203.
Overbey, Eliah G., Amber M. Paul, Willian A. da Silveira, Candice G. T. Tahimic, Sigrid S. Reinsch, Nathaniel Szewczyk, Seta Stanbouly, Charles Wang, Jonathan M. Galazka, and Xiao Wen Mao. 2019. "Mice Exposed to Combined Chronic Low-Dose Irradiation and Modeled Microgravity Develop Long-Term Neurological Sequelae." *International Journal of Molecular Sciences* 20 (17). https://doi.org/10.3390/ijms20174094.
Patro, Rob, Geet Duggal, Michael I. Love, Rafael A. Irizarry, and Carl Kingsford. 2017. "Salmon Provides Fast and Bias-Aware Quantification of Transcript Expression." *Nature Methods* 14 (4): 417–19.
Paul, Amber M., Dhiraj Acharya, Laurel Duty, E. Ashley Thompson, Linda Le, Dobrivoje S. Stokic, A. Arturo Leis, and Fengwei Bai. 2017. "Osteopontin Facilitates West Nile Virus Neuroinvasion via Neutrophil 'Trojan Horse' Transport." *Scientific Reports* 7 (1): 4722.
Pérez de Sevilla Müller, Luis, Shaghauyegh S. Azar, Janira de Los Santos, and Nicholas C. Brecha. 2017. "Prox1 Is a Marker for AII Amacrine Cells in the Mouse Retina." *Frontiers in Neuroanatomy* 11 (May): 39.
Piñero, Janet, Àlex Bravo, Núria Queralt-Rosinach, Alba Gutiérrez-Sacristán, Jordi Deu-Pons, Emilio Centeno, Javier García-García, Ferran Sanz, and Laura I. Furlong. 2017. "DisGeNET: A Comprehensive Platform Integrating Information on Human Disease-Associated Genes and Variants." *Nucleic Acids Research* 45 (D1): D833–39.
Raplee, Isaac D., Alexei V. Evsikov, and Caralina Marín de Evsikova. 2019. "Aligning the Aligners: Comparison of RNA Sequencing Data Alignment and Gene Expression Quantification Tools for Clinical Breast Cancer Research." *Journal of Personalized Medicine* 9 (2). https://doi.org/10.3390/jpm9020018.
Rebhan, M. 1997. "GeneCards: Integrating Information about Genes, Proteins and Diseases." *Trends in Genetics*. https://doi.org/10.1016/s0168-9525(97)01103-7.
Risso, Davide, John Ngai, Terence P. Speed, and Sandrine Dudoit. 2014. "Normalization of RNA-Seq Data Using Factor Analysis of Control Genes or Samples." *Nature Biotechnology* 32 (9): 896–902.
Risso, Davide, Katja Schwartz, Gavin Sherlock, and Sandrine Dudoit. 2011. "GC-Content Normalization for RNA-Seq Data." *BMC Bioinformatics* 12 (December): 480.
Roberts, Joan E., Barbara M. Kukielczak, Colin F. Chignell, Bob H. Sik, Dan-Ning Hu, and

Mary Ann Principato. 2006. “Simulated Microgravity Induced Damage in Human Retinal Pigment Epithelial Cells.” *Molecular Vision* 12 (May): 633–38.
Roberts, Richard J., Mauricio O. Carneiro, and Michael C. Schatz. 2013. “The Advantages of SMRT Sequencing.” *Genome Biology* 14 (7): 405.
Rocca-Serra, P., M. Brandizi, E. Maguire, N. Sklyar, C. Taylor, K. Begley, D. Field, et al. 2010. “ISA Software Suite: Supporting Standards-Compliant Experimental Annotation and Enabling Curation at the Community Level.” *Bioinformatics*. https://doi.org/10.1093/bioinformatics/btq415.
Romero-Puertas, María C., and Luisa M. Sandalio. 2016. “Nitric Oxide Level Is Self-Regulating and Also Regulates Its ROS Partners.” *Frontiers in Plant Science* 7 (March): 316.
Ronca, April E., Eric L. Moyer, Yuli Talyansky, Moniece Lowe, Shreejit Padmanabhan, Sungshin Choi, Cynthia Gong, Samuel M. Cadena, Louis Stodieck, and Ruth K. Globus. 2019. “Behavior of Mice Aboard the International Space Station.” *Scientific Reports* 9 (1): 4717.
Rutter, Lindsay, Richard Barker, Daniela Bezdan, Henry Cope, Sylvain V. Costes, Lovorka Degoricija, Kathleen M. Fisch, et al. n.d. “A New Era for Space Life Science: International Standards for Space Omics Processing (ISSOP).” *Cell*. Accessed August 26, 2020. https://drive.google.com/drive/u/0/folders/1BQ72FMlHZ_GR777KaZ6nCL3uTe4Q07x4.
Sansone, Susanna-Assunta, Philippe Rocca-Serra, Dawn Field, Eamonn Maguire, Chris Taylor, Oliver Hofmann, Hong Fang, et al. 2012. “Toward Interoperable Bioscience Data.” *Nature Genetics* 44 (2): 121–26.
Schaarschmidt, Stephanie, Axel Fischer, Ellen Zuther, and Dirk K. Hincha. 2020. “Evaluation of Seven Different RNA-Seq Alignment Tools Based on Experimental Data from the Model Plant Arabidopsis Thaliana.” *International Journal of Molecular Sciences* 21 (5). https://doi.org/10.3390/ijms21051720.
Schurch, Nicholas J., Pietá Schofield, Marek Gierliński, Christian Cole, Alexander Sherstnev, Vijender Singh, Nicola Wrobel, et al. 2016. “How Many Biological Replicates Are Needed in an RNA-Seq Experiment and Which Differential Expression Tool Should You Use?” *RNA*. https://doi.org/10.1261/rna.053959.115.
Scott, Ryan T., Kirill Grigorev, Graham Mackintosh, Samrawit G. Gebre, Christopher E. Mason, Martha E. Del Alto, and Sylvain V. Costes. 2020. “Advancing the Integration of Biosciences Data Sharing to Further Enable Space Exploration.” *Cell Reports*.
Seawright, John W., Yusra Samman, Vijayalakshmi Sridharan, Xiao Wen Mao, Maohua Cao, Preeti Singh, Stepan Melnyk, et al. 2017. “Effects of Low-Dose Rate γ-Irradiation Combined with Simulated Microgravity on Markers of Oxidative Stress, DNA Methylation Potential, and Remodeling in the Mouse Heart.” *PloS One* 12 (7): e0180594.
Smart, Deedee. 2017. “Radiation Toxicity in the Central Nervous System: Mechanisms and Strategies for Injury Reduction.” *Seminars in Radiation Oncology* 27 (4): 332–39.
Smith, Cynthia L., Judith A. Blake, James A. Kadin, Joel E. Richardson, Carol J. Bult, and the Mouse Genome Database Group. 2018. “Mouse Genome Database (MGD)-2018: Knowledgebase for the Laboratory Mouse.” *Nucleic Acids Research*. https://doi.org/10.1093/nar/gkx1006.
Soneson, C., M. Love, and M. Robinson. n.d. “Differential Analyses for RNA-Seq: Transcript-Level Estimates Improve Gene-Level Inferences [version 2; Peer Review: 2 Approved].” F1000Research.

Song, Ting-Ting, Yan-Hua Bi, Yu-Qi Gao, Rui Huang, Ke Hao, Gang Xu, Jia-Wei Tang, et al. 2016. “Systemic pro-Inflammatory Response Facilitates the Development of Cerebral Edema during Short Hypoxia.” *Journal of Neuroinflammation* 13 (1): 63.

Stenger, Michael B., William J. Tarver, Tyson Brunstetter, Charles Robert Gibson, Steven S. Laurie, Stuart M. C. Lee, Brandon R. Macias, et al. 2017. “Evidence Report: Risk of Spaceflight Associated Neuro-Ocular Syndrome (SANS),” November. https://ntrs.nasa.gov/search.jsp?R=20180000936.

Subramanian, Aravind, Pablo Tamayo, Vamsi K. Mootha, Sayan Mukherjee, Benjamin L. Ebert, Michael A. Gillette, Amanda Paulovich, et al. 2005. “Gene Set Enrichment Analysis: A Knowledge-Based Approach for Interpreting Genome-Wide Expression Profiles.” *Proceedings of the National Academy of Sciences of the United States of America* 102 (43): 15545–50.

Sun, Dong, M. Ross Bullock, Melissa J. McGinn, Zhengwen Zhou, Nabil Altememi, Sarah Hagood, Robert Hamm, and Raymond J. Colello. 2009. “Basic Fibroblast Growth Factor-Enhanced Neurogenesis Contributes to Cognitive Recovery in Rats Following Traumatic Brain Injury.” *Experimental Neurology*. https://doi.org/10.1016/j.expneurol.2008.11.011.

Sun, X. Z., M. Inouye, Y. Fukui, S. Hisano, K. Sawada, H. Muramatsu, and T. Muramatsu. 1997. “An Immunohistochemical Study of Radial Glial Cells in the Mouse Brain Prenatally Exposed to Gamma-Irradiation.” *Journal of Neuropathology and Experimental Neurology* 56 (12): 1339–48.

Szklarczyk, D., A. L. Gable, D. Lyon, and A. Junge. 2019. “STRING v11: Protein–protein Association Networks with Increased Coverage, Supporting Functional Discovery in Genome-Wide Experimental Datasets.” *Nucleic Acids*. https://academic.oup.com/nar/article-abstract/47/D1/D607/5198476.

Teng, Mingxiang, Michael I. Love, Carrie A. Davis, Sarah Djebali, Alexander Dobin, Brenton R. Graveley, Sheng Li, et al. 2016. “A Benchmark for RNA-Seq Quantification Pipelines.” *Genome Biology* 17 (April): 74.

Terada, Koji, and Takahisa Furukawa. 2010. “Sumoylation Controls Retinal Progenitor Proliferation by Repressing Cell Cycle Exit in Xenopus Laevis.” *Developmental Biology* 347 (1): 180–94.

Thirsk, Robert, Andre Kuipers, Chiaki Mukai, and David Williams. 2009. “The Space-Flight Environment: The International Space Station and beyond.” *CMAJ: Canadian Medical Association Journal = Journal de l’Association Medicale Canadienne* 180 (12): 1216–20.

Thomas, P. D. 2003. “PANTHER: A Library of Protein Families and Subfamilies Indexed by Function.” *Genome Research*. https://doi.org/10.1101/gr.772403.

Tymko, Michael M., Lindsey M. Boulet, and Joseph Donnelly. 2017. “Intracranial Pressure in Outer Space: Preparing for the Mission to Mars.” *The Journal of Physiology*.

Ueda, Kaori, Tsutomu Nakahara, Maya Hoshino, Asami Mori, Kenji Sakamoto, and Kunio Ishii. 2010. “Retinal Blood Vessels Are Damaged in a Rat Model of NMDA-Induced Retinal Degeneration.” *Neuroscience Letters* 485 (1): 55–59.

Väremo, Leif, Jens Nielsen, and Intawat Nookaew. 2013. “Enriching the Gene Set Analysis of Genome-Wide Data by Incorporating Directionality of Gene Expression and Combining Statistical Hypotheses and Methods.” *Nucleic Acids Research* 41 (8): 4378–91.

Walker, Steven A., Lawrence W. Townsend, and John W. Norbury. 2013. “Heavy Ion

Contributions to Organ Dose Equivalent for the 1977 Galactic Cosmic Ray Spectrum." *Advances in Space Research*. https://doi.org/10.1016/j.asr.2012.12.011.

Wang, Jing, Suhas Vasaikar, Zhiao Shi, Michael Greer, and Bing Zhang. 2017. "WebGestalt 2017: A More Comprehensive, Powerful, Flexible and Interactive Gene Set Enrichment Analysis Toolkit." *Nucleic Acids Research* 45 (W1): W130–37.

Weiss, E. R., Y. Hao, C. D. Dickerson, and S. Osawa. 1995. "Altered cAMP Levels in Retinas from Transgenic Mice Expressing a Rhodopsin Mutant." *Biochemical and*. https://www.sciencedirect.com/science/article/pii/S0006291X85726861.

Williams, Claire R., Alyssa Baccarella, Jay Z. Parrish, and Charles C. Kim. 2016. "Trimming of Sequence Reads Alters RNA-Seq Gene Expression Estimates." *BMC Bioinformatics* 17 (February): 103.

Wu, Douglas C., Jun Yao, Kevin S. Ho, Alan M. Lambowitz, and Claus O. Wilke. 2018. "Limitations of Alignment-Free Tools in Total RNA-Seq Quantification." *BMC Genomics* 19 (1): 510.

Yamazaki, Yu, Mitsuru Shinohara, Motoko Shinohara, Akari Yamazaki, Melissa E. Murray, Amanda M. Liesinger, Michael G. Heckman, et al. 2019. "Selective Loss of Cortical Endothelial Tight Junction Proteins during Alzheimer's Disease Progression." *Brain*. https://doi.org/10.1093/brain/awz011.

Yang, Cheng, Po-Yen Wu, Li Tong, John H. Phan, and May D. Wang. 2015. "The Impact of RNA-Seq Aligners on Gene Expression Estimation." *ACM-BCB ... ... : The ... ACM Conference on Bioinformatics, Computational Biology and Biomedicine. ACM Conference on Bioinformatics, Computational Biology and Biomedicine* 2015 (September): 462–71.

Yasuda, Masayuki, Yuji Tanaka, Morin Ryu, Satoru Tsuda, and Toru Nakazawa. 2014. "RNA Sequence Reveals Mouse Retinal Transcriptome Changes Early after Axonal Injury." *PloS One* 9 (3): e93258.

Zhang, Juan, and Jianwei Jiao. 2015. "Molecular Biomarkers for Embryonic and Adult Neural Stem Cell and Neurogenesis." *BioMed Research International* 2015 (September): 727542.

Zhao, Chunmei, Wei Deng, and Fred H. Gage. 2008. "Mechanisms and Functional Implications of Adult Neurogenesis." *Cell* 132 (4): 645–60.

Zhao, Ying, Kai Feng, Ruibao Liu, Jinhua Pan, Lailin Zhang, and Xuejing Lu. 2019. "Vitamins and Mineral Supplements for Retinitis Pigmentosa." *Journal of Ophthalmology* 2019 (February): 8524607.

Zheng, Ling, Bendi Gong, Denise A. Hatala, and Timothy S. Kern. 2007. "Retinal Ischemia and Reperfusion Causes Capillary Degeneration: Similarities to Diabetes." *Investigative Ophthalmology & Visual Science* 48 (1): 361–67.

Zhuo, Yeye, Haibiao Guo, Yufang Cheng, Chuang Wang, Canmao Wang, Jingang Wu, Zhengqiang Zou, Danna Gan, Yiwen Li, and Jiangping Xu. 2016. "Inhibition of Phosphodiesterase-4 Reverses the Cognitive Dysfunction and Oxidative Stress Induced by Aβ25–35 in Rats." *Metabolic Brain Disease* 31 (4): 779–91.

Chapter 5: Epitranscriptomic Landscape of Naive and Primed Pluripotent Stem Cells

5.1 Abstract

Compared to DNA and protein modifications, the role and function of RNA modifications is understudied and poorly understood. Methods of profiling multiple RNA modifications in parallel is necessary to scale studies of RNA modifications. To address this, we have used direct-RNA sequencing with the Oxford Nanopore MinION to create universal modification maps of three distinct states of pluripotency (5i/L/A, 2i/L/I/F, mTeSR) in human embryonic stem cells. Over 1900 genes were modified in each state and these modified genes were enriched for known functions of individual modifications, allowing us to create a universal modification map. Additionally, enriched modified 5-mer sequences were identified across all three states and showed variation between states. Many of these 5-mers were associated with RNA binding proteins, particularly EIF4B which is known to associate with the motif for the m^6A modification. Of special interest is the landscape of the m^6A modification, the role of which has had conflicting results in pluripotent stem cells. We found unique modified genes between the different pluripotent states, which offers an explanation to previous conflicting results.

5.2 Introduction

Over 170 RNA modifications have been detected (Boccaletto et al., 2018), yet the epitranscriptomic landscape and its role in human health and disease remains largely unknown. To date, RNA modifications have connections to cancer (Lin et al. 2016; Jun Liu et al., 2018; Ma et al., 2017), infertility (Zheng et al., 2013; Ivanova et al., 2017; Hsu et al., 2017), obesity (Church et al., 2010; Jia et al., 2011), among many other diseases (Jonkhout et al., 2017) In order to further our understanding of the epitranscriptome landscape and its impact on disease, the molecular effects of RNA modifications need to be further characterized.

The most common epitranscriptomic modifications identified in humans are m^6A, m^1A, m^5C, hm^5C, 8-oxoG, and pseudouridine (Ψ) (Roundtree et al., 2017). These modifications are reversible and have specific proteins related to dynamic writing and erasing of RNA modifications (F. Liu et al., 2016; Jia et al., 2011; Jianzhao Liu et al., 2014). Moreover, these modifications are linked to changes in translation efficiency (Dominissini et al., 2016; X. Li et al., 2017; Simms et al., 2014), RNA degradation (Karijolich and Yu 2011; Lee et al., 2020),

stability (X. Zhang et al., 2012), structure (Tanzer, Hofacker, and Lorenz 2019), and splicing (Zhou et al., 2019).

Of particular interest is the influence of m^6A on cell fate determination and pluripotency. Previous studies in stem cells have linked m^6A to differentiation potential, yet have presented contradictory results. Two studies have concluded that m^6A is necessary for transcriptome turnover and elimination of this modification will inhibit differentiation out of the pluripotent state (Batista et al., 2014; Geula et al., 2015). In another set of studies it was found that inhibiting the deposition of m^6A led to decreased self-renewal and triggered differentiation (Wen et al., 2018; Wu et al., 2019; Aguilo et al., 2015). One hypothesis that has been put forward is that differences in the state of pluripotency may be responsible for the contradictions in differentiation potential (M. Zhang et al., 2020).

Pluripotency can be defined by two distinct states: naive and primed (Nichols and Smith 2009). The naive state is an earlier developmental state and corresponds to the pre-implantation embryo *in vivo.* In contrast, the primed state occurs when stem cells are ready for lineage-specific differentiation and correspond to the post-implantation embryo. Pluripotent stem cells exist in these states along a pluripotency spectrum depending on their growth conditions (Sperber et al., 2015). One of the earliest growth conditions is the 5i/L/A state, which contains a mixture of five kinase inhibitors + leukemia inhibitory factor + Activin A (Theunissen et al., 2014). On the other end of the spectrum is the mTeSR growth condition, which is the end stage of primed pluripotency (Ludwig and Thomson 2007). Located between these two states is the 2i/L/I/F growth condition, which contains two inhibitors + leukemia inhibitory factor + insulin-like growth factor + fibroblast growth factor and represent a state of late naive pluripotency (Ware et al., 2014). It has been well documented that different pluripotent states have unique transcriptional and epigenetic profiles (Battle et al., 2019; Sperber et al., 2015).

Previous studies characterizing RNA modifications in stem cells used immunoprecipitation assays. Batista *et al.,* used m^6A RNA immunoprecipitation sequencing (RIP-seq), which fragments RNA fragments, uses an antibody to select for fragments with the m^6A modification, and peak calling to find modified regions (Batista et al., 2014). In contrast, Molinie *et al*., have used m6A-level and isoform-characterization sequencing (m^6A-LAIC), which is similar to m^6A RIP-seq, but skips the fragmentation step, allowing for the detection of ratios of modified-to-unmodified transcripts of a single gene (Molinie et al., 2016). The drawback of these methods is that the technical skill required, along with the cost of reagents to perform the immunoprecipitation step.

Nanopore sequencing has been put forth as an alternative method for detecting RNA modifications. In nanopore sequencing, poly(A) RNA is passed through a small pore

simultaneous to an electrical current (Jain et al., 2016). The fluctuations in electrical current can be analyzed using deep learning in order to basecall the poly(A) RNA molecule (Garalde et al., 2018). Recently, it has been demonstrated that these current fluctuations are sensitive enough to detect base modifications. The software tool 'epitranscriptional landscape inferring from glitches of ONT signals' (ELIGOS) is capable of predicting the location of RNA modifications using basecalling errors, where the sequenced base differs from the reference genome (Jenjaroenpun et al., 2021). The accuracy of ELIGOS is variable based on the modification being detected. The m^6A modification in the context of the DRACH motif (AGT/AG/A/C/ACT) demonstrated a ROC (receiver operating characteristic) value of 0.973, achieving high prediction performance.

Alternate calculation methods using basecalling error rates have been published for viral RNA (Price et al., 2020; Smith et al., 2019), *Arabidopsis thaliana* (Parker et al., 2020), *Saccharomyces cerevisiae (Garalde et al. 2018)*, and a human B-lymphocyte cell line (Workman et al., 2019). We present below the first study to use direct-RNA nanopore sequencing to derive novel findings about the epitranscriptome of stem cells. Here, we demonstrate that the results from previous, distinct studies of different RNA modifications are present within our single epitranscriptomic landscape, that our modification sites are enriched for RNA binding protein motifs, and we have created a catalog of changes in the m^6A landscape through different phases of pluripotency.

5.3 Results

5.3.1 Developmental States and their Epitranscriptomic Modification Landscape

Human embryonic stem cells (hESCs) from two cell lines, Elf1 and WIBR3, were grown in three distinct media conditions, 5i/L/A, 2i/L/I/F, and mTeSR (Figure 1A). Transcript per million (TPM) values for gene expression were correlated between cell lines grown in the same media condition, having a Pearson correlation of 0.93 or higher (Figure 1B). Principal component analysis of the top 500 most variable genes clusters samples by media condition (Figure 1C). Similarly, a heatmap quantifying samples by Euclidean distance also clusters samples by media condition, with 5i/L/A samples having the largest distance from the other media conditions.

While biological replicates are highly correlated with one another, they are not as high as correlations we expect from sequencers with high base-calling accuracy. This is due, in part, to the effect that RNA modifications cause on the recorded current signal during sequencing as current ONT basecaller does not have the capacity to detect RNA modifications. The tool ELIGOS was used to detect current aberrations that are caused by RNA modifications, which

uses the error rate during alignment to make these predictions. An unmodified stretch of RNA will follow the typical standard model closely, only occasionally producing noisy aberrations (Fig 2A). Compare this to modified 5-mer predictions, which have observably noisier signal patterns (Fig 2B).

In particular, we notice modified 5-mers on transcribed genes responsible for pluripotency maintenance (Fig 2B). There are (216, 144, 263) predicted modifications on *DNMT3B*, (145, 18, 127) on *NANOG*, and (81, 0, 265) on *SOX2* in the (5i/L/A, 2i/L/I/F, mTeSR) states respectively. In total, modifications were detected on (1930, 2168, 2727) genes in the (5i/L/A, 2i/L/I/F, mTeSR) states respectively. We then used the genes from PluriNet (Müller et al., 2008), a set of 299 genes linked to pluripotency, to broaden our search for modifications on stem cell specific genes. We find that modifications are present on nearly half (49.5%) of the genes during at least one of our pluripotency stages (Fig 2C). In Fig 2C, the log2-transformed modification count for each gene in the PluriNet gene set is displayed (Fig 2C rows), along with the average number of modifications across all expressed genes, including genes outside the PluriNet set. PluriNet genes with no detectable expression in our data are shaded gray. 104 (34.8%) of PluriNet genes have gene expression across all three states, yet have no predicted modifications. This is lower than the proportion of all genes, including genes outside the PluriNet set, that are expressed in all three states and have no modifications in any state, 48.0%. From comparing the distribution of modification counts between PluriNet and non-PluriNet genes, we observe that pluripotency related genes show greater enrichment for RNA modifications than the set of all expressed genes (Fig 2D).

Locationally, modifications are distributed near 5'UTR and 3'UTR boundaries with the coding sequence, with a smaller number of modifications located within the coding sequence (Fig 2E). To determine similarities in the modification landscape between developmental states, the number of modifications per gene in each state was summarized into a matrix where each row represents a gene and each column represents a biological replicate (Elf1 or WIBR3). The matrix was filtered for genes expressed in all three developmental states to eliminate modifications only present due to gene expression differences. The rows with the top 200 variance were selected and PCA was performed on the log2 transform of this matrix (Fig 2F). Samples form clusters that are linearly separable from one another, indicating a unique gene modification landscape for each state. The modification landscape of 5i/L/A state samples were more tightly clustered and distant from the 2i/L/I/F and mTeSR states, indicating a more unique modification landscape than that between the 2i/L/I/F and mTeSR states.

5.3.2 Biological and Functional Predictions of Highly Modified Transcripts

Of notice for each developmental state, there is a set of highly modified genes. Most genes had an average of 1-2 modifications per 100 bases, however, each state had a pool of genes with between 5-40 modifications per 100 bases (Fig 3A). The top 10% of highly modified genes were identified and compared between each developmental state (Fig 3B). Most highly modified genes were shared between all three developmental states. Enriched functions of all highly modified genes were determined using overrepresentation analysis in WebGestalt (Liao et al., 2019) to find enriched gene ontology (GO) terms and Reactome pathways. Among enriched GO biological process categories are "cell cycle" (FDR=$2.0x10^{-4}$), "regulation of cell cycle G1/S phase transition" (FDR=$5.2x10^{-3}$), "positive regulation of nucleic acid-templated transcription" (FDR=$4.2x10^{-4}$), "negative regulation of nucleic acid-templated transcription" (FDR=$2.5x10^{-2}$), and "positive regulation of developmental process" (FDR=$3.9x10^{-3}$). The GO molecular function categories are composed almost entirely of various binding functions, including "RNA binding" (FDR=$2.5x10^{-12}$), "transcription factor binding" (FDR=$1.0x10^{-4}$), and "double stranded DNA binding" (FDR=$1.6x10^{-4}$). The Reactome pathways that are enriched include "G2/M Transition" (FDR=$3.3x10^{-2}$), "PCP/CE Pathway" (FDR=$8.5x10^{-3}$), "Translation" (FDR=$1.7x10^{-3}$), and "Metabolism of RNA" (FDR=$1.3x10^{-4}$) (Fig 3C).

There were also small sets of genes unique to each developmental state. Within these sets, PluriNet genes were identified. The 5iLA-unique set had one gene from the PluriNet set, *TRIM28,* which acts as an epigenetic corepressor during stem cell reprogramming (Klimczak et al., 2017). Similarly, the 2iLIF-unique set had a sole gene, *GADD45A*, which promotes DNA demethylation (Schüle et al., 2019). The mTeSR-unique set had the seven genes *CDT1*, *HMGA1, MYC, NANOG, PMAIP1, PNN,* and *RASL11B* and the intersection of all three states had the three genes *COPS6, NOP56*, and *POP7*. The full sets of genes for each state were input into GeneWalk (Ietswaart et al., 2021) to determine significant functions of these genes by assigning them enriched GO terms. For 5i/LA state, *TRIM28* is related to binding functions, as well as the RAS signaling pathway (Fig 3D). *GADD45A* from the 2i/L/I/F-unique set performs various regulatory activities related to cell migration, mechanical stimulus, and cell cycle arrest (Fig 3E). From the set of mTeSR-unique genes, *NANOG* is known for its role as a core transcriptional regulator in pluripotency (Pan and Thomson 2007; Heurtier et al., 2019; Loh et al., 2006). Genewalk found that *NANOG* is involved in various DNA binding functions in addition to its known function in stem cell and pluripotency maintenance (Fig 3F).

5.3.3 Common Motifs and Associated RNA Binding Proteins

RNA transcripts are known to interact with RNA binding proteins, some of which have been associated with specific sequence motifs. While some RNA modifications have been linked to specific protein interactions, the role of RNA modifications in protein interactions is largely

unknown. To investigate whether there were differences in motifs across developmental states, fold-changes and p-values were calculated for 5-mer frequencies identified by ELIGOS compared to the background frequency of 5-mers in from all sequenced reads. Fold-change values and are plotted in Fig 4A, and enriched 5-mers (p-values < 0.001) are indicated in color. Homopolymer sequences of 3-bases or longer were removed prior to calculations, reducing the set of unique 5-mers from 1024 to 864. A matrix of fold-change values for each developmental state was generated for PCA. The 5-mers with the top 200 variance scores were used for analysis (Fig 4B). Samples are not tightly clustered by developmental state, but they do trend in different directions, indicating unique shifts in modification enrichment among the different states.

Enriched 5-mers were compared for similarity across developmental states (Fig 4C). The enriched 5-mer counts are (279, 279, 272) for the (5i/L/A, 2iL/I/F, mTeSR) states respectively. 202 5-mers are shared across all three states, which covers the majority of 5-mers. However, each state did have smaller sets of unique 5-mers, which explain what drives the subtle differences observed in PCA. Enriched 5-mers were searched for RNA-binding motifs using the database the RNA-binding Proteins Database (RBPBDB) (Cook et al., 2011) and the motifs listed in the supplement of Van Nostrand *et al.* (Van Nostrand et al., 2020). 107 k-mers with associated RNA binding motifs were identified and their fold-change values for each state were summarized in a heatmap (Fig 4D), where grey boxes indicate the associated k-mer was not enriched in that particular developmental state. Of note, three 5-mers matching the DRACH motif, which is associated with the m^6A modification, were enriched across all three developmental states. The mTeSR state in particular had the most enrichment of the DRACH motif across all three 5-mers. Corresponding to the 107 enriched k-mers are 38 RNA binding proteins. The EIF4B protein was related to the three DRACH motifs. The protein HNRNPA0 is associated with the enriched 5-mers CTAGG and TTAGG, particularly in the 2i/L/I/F state. The RBM4 protein has two associated 5-mers that are enriched, CGCGA in the 2iL/I/F state and ACGCG in the 5i/L/A state. The EIF4B, HNRNPA0, and RBM4 proteins are the only RNA binding proteins associated with enriched 5-mers that have a fold-change score greater than 3. These 3 proteins all have expression detected in the RNA-seq data across all three states (Fig 4E). The *EIF4B* gene is highly expressed across all three states, with the highest expression in the mTeSR state. *HNRNPA0* is highly expressed across all three states, particularly in the 2i/L/I/F and mTeSR states. *RBM4* has moderate expression across all states. Overall, expression was detected for 26 of the 38 RNA binding proteins.

A variety of proteins related to writing, reading, and erasing proteins have been previously identified (Nie et al., 2020). Each cell state shows expression of at least one writer, reader, or eraser protein (Fig 4F). Proteins for writing, reading, and erasing the m^6A modification was detected in each cell state. The proteins with the highest gene expression are HNRNPA2B1 and HNRNPC, which are known m^6A reader proteins. A single writer protein for m^1A was

detected across all three states, TRMT61B, which the other known writer proteins had no expression (TRMT61B, SUN3). Expression of the m^{6}A reader proteins YTHDC1 and YTHDC2 also had no expression, though five other reader proteins associated with this modification were expressed.

5.3.4 DRACH Motifs and the m^{6}A Modification

The m^{6}A modification has been previously characterized in the mTeSR developmental state in the human embryonic stem cell line H1 using m^{6}A RIP-seq and m^{6}A-LAIC (Batista et al., 2014; Molinie et al., 2016). Collectively, the m^{6}A RIP-seq and m^{6}A-LAIC data cover 11,039 out of 55,765 genes (19.8%) in the hg19 GENCODE annotation of the human genome. This breaks down into 2814 genes detected by m^{6}A RIP-seq, 3045 genes detected in m^{6}A-LAIC and 5180 genes detected in both assays (Fig 5A). All m^{6}A RIP-seq peaks and m^{6}A-LAIC peaks with detection on transcripts 25% of transcripts were used for analysis. To compare our ability to detect previously identified m^{6}A modifications, we sequenced H1 cells grown in the mTeSR state using direct-RNA nanopore sequencing and ran ELIGOS to detect modified DRACH motifs. Our detected H1 expression clustered with mTeSR data from Elf1 and WIBR3 cells (Sup Fig S1). We detected a total of 795 DRACH modified genes with enriched representation of previously identified m^{6}A modified genes. 528 (66.4%) of these genes were previously identified by either m^{6}A RIP-seq or m^{6}A-LAIC. Specifically, 258 genes were uniquely identified by m^{6}A RIP-seq, 42 by m^{6}A-LAIC and 228 by both assays. The positions of these modifications are distributed around the end of the coding sequence and the beginning of the 3' UTR (Fig 5B). Overrepresentation analysis of GO biological process and molecular function categories was performed on the novel genes and in the genes previously identified m^{6}A RIP-seq + m^{6}A-LAIC to compare the functions of these distinct gene sets. Multiple similarities in gene processes and functions were identified, and included "RNA binding", "structural molecule activity", and "cellular protein localization" (Fig 5C).

Modified DRACH motifs were identified in the 5i/L/A, 2i/L/I/F, and mTeSR states for Elf1 and WIBR3 cell lines. Similar to the H1 data, these modifications were distributed around the end of the coding sequence and the beginning of the 3' UTR (Fig 5D). The different developmental states have (785, 871, 1012) genes with these motifs in the (5i/L/A, 2i/L/I/F, mTeSR) states respectively. To identify genes that are modified in one cell state but absent in the others, m^{6}A-modified genes were filtered for genes with detected expression in all three states. 549 of these genes have m^{6}A modifications across all three developmental states (Fig 5E). There are (82, 72, 177) state-specific genes that are m^{6}A-modified in the (5i/L/A, 2i/L/I/F, mTeSR) states respectively (Fig 5E) despite having expression across all three states (Fig 5F). GO overrepresentation analysis was performed on these distinct gene sets to identify enriched

biological processes. There were no GO sets with an FDR < 0.05 in the 5i/L/A and 2i/L/I/F states, however, there were GO sets with high enrichment ratios. The GO sets with the 10 lowest FDR values for each state are displayed in Fig 5G. The GO categories were primarily unique among the different cell states, with a few exceptions. The categories “amide biosynthetic process”, “peptide biosynthetic process”, and “translation” were shared between the 5i/L/A and mTeSR states. All states also had enriched functions related to various mitochondrial functions.

5.4 Discussion

Characterization and annotation of the epitranscriptome is still in its infancy. In this study, we have demonstrated that a direct-RNA nanopore sequencing is a potential method for accelerating our understanding of epitranscriptomics. In a single experiment, we were able to determine precise locations of candidate modification sites that are not constrained to any specific modification.

We observed that the locational distribution of RNA modifications is consistent with previous observations. Specifically, the m^6A modification is heavily distributed towards the 3’UTR with more moderate representation throughout the rest of the sequence (Mao et al., 2019). Additionally, there is a spike around the 5’UTR in the global modification distribution that is not as prominent in the DRACH distribution, which can be explained by the typical binding sites for m^1A and m^5C (X. Li et al., 2016; Squires et al., 2012). Modifications in the coding sequence that do not correspond to the DRACH motif could be explained by the hm^5C or Ψ modifications, which as been associated with exonic regions and splice sites respectively (Delatte et al., 2016; Chen et al., 2010).

We determined that the modification landscape of individual genes varies based on developmental state and is significant enough to drive differences in cell clustering (Fig 2E). This is consistent with previous studies that have used immunoprecipitation methods to compare the m^6A profiles between different cell and tissue types. Zhang *et al*. have demonstrated that different human tissues including the brain, liver, lung, spleen, heart, muscle, testis, and kidneys have distinct m^6A profiles (H. Zhang et al., 2020). The m^6A modification has also been shown to be necessary for various differentiation processes. Mice T cells that are deficient for METTL3, the protein responsible for writing m^6A modifications, have their homeostasis patterns disrupted and undergo proliferation (H.-B. Li et al., 2017). Conversely, in order for human embryonic stem cells to exit the primed state and undergo endoderm differentiation, the m^6A modification is required and elimination of this modification will maintain cellular homeostasis (Batista et al., 2014). We have shown here that both m^6A and global modification patterns change across the naive to primed pluripotency spectrum, occurring earlier in human development than previously

demonstrated. It is likely that changes in m^6A landscape occur even earlier in development than the naïve pluripotency states captured by the 5i/L/A and 2i/L/I/F growth conditions. While it has not yet been demonstrated in human cells, it has been observed that zebrafish require m^6A in order to facilitate the maternal-to-zygotic transition. The removal of m^6A binding protein Ythdf2 decreases the turnover of m^6A modified transcripts in the maternal state and inhibits activation of the zygotic genome (Zhao et al., 2017).

In addition to the unique modification landscapes detected, there was also a large set of overlapping, highly modified genes shared between each of the three developmental states. Common themes among their GO biological processes and Reactome pathways were related to the cell cycle, transcription, and developmental processes. This correlates with the previous study of the m^6A modification in primed H1 hESCs obtained with m^6A RIP-seq, which found enrichment of the GO biological process categories "regulation of transcription" and "regulation of cell morphogenesis" (Batista et al., 2014). Additionally, gene set enrichment analysis found high enrichment of the m^6A modification on transcripts related to cell cycle (Batista et al., 2014).

Similarly, a previous study of the m^1A modification in a human cell line revealed that the m^1A modification affected genes related to binding activity using GO molecular function analysis. More specifically, protein binding, poly(A) RNA binding, RNA binding, mRNA binding, ubiquitin protein ligase binding, nucleotide binding and DNA binding are functions associated with transcripts containing the m^1A modification (X. Li et al., 2017). Our GO molecular function analysis on the set of highly modified transcripts shared between all developmental states indicated similar enrichment for binding activity.

Less studied is the m^5C modification on mRNA, which has been associated with a stabilization of mRNA and decrease of transcript turnover (X. Zhang et al., 2012). A previous study of m^5C has associated the modification with GO biological processes for "translation" and "peptide metabolic process" across multiple tissue types (Yang et al., 2017). Expression of the NSUN2 writer protein, along with writer, reader, and eraser proteins for the m^1A and m^6A modifications, was detected across all three pluripotent states (Fig 4F). Overall, the functions of modified transcripts detected in our epitranscriptome map contain known functions from previous assays of individual modifications, a sign that nanopore sequencing is a promising assay for profiling multiple modifications simultaneously.

Future studies will need to focus on differentiating the distinct modifications from the pooled modification landscape generated by nanopore sequencing. For m^6A this is already possible given its strong association with the DRACH motif. Motif sequences for other modifications are less defined. However, we have noticed that there are other enriched 5-mer sequences in our list of modified regions. Additionally, some of these enriched 5-mers are

associated with various RNA binding proteins, many of which also have gene expression in our data (Fig 4D-E). It is possible that these RNA binding proteins are influenced by modification status and that links exist between these binding proteins and RNA modifications. As these links become clearer, more RNA modifications will become associated with particular motifs.

We also observed that there are unique distributions of modified 5-mers among different developmental states. For example, we observed that certain motifs for the RNA binding proteins HNRNPL, NOVA1, PUM1, RBM4, RBM41, RBM45, and SFRS1 are enriched in only one developmental state. Similarly, we observe that the DRACH motif GGACT is almost twice as enriched in the mTeSR state (fold-change=6.4) than in the 5i/L/A state (fold-change=3.3). Previous studies using m^6A RIP-seq have shown that the DRACH 5-mer AAACH is strongly enriched in cortex tissue, whereas the GAACH motif is enriched in liver tissue (H. Zhang et al., 2020).

Additionally, we were able to detect novel m^6A modifications not detected from previous studies. Analysis of m^6A RIP-seq overlapping peaks vary 30-60% between experiments of the same cell type (McIntyre et al., 2020). Even with this limitation, 66.4% of modified DRACH positions were on genes previously reported to have the m^6A modification. The set of novel DRACH-modified genes share many functional similarities with the set of previously reported m^6A containing genes. This combined with the high accuracy rate of ELIGOS specifically for DRACH motifs, supports the premise that this is not merely a set of false positive detections and nanopore sequencing is able to detect a more expansive set of m^6A modifications.

The m^6A modification is the most well-studied RNA modification, yet its role in pluripotency is still unclear. Here, we have demonstrated that the modification landscapes between different pluripotent states is unique, which could be the cause of the conflicting results between previous experiments. Additionally, the link between RNA binding proteins and RNA modifications is poorly understood. The ability to link enriched modified k-mers to RNA binding protein sites will be an aide in discovering how modifications impact the function of these binding proteins. Overall, we believe that use of nanopore sequencing to detect RNA modifications will lead to a deeper understanding of the difference in epitranscriptome landscapes between cell states and is a promising approach for future epitranscriptomic studies.

5.5 Methods

Cell Culture

Elf1 and WIBR3 cells were cultured in 5i/L/A, 2i/L/I/F, and mTeSR1 (Stem Cell Technologies) conditions. Naive hESCs (5i/L/A and 2i/L/I/F) were grown on g-irradiated primary mouse

embryonic fibroblasts (MEFs), dose 3000g. Before hESC plating, the irradiated fibroblasts were "ghosted" with a lysis mix (PBS containing 0.5% SDS, 5mM NH_4OH) (Abbondanzo et al., 1993) to create a cell-free feeder matrix. Primed hESCs (mTeSR1 from StemCell Technologies) were grown on Matrigel. All cultures were incubated at 37ºC in 5% $C0_2$, 5% O_2, 90% N_2 and were passaged with accutase. Specific inhibitors used in the naive conditions were performed as previously described (Sperber et al., 2015). Cells were harvested with trypsin, rinsed with PBS, and snap frozen before library preparation.

Library Preparation and Sequencing

Total RNA was isolated using the AllPrep DNA/RNA Mini Kit (Qiagen). mRNA was purified by performing two rounds of cleanup with Dynabeads mRNA Purification kit (Invitrogen). Direct RNA sequencing is performed using the SQK-RNA001 kit from Oxford Nanopore Technologies (ONT). 500 ng of each sample was used for each run. Library preparation was performed according to manufacturer's instructions (version: DRS_9026_v1_revP_15Dec2016, ONT). In brief, this protocol added adapters for reverse transcription off of the poly-A tail. Reverse transcription was performed with SuperScript III reverse transcriptase (ThermoFisher Scientific) to unravel and RNA secondary structure. ONT adapters were ligated to the RNA and the primary RNA strand was sequenced on the MinION. Each sample was run on its own flow cell.

Bioinformatic Pipeline

Basecalling the raw data is performed using Guppy (v3.2.2) from ONT. Mapping was performed with Minimap2 by aligning transcripts to the human reference genome hg19 (options -ax --secondary=no) (v2.17) (H. Li 2018). Modified bases were detected using ELIGOS (function: eligos2 rna_mod) (v2.0.0) (Jenjaroenpun et al., 2021). Signal visualizations were generated with Tombo (function: tombo plot genome_locations) (v1.5.1) (Stoiber et al., 2016). All gene annotations are from GENCODE GRCh37, version 19 (Ensembl 74). The list of pluripotency-relevant genes was obtained from GSEA, gene set MUELLER_PLURINET (Müller et al., 2008). Relevant functions of individual genes were calculated by GeneWalk (Ietswaart et al., 2021). Functions of gene sets were calculated with WebGestalt (Liao et al., 2019). PCA from gene expression was performed with PCAExplorer (Marini and Binder 2019).

5.6 Figures

Figure 1

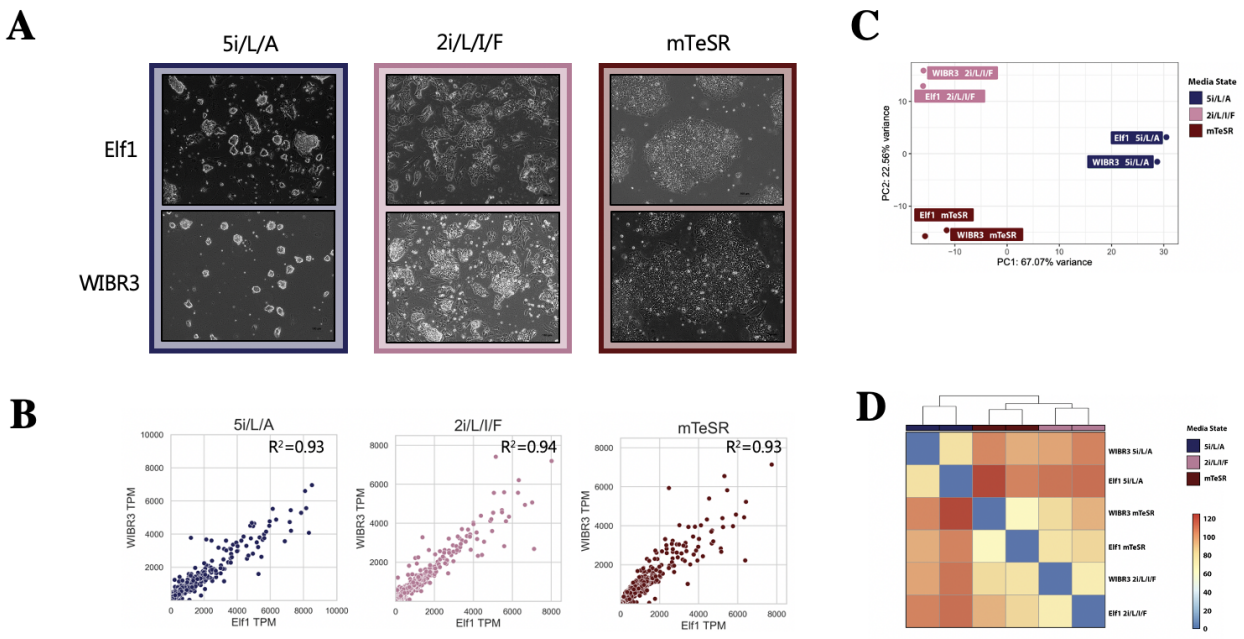


Figure 1: Distinct Stem Cell States Profiled by Direct Nanopore Sequencing. (A) Early naïve (5i/L/A), late naïve (2i/L/I/F) and primed (mTeSR) stem cells used in this study. Two cell types (Elf1 and WIBR3) were sequenced (direct RNA) for each state. 10X magnification. **(B)** Correlation of biological replicates using transcripts per million (TPM). R-squared values are between 0.93-0.94. **(C)** Principal component analysis. Samples cluster by media state; **(D)** Sample-to-sample heatmap measuring Euclidean distance between samples. Samples cluster by media state. 5i/L/A condition creates the most dissimilar gene expression profile.

Figure 2

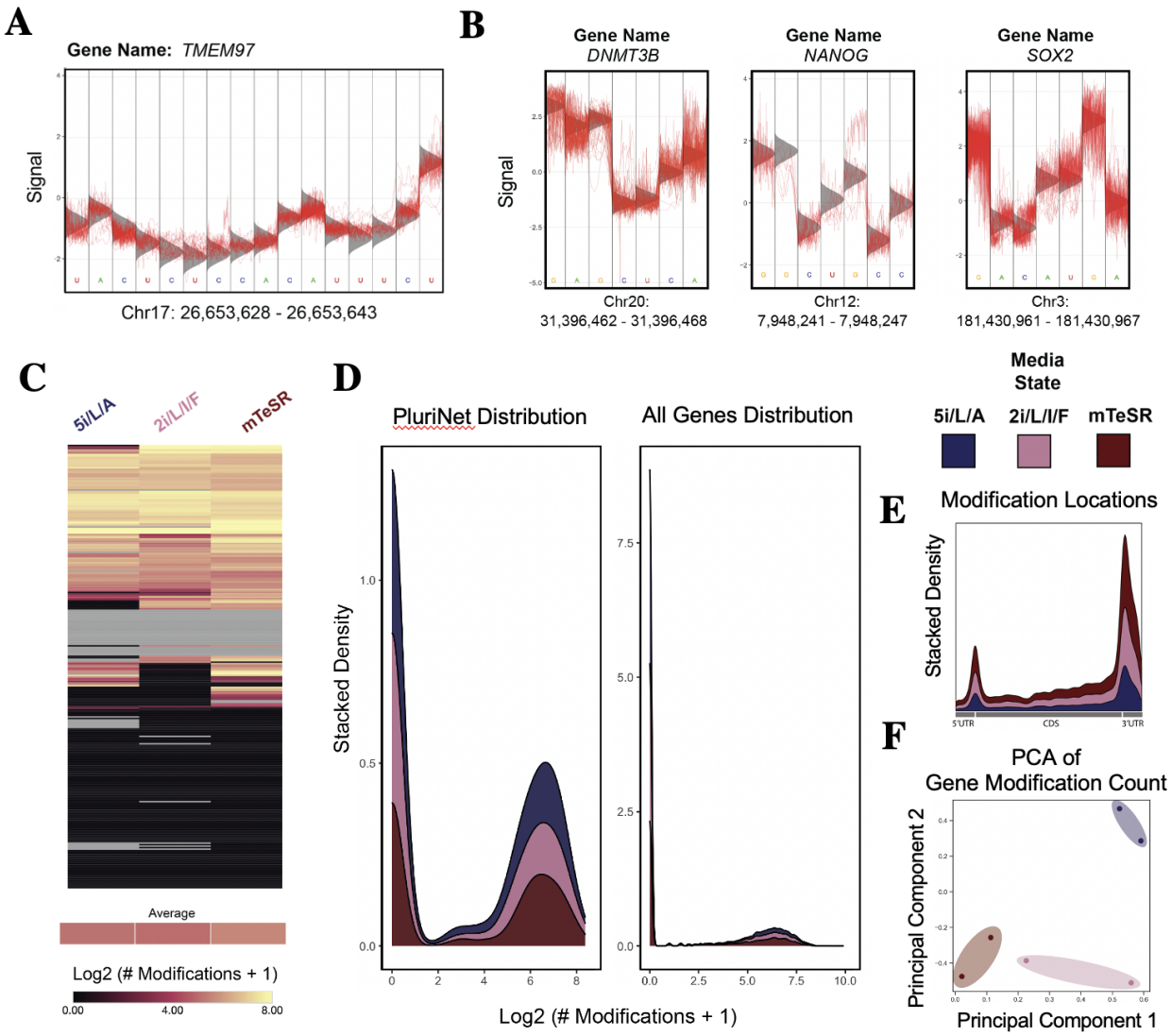


Figure 2: Epitranscriptomic Modification Landscape Detected from Current Aberrations. **(A)** An unmodified region of the transcriptome detected by ELIGOS on the gene *TMEM97*. Position given in genomic coordinates. **(B)** Three modified regions of the transcriptome detected by ELIGOS on pluripotency-related genes (*DNMT3B, NANOG, SOX2*). Position given in genomic coordinates. **(C)** Modification counts of the 299 PluriNet genes for each stem cell state, with one gene per row. The average number of modifications per gene across all expressed genes (including non-PluriNet genes) is displayed on the bottom. Gray cells indicate no expression of the PluriNet gene. **(D)** The density distributions of modification counts per gene. The left panel is specific for the PluriNet set. The right panel contains all genes not in PluriNet. **(E)** Kernel density (kde) distribution of modification locations within transcripts with respect to the boundaries of the transcript coding sequence (CDS) with the 5' UTR and 3' UTR. Reads were filtered for full length transcripts. Transcript lengths were normalized within each segment type (5'UTR, CDS, 3'UTR). Heights are "stacked" for visualization and density heights between

samples are comparable. **(F)** Principal component analysis (PCA) of samples based on the number of modifications per gene in each sample. Genes were filtered for detection of expression in all three cell state and the genes with the top 200 variances were used for analysis.

Figure 3

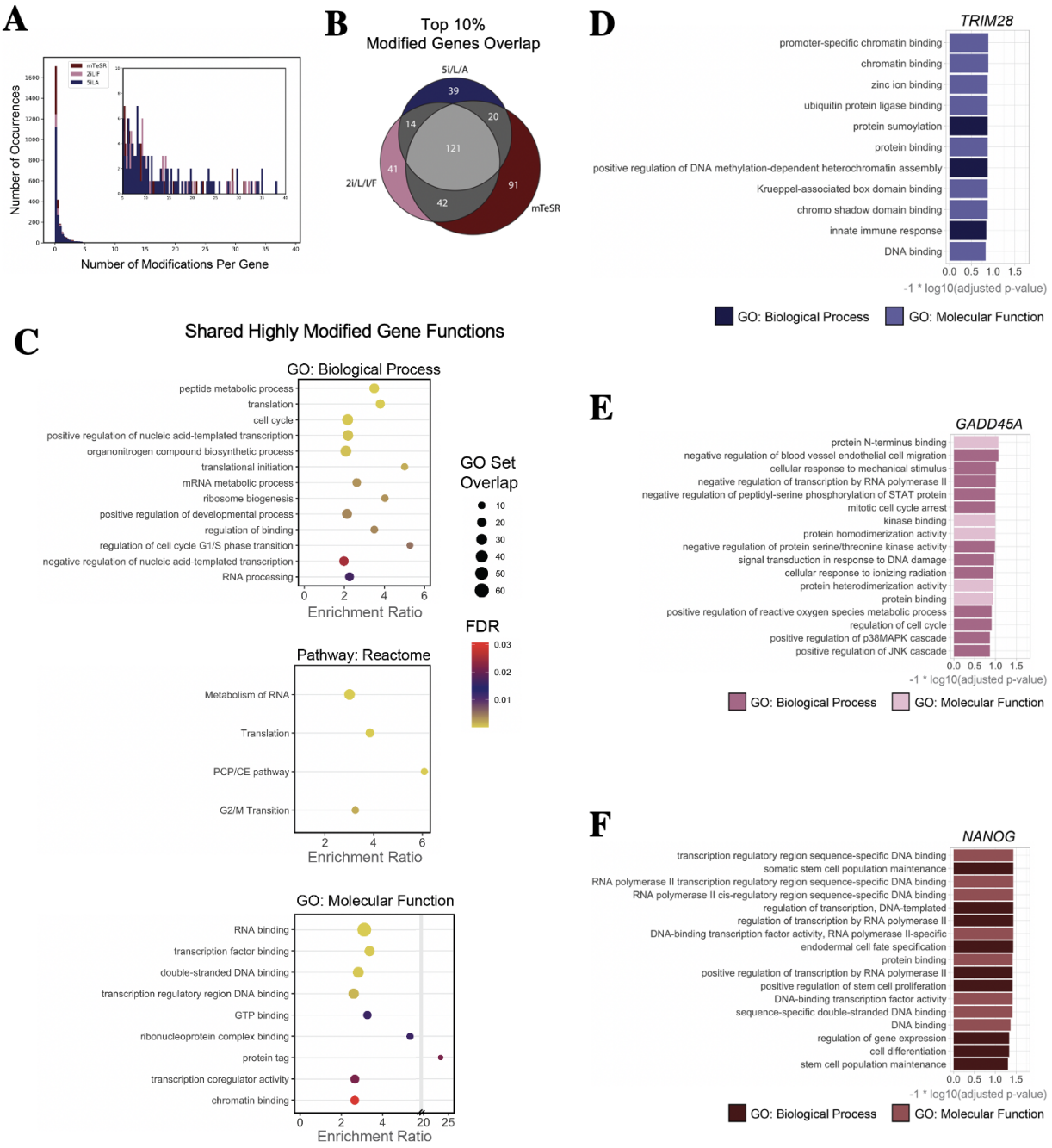


Figure 3: Functions of Highly Modified Genes. (A) Distribution of the number of modifications per 100 bases. Inset plot is zoomed-in on 5+ modifications per 100 bases. **(B)** The

number of shared and distinct genes with modification enrichment (top 10%) between cell states. **(C)** WebGestalt overrepresentation analysis for the set of all highly modified genes. **(D-F)** GeneWalk annotations for *TRIM28* (5i/L/A), *GADD45A* (2i/L/I/F), and *NANOG* (mTeSR).

Figure 4

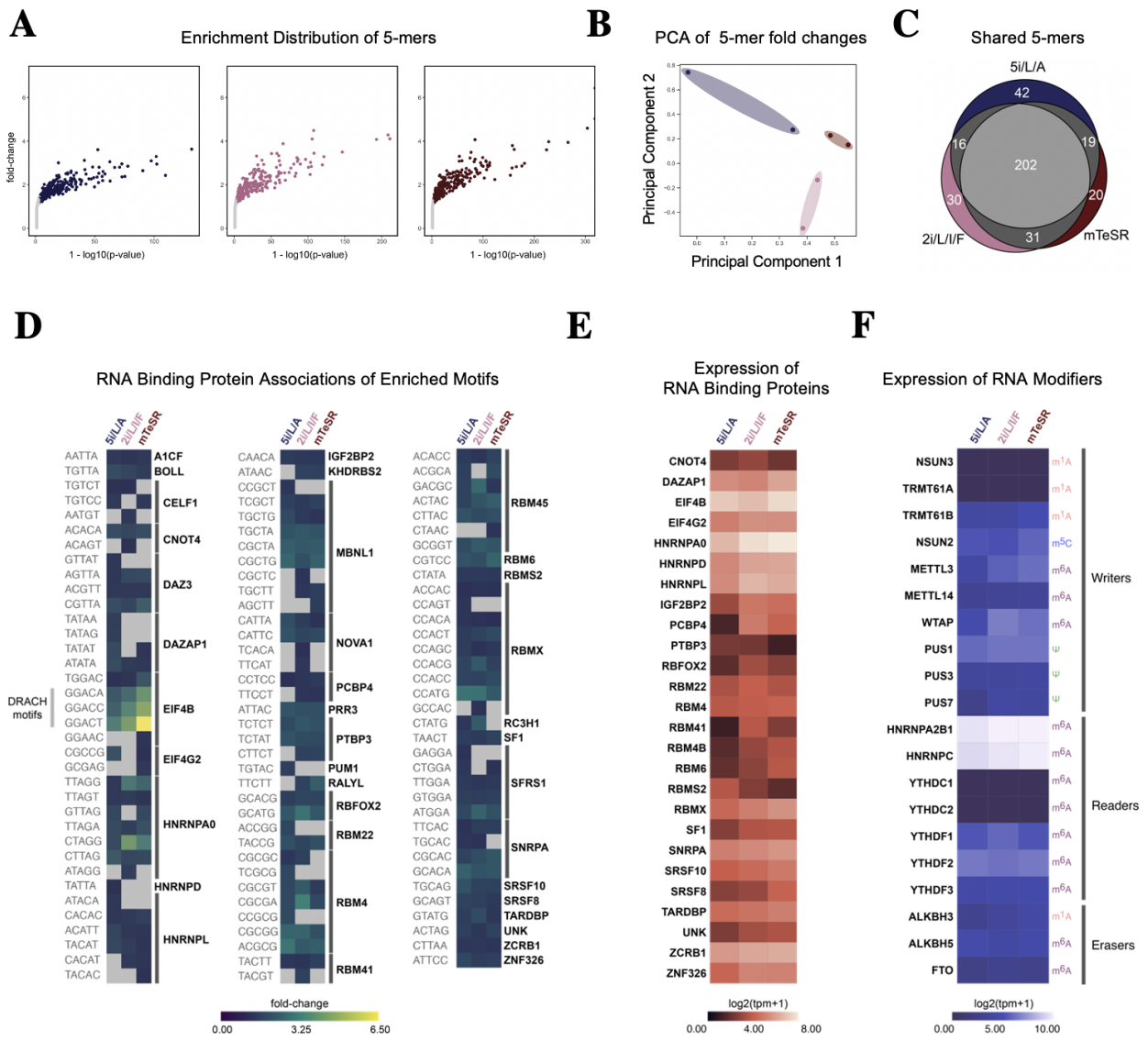


Figure 4: Motif Analysis. (A) Fold change distribution of all 864 5-mers with homopolymer sequences of 3 or greater removed. Fold change was calculated by dividing the number of modified 5-mers by the expected number of 5-mers from the sequencing experiment background. 5-mers with a p-values less than 0.001 are colored. Dashed line indicates fold-change of 1 (i.e. no fold-change) **(B)** Principal component analysis of the number of modified 5-mer occurrences

for all 864 5-mers with homopolymer sequences of 3 or greater removed. PCA was performed on fold-change values. The fold-change values with the top 200 variances were used for PCA. **(C)** The number of shared and unique 5-mers detected between cell states. **(D)** 38 RNA binding proteins associated with the enriched 5-mer motifs (p-value < 0.001). Heatmap intensity is determined by fold-change value. DRACH motifs, associated with the m^{6}A modification, are annotated. **(E)** Gene expression log-transformed TPM of the 26 RNA binding protein genes with motif representation in enriched 5-mers in each of the three cell states. **(F)** Expression of known writer, reader, and eraser proteins of RNA modifications from the RNAWRE database.

Figure 5

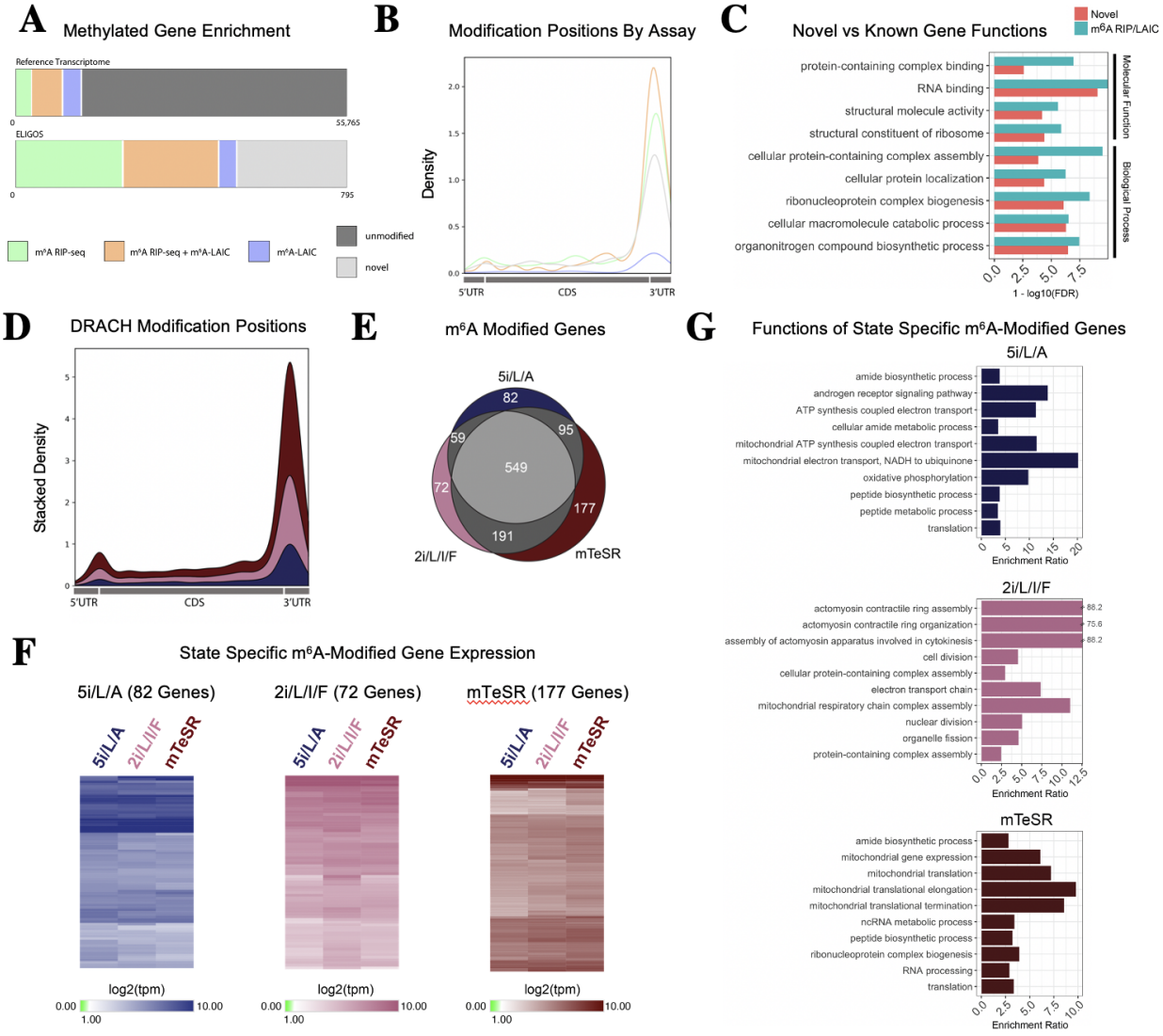
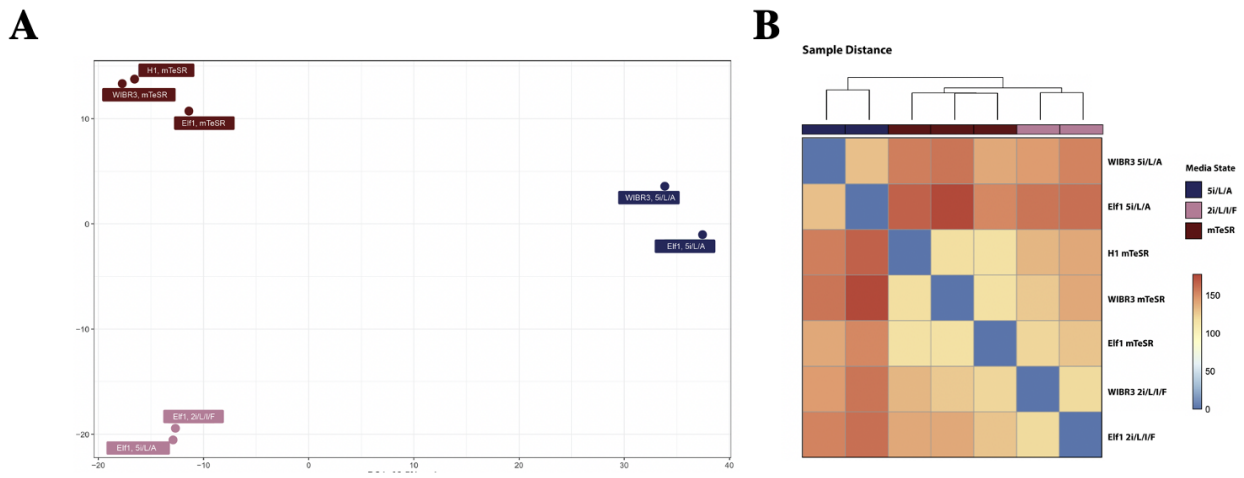


Figure 5: m^{6}A Modification Profile. (A) The number of genes with modifications compared to the number of genes in the reference genome from m^{6}A RIP-seq and m^{6}A-LAIC studies (top). The number of genes detected by ELIGOS with the DRACH motif that overlap with m^{6}A RIP-seq and m^{6}A-LAIC assays, and the number of genes that are novel modification sites in H1 cells in the mTeSR state (bottom). **(B)** Locational distribution of modified DRACH positions that are in concurrence with the m^{6}A RIP-seq, m^{6}a-LAIC assay, and novel modification candidates. Positions were normalized with respect to the 5' and 3' UTRs. **(C)** Shared GO biological process and molecular function categories between the novel genes and genes identified with m^{6}A RIP-seq and m^{6}A-LAIC studies. **(D)** Locational distribution of DRACH positions in ELIGOs predictions in the 5i/L/A, 2i/L/I/F, and mTeSR states with respect to the 5' and 3' UTRs by cell type. **(E)** Similarity of m^{6}A modified genes between cell states. Genes were first filtered for detectable expression in all three states. **(F)** Expression of state-specific m^{6}A -modified genes. **(G)** Gene ontology (GO) biological processes overrepresentation analysis on the state-specific m^{6}A modified genes

.

5.7 Supplemental Figures

Supplemental Figure 1



Supplemental Figure S1: H1 Clustering. (A-B) H1 cells cultured in the mTeSR state cluster with other cell lines grown in the mTeSR state in PCA (A) and Euclidean distance (B).

5.8 References

Abbondanzo SJ, Gadi I, Stewart CL. Derivation of embryonic stem cell lines. Methods Enzymol. 1993;225:803-23. doi: 10.1016/0076-6879(93)25052-4. PMID: 8231888.

Aguilo, Francesca, Fan Zhang, Ana Sancho, Miguel Fidalgo, Serena Di Cecilia, Ajay Vashisht, Dung-Fang Lee, et al. 2015. "Coordination of m6A mRNA Methylation and Gene

Transcription by ZFP217 Regulates Pluripotency and Reprogramming." *Cell Stem Cell* 17 (6): 689–704.
Batista, Pedro J., Benoit Molinie, Jinkai Wang, Kun Qu, Jiajing Zhang, Lingjie Li, Donna M. Bouley, et al. 2014. "m6A RNA Modification Controls Cell Fate Transition in Mammalian Embryonic Stem Cells." *Cell Stem Cell* 15 (6): 707–19.
Battle, Stephanie L., Naresh Doni Jayavelu, Robert N. Azad, Jennifer Hesson, Faria N. Ahmed, Eliah G. Overbey, Joseph A. Zoller, et al. 2019. "Enhancer Chromatin and 3D Genome Architecture Changes from Naive to Primed Human Embryonic Stem Cell States." *Stem Cell Reports* 12 (5): 1129–44.
Boccaletto, Pietro, Magdalena A. Machnicka, Elzbieta Purta, Pawel Piatkowski, Blazej Baginski, Tomasz K. Wirecki, Valérie de Crécy-Lagard, et al. 2018. "MODOMICS: A Database of RNA Modification Pathways. 2017 Update." *Nucleic Acids Research* 46 (D1): D303–7.
Chen, Chun, Xinliang Zhao, Ryszard Kierzek, and Yi-Tao Yu. 2010. "A Flexible RNA Backbone within the Polypyrimidine Tract Is Required for U2AF65 Binding and Pre-mRNA Splicing InVivo." *Molecular and Cellular Biology* 30 (17): 4108–19.
Church, Chris, Lee Moir, Fiona McMurray, Christophe Girard, Gareth T. Banks, Lydia Teboul, Sara Wells, et al. 2010. "Overexpression of Fto Leads to Increased Food Intake and Results in Obesity." *Nature Genetics* 42 (12): 1086–92.
Cook, Kate B., Hilal Kazan, Khalid Zuberi, Quaid Morris, and Timothy R. Hughes. 2011. "RBPDB: A Database of RNA-Binding Specificities." *Nucleic Acids Research* 39 (Database issue): D301–8.
Delatte, B., F. Wang, L. V. Ngoc, E. Collignon, E. Bonvin, R. Deplus, E. Calonne, et al. 2016. "Transcriptome-Wide Distribution and Function of RNA Hydroxymethylcytosine." *Science*. https://doi.org/10.1126/science.aac5253.
Dominissini, Dan, Sigrid Nachtergaele, Sharon Moshitch-Moshkovitz, Eyal Peer, Nitzan Kol, Moshe Shay Ben-Haim, Qing Dai, et al. 2016. "The Dynamic N1-Methyladenosine Methylome in Eukaryotic Messenger RNA." *Nature*. https://doi.org/10.1038/nature16998.
Garalde, Daniel R., Elizabeth A. Snell, Daniel Jachimowicz, Botond Sipos, Joseph H. Lloyd, Mark Bruce, Nadia Pantic, et al. 2018. "Highly Parallel Direct RNA Sequencing on an Array of Nanopores." *Nature Methods* 15 (3): 201–6.
Geula, Shay, Sharon Moshitch-Moshkovitz, Dan Dominissini, Abed Alfatah Mansour, Nitzan Kol, Mali Salmon-Divon, Vera Hershkovitz, et al. 2015. "Stem Cells. m6A mRNA Methylation Facilitates Resolution of Naïve Pluripotency toward Differentiation." *Science* 347 (6225): 1002–6.
Heurtier, Victor, Nick Owens, Inma Gonzalez, Florian Mueller, Caroline Proux, Damien Mornico, Philippe Clerc, Agnes Dubois, and Pablo Navarro. 2019. "The Molecular Logic of Nanog-Induced Self-Renewal in Mouse Embryonic Stem Cells." *Nature Communications*. https://doi.org/10.1038/s41467-019-09041-z.
Hsu, Phillip J., Yunfei Zhu, Honghui Ma, Yueshuai Guo, Xiaodan Shi, Yuanyuan Liu, Meijie Qi, et al. 2017. "Ythdc2 Is an N6-Methyladenosine Binding Protein That Regulates Mammalian Spermatogenesis." *Cell Research* 27 (9): 1115–27.
Ietswaart, R., B. M. Gyori, J. A. Bachman, and P. K. Sorger. 2021. "GeneWalk Identifies Relevant Gene Functions for a Biological Context Using Network Representation Learning." *Genome Biology*. https://genomebiology.biomedcentral.com/articles/10.1186/s13059-021-02264-8.

Ivanova, Ivayla, Christian Much, Monica Di Giacomo, Chiara Azzi, Marcos Morgan, Pedro N. Moreira, Jack Monahan, Claudia Carrieri, Anton J. Enright, and Dónal O'Carroll. 2017. "The RNA M 6 A Reader YTHDF2 Is Essential for the Post-Transcriptional Regulation of the Maternal Transcriptome and Oocyte Competence." *Molecular Cell*. https://doi.org/10.1016/j.molcel.2017.08.003.

Jain, Miten, Hugh E. Olsen, Benedict Paten, and Mark Akeson. 2016. "The Oxford Nanopore MinION: Delivery of Nanopore Sequencing to the Genomics Community." *Genome Biology* 17 (1): 239.

Jenjaroenpun, Piroon, Thidathip Wongsurawat, Taylor D. Wadley, Trudy M. Wassenaar, Jun Liu, Qing Dai, Visanu Wanchai, et al. 2021. "Decoding the Epitranscriptional Landscape from Native RNA Sequences." *Nucleic Acids Research*. https://doi.org/10.1093/nar/gkaa620.

Jia, Guifang, Ye Fu, Xu Zhao, Qing Dai, Guanqun Zheng, Ying Yang, Chengqi Yi, et al. 2011. "N6-Methyladenosine in Nuclear RNA Is a Major Substrate of the Obesity-Associated FTO." *Nature Chemical Biology* 7 (12): 885–87.

Jonkhout, Nicky, Julia Tran, Martin A. Smith, Nicole Schonrock, John S. Mattick, and Eva Maria Novoa. 2017. "The RNA Modification Landscape in Human Disease." *RNA* 23 (12): 1754–69.

Karijolich, John, and Yi-Tao Yu. 2011. "Converting Nonsense Codons into Sense Codons by Targeted Pseudouridylation." *Nature* 474 (7351): 395–98.

Klimczak, Marta, Patrycja Czerwińska, Sylwia Mazurek, Barbara Sozańska, Przemysław Biecek, Andrzej Mackiewicz, and Maciej Wiznerowicz. 2017. "TRIM28 Epigenetic Corepressor Is Indispensable for Stable Induced Pluripotent Stem Cell Formation." *Stem Cell Research* 23 (August): 163–72.

Lee, Yujin, Junho Choe, Ok Hyun Park, and Yoon Ki Kim. 2020. "Molecular Mechanisms Driving mRNA Degradation by m6A Modification." *Trends in Genetics: TIG* 36 (3): 177–88.

Liao, Yuxing, Jing Wang, Eric J. Jaehnig, Zhiao Shi, and Bing Zhang. 2019. "WebGestalt 2019: Gene Set Analysis Toolkit with Revamped UIs and APIs." *Nucleic Acids Research* 47 (W1): W199–205.

Li, Heng. 2018. "Minimap2: Pairwise Alignment for Nucleotide Sequences." *Bioinformatics* , May. https://doi.org/10.1093/bioinformatics/bty191.

Li, Hua-Bing, Jiyu Tong, Shu Zhu, Pedro J. Batista, Erin E. Duffy, Jun Zhao, Will Bailis, et al. 2017. "m6A mRNA Methylation Controls T Cell Homeostasis by Targeting the IL-7/STAT5/SOCS Pathways." *Nature* 548 (7667): 338–42.

Lin, Shuibin, Junho Choe, Peng Du, Robinson Triboulet, and Richard I. Gregory. 2016. "The m6A Methyltransferase METTL3 Promotes Translation in Human Cancer Cells." *Molecular Cell* 62 (3): 335–45.

Liu, Fange, Wesley Clark, Guanzheng Luo, Xiaoyun Wang, Ye Fu, Jiangbo Wei, Xiao Wang, et al. 2016. "ALKBH1-Mediated tRNA Demethylation Regulates Translation." *Cell* 167 (7): 1897.

Liu, Jianzhao, Yanan Yue, Dali Han, Xiao Wang, Ye Fu, Liang Zhang, Guifang Jia, et al. 2014. "A METTL3-METTL14 Complex Mediates Mammalian Nuclear RNA N6-Adenosine Methylation." *Nature Chemical Biology* 10 (2): 93–95.

Liu, Jun, Mark A. Eckert, Bryan T. Harada, Song-Mei Liu, Zhike Lu, Kangkang Yu, Samantha M. Tienda, et al. 2018. "m6A mRNA Methylation Regulates AKT Activity to Promote the

Proliferation and Tumorigenicity of Endometrial Cancer." *Nature Cell Biology*. https://doi.org/10.1038/s41556-018-0174-4.

Li, Xiaoyu, Xushen Xiong, Kun Wang, Lixia Wang, Xiaoting Shu, Shiqing Ma, and Chengqi Yi. 2016. "Transcriptome-Wide Mapping Reveals Reversible and Dynamic N1-Methyladenosine Methylome." *Nature Chemical Biology*. https://doi.org/10.1038/nchembio.2040.

Li, Xiaoyu, Xushen Xiong, Meiling Zhang, Kun Wang, Ying Chen, Jun Zhou, Yuanhui Mao, et al. 2017. "Base-Resolution Mapping Reveals Distinct m1A Methylome in Nuclear- and Mitochondrial-Encoded Transcripts." *Molecular Cell*. https://doi.org/10.1016/j.molcel.2017.10.019.

Loh, Yuin-Han, Qiang Wu, Joon-Lin Chew, Vinsensius B. Vega, Weiwei Zhang, Xi Chen, Guillaume Bourque, et al. 2006. "The Oct4 and Nanog Transcription Network Regulates Pluripotency in Mouse Embryonic Stem Cells." *Nature Genetics* 38 (4): 431–40.

Ludwig, Tenneille, and James A. Thomson. 2007. "Defined, Feeder-Independent Medium for Human Embryonic Stem Cell Culture." *Current Protocols in Stem Cell Biology*. https://doi.org/10.1002/9780470151808.sc01c02s2.

Ma, J., F. Yang, C. Zhou, F. Liu, J. Yuan, and F. Wang. 2017. "METTL14 Suppresses the Metastatic Potential of Hepatocellular Carcinoma by Modulating N6‐methyladenosine‐dependent Primary MicroRNA Processing." https://aasldpubs.onlinelibrary.wiley.com/doi/abs/10.1002/hep.28885.

Mao, Yuanhui, Leiming Dong, Xiao-Min Liu, Jiayin Guo, Honghui Ma, Bin Shen, and Shu-Bing Qian. 2019. "m6A in mRNA Coding Regions Promotes Translation via the RNA Helicase-Containing YTHDC2." *Nature Communications* 10 (1): 5332.

Marini, Federico, and Harald Binder. 2019. "pcaExplorer: An R/Bioconductor Package for Interacting with RNA-Seq Principal Components." *BMC Bioinformatics* 20 (1): 331.

McIntyre, Alexa B. R., Nandan S. Gokhale, Leandro Cerchietti, Samie R. Jaffrey, Stacy M. Horner, and Christopher E. Mason. 2020. "Limits in the Detection of m6A Changes Using MeRIP/m6A-Seq." *Scientific Reports* 10 (1): 6590.

Molinie, Benoit, Jinkai Wang, Kok Seong Lim, Roman Hillebrand, Zhi-Xiang Lu, Nicholas Van Wittenberghe, Benjamin D. Howard, et al. 2016. "M 6 A-LAIC-Seq Reveals the Census and Complexity of the M 6 A Epitranscriptome." *Nature Methods* 13 (8): 692–98.

Müller, Franz-Josef, Louise C. Laurent, Dennis Kostka, Igor Ulitsky, Roy Williams, Christina Lu, In-Hyun Park, et al. 2008. "Regulatory Networks Define Phenotypic Classes of Human Stem Cell Lines." *Nature* 455 (7211): 401–5.

Nichols, Jennifer, and Austin Smith. 2009. "Naive and Primed Pluripotent States." *Cell Stem Cell* 4 (6): 487–92.

Nie, Fulei, Pengmian Feng, Xiaoming Song, Meng Wu, Qiang Tang, and Wei Chen. 2020. "RNAWRE: A Resource of Writers, Readers and Erasers of RNA Modifications." *Database: The Journal of Biological Databases and Curation* 2020 (January). https://doi.org/10.1093/database/baaa049.

Pan, Guangjin, and James A. Thomson. 2007. "Nanog and Transcriptional Networks in Embryonic Stem Cell Pluripotency." *Cell Research* 17 (1): 42–49.

Parker, Matthew T., Katarzyna Knop, Anna V. Sherwood, Nicholas J. Schurch, Katarzyna Mackinnon, Peter D. Gould, Anthony Jw Hall, Geoffrey J. Barton, and Gordon G. Simpson. 2020. "Nanopore Direct RNA Sequencing Maps the Complexity of Arabidopsis mRNA

Processing and m6A Modification." *eLife* 9 (January). https://doi.org/10.7554/eLife.49658.
Price, Alexander M., Katharina E. Hayer, Alexa B. R. McIntyre, Nandan S. Gokhale, Jonathan S. Abebe, Ashley N. Della Fera, Christopher E. Mason, et al. 2020. "Direct RNA Sequencing Reveals m6A Modifications on Adenovirus RNA Are Necessary for Efficient Splicing." *Nature Communications* 11 (1): 6016.
Roundtree, Ian A., Molly E. Evans, Tao Pan, and Chuan He. 2017. "Dynamic RNA Modifications in Gene Expression Regulation." *Cell* 169 (7): 1187–1200.
Schüle, Katrin M., Manuel Leichsenring, Tommaso Andreani, Viviana Vastolo, Medhavi Mallick, Michael U. Musheev, Emil Karaulanov, and Christof Niehrs. 2019. "GADD45 Promotes Locus-Specific DNA Demethylation and 2C Cycling in Embryonic Stem Cells." *Genes & Development* 33 (13-14): 782–98.
Simms, Carrie L., Benjamin H. Hudson, John W. Mosior, Ali S. Rangwala, and Hani S. Zaher. 2014. "An Active Role for the Ribosome in Determining the Fate of Oxidized mRNA." *Cell Reports* 9 (4): 1256–64.
Smith, Andrew M., Miten Jain, Logan Mulroney, Daniel R. Garalde, and Mark Akeson. 2019. "Reading Canonical and Modified Nucleobases in 16S Ribosomal RNA Using Nanopore Native RNA Sequencing." *PloS One* 14 (5): e0216709.
Sperber, Henrik, Julie Mathieu, Yuliang Wang, Amy Ferreccio, Jennifer Hesson, Zhuojin Xu, Karin A. Fischer, et al. 2015. "The Metabolome Regulates the Epigenetic Landscape during Naive-to-Primed Human Embryonic Stem Cell Transition." *Nature Cell Biology* 17 (12): 1523–35.
Squires, Jeffrey E., Hardip R. Patel, Marco Nousch, Tennille Sibbritt, David T. Humphreys, Brian J. Parker, Catherine M. Suter, and Thomas Preiss. 2012. "Widespread Occurrence of 5-Methylcytosine in Human Coding and Non-Coding RNA." *Nucleic Acids Research* 40 (11): 5023–33.
Stoiber, M., J. Quick, R. Egan, J. E. Lee, and S. Celniker. 2016. "De Novo Identification of DNA Modifications Enabled by Genome-Guided Nanopore Signal Processing." *BioRxiv*. https://www.biorxiv.org/content/10.1101/094672v1.abstract.
Tanzer, Andrea, Ivo L. Hofacker, and Ronny Lorenz. 2019. "RNA Modifications in Structure Prediction – Status Quo and Future Challenges." *Methods*. https://doi.org/10.1016/j.ymeth.2018.10.019.
Theunissen, Thorold W., Benjamin E. Powell, Haoyi Wang, Maya Mitalipova, Dina A. Faddah, Jessica Reddy, Zi Peng Fan, et al. 2014. "Systematic Identification of Culture Conditions for Induction and Maintenance of Naive Human Pluripotency." *Cell Stem Cell* 15 (4): 524–26.
Van Nostrand, Eric L., Peter Freese, Gabriel A. Pratt, Xiaofeng Wang, Xintao Wei, Rui Xiao, Steven M. Blue, et al. 2020. "A Large-Scale Binding and Functional Map of Human RNA-Binding Proteins." *Nature* 583 (7818): 711–19.
Ware, Carol B., Angelique M. Nelson, Brigham Mecham, Jennifer Hesson, Wenyu Zhou, Erica C. Jonlin, Antonio J. Jimenez-Caliani, et al. 2014. "Derivation of Naive Human Embryonic Stem Cells." *Proceedings of the National Academy of Sciences of the United States of America* 111 (12): 4484–89.
Wen, Jing, Ruitu Lv, Honghui Ma, Hongjie Shen, Chenxi He, Jiahua Wang, Fangfang Jiao, et al. 2018. "Zc3h13 Regulates Nuclear RNA m6A Methylation and Mouse Embryonic Stem Cell Self-Renewal." *Molecular Cell* 69 (6): 1028–38.e6.
Workman, Rachael E., Alison D. Tang, Paul S. Tang, Miten Jain, John R. Tyson, Roham

Razaghi, Philip C. Zuzarte, et al. 2019. “Nanopore Native RNA Sequencing of a Human poly(A) Transcriptome.” *Nature Methods* 16 (12): 1297–1305.
Wu, Ruifan, Youhua Liu, Yuanling Zhao, Zhen Bi, Yongxi Yao, Qing Liu, Fengqin Wang, Yizhen Wang, and Xinxia Wang. 2019. “m6A Methylation Controls Pluripotency of Porcine Induced Pluripotent Stem Cells by Targeting SOCS3/JAK2/STAT3 Pathway in a YTHDF1/YTHDF2-Orchestrated Manner.” *Cell Death & Disease*. https://doi.org/10.1038/s41419-019-1417-4.
Yang, Xin, Ying Yang, Bao-Fa Sun, Yu-Sheng Chen, Jia-Wei Xu, Wei-Yi Lai, Ang Li, et al. 2017. “5-Methylcytosine Promotes mRNA Export - NSUN2 as the Methyltransferase and ALYREF as an m5C Reader.” *Cell Research* 27 (5): 606–25.
Zhang, Hui, Xinrui Shi, Tao Huang, Xueni Zhao, Wanying Chen, Nannan Gu, and Rui Zhang. 2020. “Dynamic Landscape and Evolution of m6A Methylation in Human.” *Nucleic Acids Research* 48 (11): 6251–64.
Zhang, Meng, Yanhui Zhai, Sheng Zhang, Xiangpeng Dai, and Ziyi Li. 2020. “Roles of N6-Methyladenosine (m6A) in Stem Cell Fate Decisions and Early Embryonic Development in Mammals.” *Frontiers in Cell and Developmental Biology*. https://doi.org/10.3389/fcell.2020.00782.
Zhang, Xiaotian, Zhenyun Liu, Jie Yi, Hao Tang, Junyue Xing, Minqwei Yu, Tanjun Tong, Yongfeng Shang, Myriam Gorospe, and Wengong Wang. 2012. “The tRNA Methyltransferase NSun2 Stabilizes p16 INK4 mRNA by Methylating the 3′-Untranslated Region of p16.” *Nature Communications* 3 (1): 1–12.
Zhao, Boxuan Simen, Xiao Wang, Alana V. Beadell, Zhike Lu, Hailing Shi, Adam Kuuspalu, Robert K. Ho, and Chuan He. 2017. “m6A-Dependent Maternal mRNA Clearance Facilitates Zebrafish Maternal-to-Zygotic Transition.” *Nature*. https://doi.org/10.1038/nature21355.
Zheng, Guanqun, John Arne Dahl, Yamei Niu, Peter Fedorcsak, Chun-Min Huang, Charles J. Li, Cathrine B. Vågbø, et al. 2013. “ALKBH5 Is a Mammalian RNA Demethylase That Impacts RNA Metabolism and Mouse Fertility.” *Molecular Cell* 49 (1): 18–29.
Zhou, Katherine I., Hailing Shi, Ruitu Lyu, Adam C. Wylder, Żaneta Matuszek, Jessica N. Pan, Chuan He, Marc Parisien, and Tao Pan. 2019. “Regulation of Co-Transcriptional Pre-mRNA Splicing by m6A through the Low-Complexity Protein hnRNPG.” *Molecular Cell* 76 (1): 70–81.e9.

Conclusion

Gene Expression: Natural Variation versus Spaceflight-Induced Perturbation

Throughout the course of my projects, I examined two types of variation due to gene expression. The first was natural variation in gene expression among different cell and tissue types. The second were the affects of perturbation on terminal cell states caused by the environmental factors of spaceflight. One noticeable difference between the differentially

expressed genes was the magnitude of expression differences between cell types vs environmental perturbations.

More specifically, in the chicken intestine data from Chapter 1, 3903 DEGs were observed between the jejunum and ileum, 2306 between the jejunum and proximal cecum, and 4270 between the ileum and the proximal cecum. Similarly, in the chicken immune system, there are 4911 DEGs observed between B cells and monocytes, 5907 between B cells and bursa tissue, and 3951 between bursa and thymus tissue. On the other hand, the number of differentially expressed genes that occur when mice are exposed to hindlimb unloading and low-dose irradiation is 270. However, this was after the mice had four months to recover from exposure, likely reducing the number of differentially expressed genes that would be observed right after exposure. However, for the experiment examining retinal tissue after spaceflight in Chapter 3, RNA was collected within 38 hours of terrestrial return. In this case, 600 DEGs were observed between spaceflight and ground control samples. From these experiments I have observed that the differences in the number of differentially expressed genes between cells with natural variation in gene expression is an order of magnitude larger than the differences observed due to spaceflight.

Since a relatively small number of genes are differentially expressed during spaceflight, the question remains as to whether or not these new gene expression profiles are stable cellular states or whether cells will revert back to their natural gene expression profiles after terrestrial return. Since a smaller number of genes are differentially expressed, we might anticipate changes in epigenomic and epitranscriptomic regulatory mechanisms to also be smaller and potentially capable of shifting back to their natural mechanistic state after exposure to spaceflight has ended. This is of particular interest in the case of the retinal structure changes observed after mice were flown aboard the ISS. Future studies should examine whether mice can recover from the effects of spaceflight and reestablish normal gene expression profiles after returning to Earth, or whether the environmental perturbations have permanently altered cell state and have created irreversible damage. Additionally, the tipping point between the ability of the retinal structure to recover and when a state of irreversible changes has been reached should be explored to determine the duration that astronauts can safely remain in spaceflight without irreversibly changing the structure of their retinal tissue.

For all studies mentioned, the epigenetic profiles of gene expression in *Gallus gallus* and cephalad tissue types exposed to spaceflight environmental factors are relatively unknown. Future studies will need to profile the regulatory mechanisms that create and stabilize these gene expression profiles, such as ATAC-seq, ChIP-seq, and bisulfite-seq in order to characterize chromatin accessibility, histone modifications, and DNA methylation respectively. While a select number of studies have been conducted to study epigenomic profiles of the *Gallus gallus*

genome, these studies are few in number and cover a limited set of tissues. Similarly, the epigenome of tissues exposed to the spaceflight environment are limited. Some studies do exist in the GeneLab database for bisulfite-seq profiling of DNA methylation, however this is limited to a total of 22 studies spanning *Homo sapiens*, *Mus musculus*, and *Arabidopsis thaliana*. Additionally, these studies contain a mixture of simulated spaceflight experiments (i.e. on-Earth simulations) and studies aboard the ISS as well as a mixture of different tissue types. This experimental diversity is useful for determining commonalities across tissues and species attributable to the spaceflight environment, but prevents us from commenting on the reproducibility of any particular experimental result. This problem also persists for RNA-seq data, where around 100 datasets are available. This means that, in addition to performing additional epigenome studies, the experimental capacity of the spaceflight environment must be expanded to accommodate a greater number of experiments. Given the modular nature of the ISS, this can be accomplished by adding additional scientific habitat modules or by utilizing biosatellites for experiments that do not require astronaut oversight.

Limitations of the Murine Model Organism To Study Cephalad Tissue Functions

The two spaceflight environment studies in Chapters 2 and 3 both use mice as a model organism. However, we are most interested in the impact of spaceflight on the human body in order to understand changes in astronaut health. The two cephalad tissues studied, the brain and the retina, have some key differences between humans and mice that may limit the translatability of results from the mouse model to the human astronaut.

While the general cell types composing the brain are similar between humans and mice (Hodge et al., 2019), there are some key differences that must be taken into account. The first is the differences in ion channels (Lowe 2019; Hodge et al., 2019). Given the increased intracranial pressure caused by fluid-shift, changes in ion channels could lead to differential distributions of fluids and ions within the brains of humans and mice. Understanding these changes is of particular importance due to the difference in expression of blood-brain barrier and ion channel genes, such as *Aqp1, Gria3, Grin2a, Calb2, Scn5a, Clcn5, Kcnj13, Kcnn1 and Kcnh1*, observed from hindlimb-unloading and low-dose radiation in Chapter 2. Another consideration are the changes in the ratios of expression of serotonin and glutamate receptor subtypes. Both of these receptors play a role in depression (Moriguchi et al., 2019; Albert and Benkelfat 2013), which may be significant for long duration space travel where astronauts must endure long periods of isolation and limited human interaction (Palinkas 2001). Due to the invasive nature of RNA-seq, it is not possible to perform this assay on human astronauts. Alternative, non-invasive methods for quantifying changes in human brain structure will need to be integrated with the information

that we are able to obtain from mice in order to create a clearer picture of changes in the human brain during spaceflight.

Humans and mice share similar eye and retinal structures, but have some key structural differences. One of these differences is that mice lack a macula, which is a retinal structure that sharpens visual acuity in humans (Volland et al., 2015). As a result, mice are insufficient to use as a model to study any possible links between macular degeneration and the spaceflight environment. Additionally, mice do not have a fovea, an area of concentrated cone cells present in the human retina (Volland et al., 2015). These additional cone cells present in humans may explain why humans do not have the dramatic vision changes that we may expect from mice after observing the structural integrity changes of the mouse retina in Chapter 3. From a body structure standpoint, mice may not be as affected by fluid-shift as humans. Since humans are bipedal, their body structure places their head directly above the rest of their body and results in a longer vertical distance for gravity to distribute fluid, whereas the quadrupedal body of mice places their head on a much closer geometric plane to the rest of their body. Due to these differences in body structures, the fluid-shift that occurs in mice during spaceflight may be less dramatic than the fluid-shift observed in humans. All of these differences must be taken into consideration when making predictions about human changes in retinal structure based off of murine-specific studies.

Speculation on the Stem Cell Epitranscriptome During Spaceflight

One curiosity of spaceflight is the impact it has on stem cells. Previous experiments have demonstrated the behavior of stem cells is altered by the microgravity environment aboard the ISS. Specifically, embryonic stem cells have displayed a resistance to differentiation. One study flew mouse embryonic stem cells aboard the ISS for 15 days to perform an embryoid body differentiation protocol in microgravity. Cells did not complete embryoid body differentiation in microgravity and analysis of gene expression found differential expression in stem cell related pathways, including the genes *Neurog2*, *Sox1*, and *Sox2* compared to the ground control. After the cells were returned to Earth, they displayed greater stemness and more easily differentiated into cardiomyocyte colonies (Blaber et al., 2015).

Another study cultured induced pluripotent stem cells (iPSCs) from mice aboard the TianZhou-1 spacecraft to further characterize pluripotency during spaceflight. The iPSCs exposed to spaceflight produced larger colony sizes than their ground control counterparts over the course of a 10 day experiment. The spaceflight exposed iPSCs also displayed stronger expression of *Oct4* when colonies became overgrown, as measured by a GFP reporter assay.

However, the ground control cells reported reduced expression of *Oct4* over the same timespan (Zhou et al., 2019).

To date, the connection between the epitranscriptome and pluripotent stem cells exposed to spaceflight remains unexplored. From previous studies, we know that the m^6A modification also impacts pluripotent state (Wen et al., 2018; Wu et al., 2019; Aguilo et al., 2015; Batista et al., 2014; Geula et al., 2015). Given the connection between the m^6A modification, transcript turnover, and stem cell differentiation, the m^6A landscape might differ under the environmental changes of spaceflight and be a possible contributor to the differences in stemness and differentiation potential of stem cells grown in spaceflight.

Direct-RNA Nanopore Sequencing: A New Standard Aboard ISS?

With recent advances in genomic technology, it is now possible to perform nanopore sequencing and the necessary library preparation aboard the ISS. To sequence RNA with the nanopore MinION, there are two options: cDNA sequencing and direct-RNA sequencing. Unlike direct-RNA sequencing, cDNA sequencing cannot be used to detect RNA modifications. This is because modifications do not get duplicated during PCR amplification.

There are certain advantages to cDNA sequencing. First, less RNA sample is required. This is because PCR, the part of the process that is erasing RNA modifications, will create the necessary mass of molecules needed for sequencing. As of today, 1 ng of RNA is required for cDNA sequencing. On the other hand, since it is performed without PCR, the direct-RNA sequencing protocol requires 500 ng of RNA. Second, cDNA sequencing will produce more reads. According to Oxford Nanopore Technologies, the anticipated number of reads from a one flow cell of a cDNA experiment is 7-12 million full-length reads, whereas direct-RNA sequencing will produce 1 million full-length reads per flow cell.

However, if an experiment can be accomplished with the specifications of direct-RNA sequencing, it offers clear advantages over cDNA sequencing. First, there are so few opportunities to collect spaceflight-omics data. Each experiment performed should be seeking to gain the maximum amount of information possible. Direct-RNA sequencing will add epitranscriptomic information to the differential gene expression and isoform information obtained by cDNA sequencing. Second, this additional data can be collected without any additional library preparation steps compared to cDNA experiments. In fact, without PCR, the library preparation time is less than for cDNA sequencing. This has the advantage of saving

astronaut-time, which is very limited and tightly managed by coordinators at Johnson Space Center, and could free time for astronauts to work on additional experiments.

There is still much progress to be made in order to fully utilize and understand the epitranscriptome information that is collected by nanopore sequencing. In Chapter 5, I demonstrated the ability of direct-RNA nanopore sequencing to identify differences in the global epitranscriptome and the m^6A modification landscape between different pluripotent states. Future developments in algorithms will make it easier to identify specific modifications and detect a variety of modified bases with greater accuracy. Even though we do not yet have the algorithm and analysis methods perfected, we should not waste the opportunity to begin data collection. It would be more advantageous to have the data available for when analysis methods are perfected than to be required to regenerate it from scratch, especially since experiments are rarely duplicated on the ISS.

Precise connections between spaceflight, gene expression, and the epitranscriptome remain unknown. Additionally, experiments profiling DNA methylation, chromatin state, and enhancer interactions will continue to elucidate the regulatory changes that are happening during spaceflight. We are still at the beginning of understanding the differences in gene expression during spaceflight and their connection to the epitranscriptome. Slightly more studied than the link between gene expression during spaceflight and the epitranscriptome is the impact of epitranscriptomic modifications between the natural variation of cell types. However, further improvements in modification detection with direct-RNA sequencing will accelerate our ability to profile these modifications.

References

Aguilo, Francesca, Fan Zhang, Ana Sancho, Miguel Fidalgo, Serena Di Cecilia, Ajay Vashisht, Dung-Fang Lee, et al. 2015. “Coordination of m6A mRNA Methylation and Gene Transcription by ZFP217 Regulates Pluripotency and Reprogramming.” *Cell Stem Cell* 17 (6): 689–704.

Albert, Paul R., and Chawki Benkelfat. 2013. “The Neurobiology of Depression—revisiting the

Serotonin Hypothesis. II. Genetic, Epigenetic and Clinical Studies†." *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences* 368 (1615): 20120535.

Batista, Pedro J., Benoit Molinie, Jinkai Wang, Kun Qu, Jiajing Zhang, Lingjie Li, Donna M. Bouley, et al. 2014. "m6A RNA Modification Controls Cell Fate Transition in Mammalian Embryonic Stem Cells." *Cell Stem Cell* 15 (6): 707–19.

Blaber, Elizabeth A., Hayley Finkelstein, Natalya Dvorochkin, Kevin Y. Sato, Rukhsana Yousuf, Brendan P. Burns, Ruth K. Globus, and Eduardo A. C. Almeida. 2015. "Microgravity Reduces the Differentiation and Regenerative Potential of Embryonic Stem Cells." *Stem Cells and Development* 24 (22): 2605–21.

Geula, Shay, Sharon Moshitch-Moshkovitz, Dan Dominissini, Abed Alfatah Mansour, Nitzan Kol, Mali Salmon-Divon, Vera Hershkovitz, et al. 2015. "Stem Cells. m6A mRNA Methylation Facilitates Resolution of Naïve Pluripotency toward Differentiation." *Science* 347 (6225): 1002–6.

Hodge, Rebecca D., Trygve E. Bakken, Jeremy A. Miller, Kimberly A. Smith, Eliza R. Barkan, Lucas T. Graybuck, Jennie L. Close, et al. 2019. "Conserved Cell Types with Divergent Features in Human versus Mouse Cortex." *Nature* 573 (7772): 61–68.

Lowe, Derek. 2019. "Human Brains and Mouse Brains: So Similar, So Different." August 22, 2019. https://blogs.sciencemag.org/pipeline/archives/2019/08/22/human-brains-and-mouse-brains-so-similar-so-different.

Moriguchi, Sho, Akihiro Takamiya, Yoshihiro Noda, Nobuyuki Horita, Masataka Wada, Sakiko Tsugawa, Eric Plitman, et al. 2019. "Glutamatergic Neurometabolite Levels in Major Depressive Disorder: A Systematic Review and Meta-Analysis of Proton Magnetic Resonance Spectroscopy Studies." *Molecular Psychiatry*. https://doi.org/10.1038/s41380-018-0252-9.

Palinkas, L. A. 2001. "Psychosocial Issues in Long-Term Space Flight: Overview." *Gravitational and Space Biology Bulletin: Publication of the American Society for Gravitational and Space Biology* 14 (2): 25–33.

Volland, Stefanie, Julian Esteve-Rudd, Juyea Hoo, Claudine Yee, and David S. Williams. 2015. "A Comparison of Some Organizational Characteristics of the Mouse Central Retina and the Human Macula." *PloS One* 10 (4): e0125631.

Wen, Jing, Ruitu Lv, Honghui Ma, Hongjie Shen, Chenxi He, Jiahua Wang, Fangfang Jiao, et al. 2018. "Zc3h13 Regulates Nuclear RNA m6A Methylation and Mouse Embryonic Stem Cell Self-Renewal." *Molecular Cell* 69 (6): 1028–38.e6.

Wu, Ruifan, Youhua Liu, Yuanling Zhao, Zhen Bi, Yongxi Yao, Qing Liu, Fengqin Wang, Yizhen Wang, and Xinxia Wang. 2019. "m6A Methylation Controls Pluripotency of Porcine Induced Pluripotent Stem Cells by Targeting SOCS3/JAK2/STAT3 Pathway in a YTHDF1/YTHDF2-Orchestrated Manner." *Cell Death & Disease*. https://doi.org/10.1038/s41419-019-1417-4.

Zhou, Jin, Xiao-Hui Dong, Feng-Zhi Zhang, Hui-Min Zhu, Tong Hao, Xiao-Xia Jiang, Wei-Bo Zheng, et al. 2019. "Real Microgravity Condition Promoted Regeneration Capacity of Induced Pluripotent Stem Cells during the TZ-1 Space Mission." *Cell Proliferation* 52 (3): e12574.

Appendix A: Chapter 4 Supplemental Tables and Scripts

Supplemental Table 1: GeneLab RNA-Seq Datasets

Organism	Method	Identifier	Study Name	Link
Mus Musculus	Spaceflight			
		GLDS-47	Rodent Research-1 (RR1) National Lab Validation Flight: Mouse liver transcriptomic, proteomic, and epigenomic data	https://genelab-data.ndc.nasa.gov/genelab/accession/GLDS-47/
		GLDS-48	Rodent Research-1 (RR1) NASA Validation Flight: Mouse liver transcriptomic, proteomic, and epigenomic data	https://genelab-data.ndc.nasa.gov/genelab/accession/GLDS-48/
		GLDS-98	Rodent Research-1 (RR1) NASA Validation Flight: Mouse adrenal gland transcriptomic, proteomic, and epigenomic data	https://genelab-data.ndc.nasa.gov/genelab/accession/GLDS-98/
		GLDS-99	Rodent Research-1 (RR1) NASA Validation Flight: Mouse extensor digitorum longus muscle transcriptomic and epigenomic data	https://genelab-data.ndc.nasa.gov/genelab/accession/GLDS-99/
		GLDS-100	Rodent Research-1 (RR1) NASA Validation Flight: Mouse eye transcriptomic and epigenomic data	https://genelab-data.ndc.nasa.gov/genelab/accession/GLDS-100/
		GLDS-101	Rodent Research-1 (RR1) NASA Validation Flight: Mouse gastrocnemius muscle transcriptomic, proteomic, and epigenomic data	https://genelab-data.ndc.nasa.gov/genelab/accession/GLDS-101/
		GLDS-103	Rodent Research-1 (RR1) NASA Validation Flight: Mouse quadriceps muscle transcriptomic, proteomic, and epigenomic data	https://genelab-data.ndc.nasa.gov/genelab/accession/GLDS-103/
		GLDS-104	Rodent Research-1 (RR1) NASA Validation Flight: Mouse soleus muscle transcriptomic and epigenomic data	https://genelab-data.ndc.nasa.gov/genelab/accession/GLDS-104/
		GLDS-105	Rodent Research-1 (RR1) NASA Validation Flight: Mouse tibialis anterior muscle transcriptomic, proteomic, and epigenomic data	https://genelab-data.ndc.nasa.gov/genelab/accession/GLDS-105/
		GLDS-137	Rodent Research-3-CASIS: Mouse liver transcriptomic, proteomic, and epigenomic data	https://genelab-data.ndc.nasa.gov/genelab/accession/GLDS-137/

		GLDS-161	Rodent Research-3-CASIS: Mouse adrenal gland transcriptomic, proteomic, and epigenomic data	https://genelab-data.ndc.nasa.gov/genelab/accession/GLDS-161/
		GLDS-162	Rodent Research-3-CASIS: Mouse eye transcriptomic and proteomic data	https://genelab-data.ndc.nasa.gov/genelab/accession/GLDS-162/
		GLDS-163	Rodent Research-3-CASIS: Mouse kidney transcriptomic, proteomic, and epigenomic data	https://genelab-data.ndc.nasa.gov/genelab/accession/GLDS-163/
		GLDS-164	Effects of spaceflight on the immunoglobulin repertoire of unimmunized C57BL/6 mice	https://genelab-data.ndc.nasa.gov/genelab/accession/GLDS-164/
		GLDS-168	RR-1 and RR-3 mouse liver transcriptomics with and without ERCC control RNA spike-ins	https://genelab-data.ndc.nasa.gov/genelab/accession/GLDS-168/
		GLDS-173	STS-135: Mouse Liver Transcriptomics using RNA-Seq	https://genelab-data.ndc.nasa.gov/genelab/accession/GLDS-173/
		GLDS-194	Rodent Research-3-CASIS: Mouse retina transcriptomic data	https://genelab-data.ndc.nasa.gov/genelab/accession/GLDS-194/
		GLDS-238	Transcriptomic analysis of dorsal skin from mice flown on the MHU-2 mission	https://genelab-data.ndc.nasa.gov/genelab/accession/GLDS-238/
		GLDS-239	Transcriptomic analysis of femoral skin from mice flown on the MHU-2 mission	https://genelab-data.ndc.nasa.gov/genelab/accession/GLDS-239/
		GLDS-240	Transcriptional analysis of dorsal skin from mice flown on the RR-5 mission	https://genelab-data.ndc.nasa.gov/genelab/accession/GLDS-240/
		GLDS-241	Transcriptional analysis of femoral skin from mice flown on the RR-5 mission	https://genelab-data.ndc.nasa.gov/genelab/accession/GLDS-241/
		GLDS-242	Effect of spaceflight on liver from mice flown on the ISS for 33 days: transcriptional analysis	https://genelab-data.ndc.nasa.gov/genelab/accession/GLDS-242/
		GLDS-243	Transcriptional analysis of dorsal skin from mice flown on the RR-6 mission	https://genelab-data.ndc.nasa.gov/genelab/accession/GLDS-243/
		GLDS-244	Transcriptional analysis of thymus from mice flown on the RR-6 Mission	https://genelab-data.ndc.nasa.gov/genelab/accession/GLDS-244/
		GLDS-245	Transcriptional analysis of liver from mice flown on the RR-6 mission	https://genelab-data.ndc.nasa.gov/genelab/accession/GLDS-245/
		GLDS-246	Transcriptional analysis of spleen from mice flown on the RR-6 mission	https://genelab-data.ndc.nasa.gov/genelab/accession/GLDS-246/
		GLDS-247	Transcriptional analysis of colon from mice flown on the RR-6 mission	https://genelab-data.ndc.nasa.gov/genelab/accession/GLDS-247/

		GLDS-248	Transcriptional analysis of lung from mice flown on the RR-6 mission	https://genelab-data.ndc.nasa.gov/genelab/accession/GLDS-248/
		GLDS-253	Transcriptional analysis of kidneys from mice flown on the RR-7 mission	https://genelab-data.ndc.nasa.gov/genelab/accession/GLDS-253/
		GLDS-254	Transcriptional analysis of dorsal skin from mice flown on the RR-7 mission	https://genelab-data.ndc.nasa.gov/genelab/accession/GLDS-254/
		GLDS-255	Spaceflight influences gene expression, photoreceptor integrity, and oxidative stress-related damage in the murine retina	https://genelab-data.ndc.nasa.gov/genelab/accession/GLDS-255/
		GLDS-270	Spatially resolved transcriptional analysis of hearts from mice flown on the RR-3 mission	https://genelab-data.ndc.nasa.gov/genelab/accession/GLDS-270/
		GLDS-288	Transcriptome analysis of murine spleen in space	https://genelab-data.ndc.nasa.gov/genelab/accession/GLDS-288/
		GLDS-289	Impact of spaceflight on gene expression in the thymus	https://genelab-data.ndc.nasa.gov/genelab/accession/GLDS-289/
	Ground Study			
		GLDS-49	Multi-omic investigations of mouse liver subjected to simulated spaceflight freezing and storage protocols	https://genelab-data.ndc.nasa.gov/genelab/accession/GLDS-49/
		GLDS-141	Validation of Methods to Assess the Immunoglobulin Gene Repertoire in Tissues Obtained from Mice on the International Space Station	https://genelab-data.ndc.nasa.gov/genelab/accession/GLDS-141/
		GLDS-201	Impact of Antiorthostatic Suspension on Mouse response to Tetanus Toxoid and CpG	https://genelab-data.ndc.nasa.gov/genelab/accession/GLDS-201/
		GLDS-202	Low dose (0.04 Gy) irradiation (LDR) and hindlimb unloading (HLU) microgravity in mice: brain transcriptomic and epigenomic data	https://genelab-data.ndc.nasa.gov/genelab/accession/GLDS-202/
		GLDS-203	Low dose (0.04 Gy) irradiation (LDR) and hindlimb unloading (HLU) microgravity in mice: retina transcriptomic and epigenomic data	https://genelab-data.ndc.nasa.gov/genelab/accession/GLDS-203/
		GLDS-211	Transcriptomic analysis of spleens from mice subjected to chronic low-dose radiation, hindlimb unloading or a combination of both	https://genelab-data.ndc.nasa.gov/genelab/accession/GLDS-211/
		GLDS-214	Impact of Antiorthostatic Suspension on Mouse response to Tetanus Toxoid	https://genelab-data.ndc.nasa.gov/genelab/accession/GLDS-214/

			and CpG: bone marrow transcriptomic data	
		GLDS-235	Transcriptomic analysis of liver from mice subjected to simulated spaceflight euthanasia, freezing, and tissue preservation protocols	https://genelab-data.ndc.nasa.gov/genelab/accession/GLDS-235/
		GLDS-236	Transcriptomic analysis of quadriceps from mice subjected to simulated spaceflight euthanasia, freezing, and tissue preservation protocols	https://genelab-data.ndc.nasa.gov/genelab/accession/GLDS-236/
		GLDS-237	Transcriptomic analysis of skin from mice subjected to chronic low-dose radiation, hindlimb unloading or a combination of both	https://genelab-data.ndc.nasa.gov/genelab/accession/GLDS-237/
		GLDS-272	Transcriptional analysis of spleens from mice preserved with the Rapid Freeze hardware	https://genelab-data.ndc.nasa.gov/genelab/accession/GLDS-272/
		GLDS-273	Transcriptional analysis of livers from mice preserved with the Rapid Freeze hardware	https://genelab-data.ndc.nasa.gov/genelab/accession/GLDS-273/
		GLDS-274	Transcriptional analysis of soleus from mice preserved with the Rapid Freeze hardware	https://genelab-data.ndc.nasa.gov/genelab/accession/GLDS-274/
		GLDS-287	Murine TCR-beta repertoire sequencing	https://genelab-data.ndc.nasa.gov/genelab/accession/GLDS-287/
		GLDS-294	Efficient Identification of Multiple Pathways: RNA-Seq Analysis of Livers from 56Fe Ion Irradiated Mice	https://genelab-data.ndc.nasa.gov/genelab/accession/GLDS-294/
		GLDS-295	RNA-Seq transcriptome profiling of soleus from adult and aged mice following disuse muscle atrophy and recovery	https://genelab-data.ndc.nasa.gov/genelab/accession/GLDS-295/
		GLDS-308	Differential expression profiles of long non-coding RNAs during the mouse pronucleus stage under normal gravity and simulated microgravity	https://genelab-data.ndc.nasa.gov/genelab/accession/GLDS-308/
		GLDS-318	RNA seq of tumors derived from irradiated versus sham hosts transplanted with Trp53 null mammary tissue and fed either Control diet versus Caffeic Acid Phenethyl Ester (CAPE) diet	https://genelab-data.ndc.nasa.gov/genelab/accession/GLDS-318/
		GLDS-319	Muscle atrophy, osteoporosis prevention in hibernating mammals	https://genelab-data.ndc.nasa.gov/genelab/accession/GLDS-319/

		GLDS-322	Comparative RNA-Seq transcriptome analyses reveal dynamic time dependent effects of 56Fe, 16O, and 28Si irradiation on the induction of murine hepatocellular carcinoma	https://genelab-data.ndc.nasa.gov/genelab/accession/GLDS-322/
Homo sapiens	Spaceflight			
		GLDS-258	Effects of Spaceflight on Human Induced Pluripotent Stem Cell-Derived Cardiomyocyte Structure and Function	https://genelab-data.ndc.nasa.gov/genelab/accession/GLDS-258/
	Ground Study			
		GLDS-91	A study of gene expression influenced by simulated microgravity in human lymphoblastoid cells	https://genelab-data.ndc.nasa.gov/genelab/accession/GLDS-91/
		GLDS-127	Global gene expression profiles of cardiac progenitors differentiated from human pluripotent stem cells in 3D culture under simulated microgravity	https://genelab-data.ndc.nasa.gov/genelab/accession/GLDS-127/
Drosophila melanogaster	Spaceflight			
		GLDS-96	The development of Drosophila melanogaster during space flight	https://genelab-data.ndc.nasa.gov/genelab/accession/GLDS-96/
		GLDS-207	Correlated Gene and Protein Expression in heads from Drosophila reared in microgravity	https://genelab-data.ndc.nasa.gov/genelab/accession/GLDS-207/
		GLDS-278	Spaceflight and simulated microgravity conditions increase virulence of Serratia marcescens in the Drosophila melanogaster infection model	https://genelab-data.ndc.nasa.gov/genelab/accession/GLDS-278/
	Ground Study			
		GLDS-85	Transcriptomic response of Drosophila melanogaster pupae developed in hypergravity	https://genelab-data.ndc.nasa.gov/genelab/accession/GLDS-85/
Arabidopsis thaliana	Spaceflight			
		GLDS-37	Comparison of the spaceflight transcriptome of four commonly used Arabidopsis thaliana ecotypes	https://genelab-data.ndc.nasa.gov/genelab/accession/GLDS-37/
		GLDS-38	Proteomics and Transcriptomics analysis of Arabidopsis Seedlings in Microgravity	https://genelab-data.ndc.nasa.gov/genelab/accession/GLDS-38/

		GLDS-120	Genetic Dissection of the Spaceflight Transcriptome Responses in Plants: are some responses unnecessary?	https://genelab-data.ndc.nasa.gov/genelab/accession/GLDS-120/
		GLDS-218	Characterization of Epigenetic Regulation in an Extraterrestrial Environment: The Arabidopsis Spaceflight Methylome [RNA-seq]	https://genelab-data.ndc.nasa.gov/genelab/accession/GLDS-218/
		GLDS-223	The effect of spaceflight on transgenic Arabidopsis plants with compromised signaling	https://genelab-data.ndc.nasa.gov/genelab/accession/GLDS-223/
		GLDS-251	RNAseq analysis of the response of Arabidopsis thaliana to fractional gravity under blue-light stimulation during spaceflight	https://genelab-data.ndc.nasa.gov/genelab/accession/GLDS-251/
	Ground Study			
		GLDS-208	Comparative gene expression analysis in the Arabidopsis thaliana root apex using RNA-seq and microarray transcriptome profiles	https://genelab-data.ndc.nasa.gov/genelab/accession/GLDS-208/
Oryzias latipes	Spaceflight			
		GLDS-83	Comparative Transcriptomic Analysis of Adult Medaka Tissues Sampled after Adaptation to a Space Environment	https://genelab-data.ndc.nasa.gov/genelab/accession/GLDS-83/
		GLDS-133	Medaka Osteoclast	https://genelab-data.ndc.nasa.gov/genelab/accession/GLDS-133/
Helix lucorum	Spaceflight			
		GLDS-192	Adaptive changes in the vestibular system of land snail to a 30-day spaceflight and readaptation on return to Earth	https://genelab-data.ndc.nasa.gov/genelab/accession/GLDS-192/
Brassica rapa	Spaceflight			
		GLDS-59	RNA-Seq transcriptome analysis of reactive oxygen species gene network in Mizuna plants grown in long-term space flight	https://genelab-data.ndc.nasa.gov/genelab/accession/GLDS-59/
Eruca vesicaria	Spaceflight			
		GLDS-301	Eruca sativa Rocket Science RNA-seq	https://genelab-data.ndc.nasa.gov/genelab/accession/GLDS-301/

Euprymna scolopes	Ground Study			
		GLDS-119	Effect of microgravity on an animal-bacteria symbiosis	https://genelab-data.ndc.nasa.gov/genelab/accession/GLDS-119/
Ceratopteris richardii	Ground Study			
		GLDS-57	RNA-Seq analysis identifies potential modulators of gravity response in Ceratopteris spores: Evidence for modulation by calcium pumps and apyrase activity	https://genelab-data.ndc.nasa.gov/genelab/accession/GLDS-57/
Bacillus subtilis	Spaceflight			
		GLDS-138	BRIC-23 GeneLab Process Verification Test: Bacillus subtilis transcriptomic, proteomic, and metabolomic data	https://genelab-data.ndc.nasa.gov/genelab/accession/GLDS-138/
		GLDS-185	BRIC-21 Bacillus subtilis transcriptome profile data	https://genelab-data.ndc.nasa.gov/genelab/accession/GLDS-185/
Escherichia coli	Spaceflight			
		GLDS-95	A Molecular Genetic Basis Explaining Altered Bacterial Behavior in Space	https://genelab-data.ndc.nasa.gov/genelab/accession/GLDS-95/

Supplemental Table 2: Reference Genome and Annotation Files

GeneLab RNAseq Reference Genome and Annotation Files	
Organism	**Genome / Annotation Files Used**
Arabidopsis thaliana	**Ensembl Plants release 44**
	Genome file:
	ftp://ftp.ensemblgenomes.org/pub/plants/release-44/fasta/arabidopsis_thaliana/dna/Arabidopsis_thaliana.TAIR10.dna.toplevel.fa.gz
	Annotation file:
	ftp://ftp.ensemblgenomes.org/pub/plants/release-44/gtf/arabidopsis_thaliana/Arabidopsis_thaliana.TAIR10.44.gtf.gz
Bacillus subtilis	**Ensembl Bacteria release 46**
	Genome file:
	ftp://ftp.ensemblgenomes.org/pub/bacteria/release-46/fasta/bacteria_0_collection/bacillus_subtilis_subsp_subtilis_str_168/dna/Bacillus_subtilis_subsp_subtilis_str_168.ASM904v1.dna.toplevel.fa.gz
	Annotation file:
	ftp://ftp.ensemblgenomes.org/pub/bacteria/release-46/gtf/bacteria_0_collection/bacillus_subtilis_subsp_subtilis_str_168/Bacillus_subtilis_subsp_subtilis_str_168.ASM904v1.46.gtf.gz
Brassica rapa	**Ensembl Plants release 43**
	Genome file:
	ftp://ftp.ensemblgenomes.org/pub/plants/release-43/fasta/brassica_rapa/dna/Brassica_rapa.Brapa_1.0.dna.toplevel.fa.gz
	Annotation file:
	ftp://ftp.ensemblgenomes.org/pub/plants/release-43/gtf/brassica_rapa/Brassica_rapa.Brapa_1.0.43.gtf.gz
Drosophila melanogaster	**Ensembl release 98**
	Genome file:
	ftp://ftp.ensembl.org/pub/release-98/fasta/drosophila_melanogaster/dna/Drosophila_melanogaster.BDGP6.22.dna.toplevel.fa.gz
	Annotation file:
	ftp://ftp.ensembl.org/pub/release-98/gtf/drosophila_melanogaster/Drosophila_melanogaster.BDGP6.22.98.gtf.gz
ERCC	**Thermo Fisher**
	Genome file:
	https://assets.thermofisher.com/TFS-Assets/LSG/manuals/ERCC92.zip
	(Note it's the ERCC92.fa file in the zipped directory)
	Annotation file:

	https://assets.thermofisher.com/TFS-Assets/LSG/manuals/ERCC92.zip
	(Note it's the ERCC92.gtf file in the zipped directory)
Escherichia coli	**Ensembl Bacteria release 47**
	Genome file:
	ftp://ftp.ensemblgenomes.org/pub/bacteria/release-47/fasta/bacteria_159_collection/escherichia_coli_gca_001284565/dna/Escherichia_coli_gca_001284565.7790_1_86.dna.toplevel.fa.gz
	Annotation file:
	ftp://ftp.ensemblgenomes.org/pub/bacteria/release-47/gtf/bacteria_159_collection/escherichia_coli_gca_001284565/Escherichia_coli_gca_001284565.7790_1_86.47.gtf.gz
Homo sapiens	**Ensembl release 98**
	Genome file:
	ftp://ftp.ensembl.org/pub/release-98/fasta/homo_sapiens/dna/Homo_sapiens.GRCh38.dna.primary_assembly.fa.gz
	Annotation file:
	ftp://ftp.ensembl.org/pub/release-98/gtf/homo_sapiens/Homo_sapiens.GRCh38.98.gtf.gz
Mus musculus	**Ensembl release 96**
	Genome file:
	ftp://ftp.ensembl.org/pub/release-96/fasta/mus_musculus/dna/Mus_musculus.GRCm38.dna.toplevel.fa.gz
	Annotation file:
	ftp://ftp.ensembl.org/pub/release-96/gtf/mus_musculus/Mus_musculus.GRCm38.96.gtf.gz
Mycobacterium marinum	**Ensembl Bacteria release 47**
	Genome file:
	ftp://ftp.ensemblgenomes.org/pub/bacteria/release-47/fasta/bacteria_8_collection/mycobacterium_marinum_m/dna/Mycobacterium_marinum_m.ASM1834v1.dna.toplevel.fa.gz
	Annotation file:
	ftp://ftp.ensemblgenomes.org/pub/bacteria/release-47/gtf/bacteria_8_collection/mycobacterium_marinum_m/Mycobacterium_marinum_m.ASM1834v1.47.gtf.gz
Oryzias latipes	**Ensembl release 99**
	Genome file:
	ftp://ftp.ensembl.org/pub/release-99/fasta/oryzias_latipes/dna/Oryzias_latipes.ASM223467v1.dna.toplevel.fa.gz
	Annotation file:
	ftp://ftp.ensembl.org/pub/release-99/gtf/oryzias_latipes/Oryzias_latipes.ASM223467v1.99.gtf.gz

Supplemental Table 3: Organism to Annotation Mapping

name	species	taxon	annotations
HUMAN	Homo sapiens	9606	org.Hs.eg.db
MOUSE	Mus musculus	10090	org.Mm.eg.db
RAT	Rattus norvegicus	10116	org.Rn.eg.db
ZEBRAFISH	Danio rerio	7955	org.Dr.eg.db
FLY	Drosophila melanogaster	7227	org.Dm.eg.db
WORM	Caenorhabditis elegans	6239	org.Ce.eg.db
YEAST	Saccharomyces cerevisiae (strain ATCC 204508 / S288c)	559292	org.Sc.sgd.db
ARABIDOPSIS	Arabidopsis thaliana	3702	org.At.tair.db
ECOLI	Escherichia coli (strain K12)	83333	org.EcK12.eg.db
BACSU	Bacillus subtilis (strain 168)	224308	*org.Bsubtilis.eg.db
*Since there is no publicly available annotation package that contains ensebml IDs, the `makeOrgPackageFromNCBI` function of the AnnotationForge Bioconductor Package was used to create an annotation package for Bacillus subtilis. The exact script used to create this package along with the package created for GeneLab datasets involving Bacillus subtilis are provided in the GitHub repository (https://github.com/nasa/GeneLab_Data_Processing/tree/master/RNAseq/GLDS_Processing_Scripts) for each respective dataset. See the GLDS-138 processing scripts for an example (https://github.com/nasa/GeneLab_Data_Processing/tree/master/RNAseq/GLDS_Processing_Scripts/GLDS-138).			

Supplemental Table 4: Pipeline Tools and Links

Program	Relevant Links
FastQC	https://www.bioinformatics.babraham.ac.uk/projects/fastqc/
MultiQC	https://multiqc.info/
Cutadapt	https://cutadapt.readthedocs.io/en/stable/
TrimGalore	https://www.bioinformatics.babraham.ac.uk/projects/trim_galore/
STAR	https://github.com/alexdobin/STAR
RSEM	https://github.com/deweylab/RSEM
Bioconductor	https://bioconductor.org
DESeq2	https://bioconductor.org/packages/release/bioc/html/DESeq2.html
tximport	https://bioconductor.org/packages/release/bioc/html/tximport.html
tidyverse	https://www.tidyverse.org
Risa	https://www.bioconductor.org/packages/release/bioc/html/Risa.html
STRINGdb	https://www.bioconductor.org/packages/release/bioc/html/STRINGdb.html
PANTHER.db	https://bioconductor.org/packages/release/data/annotation/html/PANTHER.db.html
org.Hs.eg.db	https://bioconductor.org/packages/release/data/annotation/html/org.Hs.eg.db.html
org.Mm.eg.db	https://bioconductor.org/packages/release/data/annotation/html/org.Mm.eg.db.html
org.Dm.eg.db	https://bioconductor.org/packages/release/data/annotation/html/org.Dm.eg.db.html
org.Ce.eg.db	https://bioconductor.org/packages/release/data/annotation/html/org.Ce.eg.db.html
org.At.tair.db	https://bioconductor.org/packages/release/data/annotation/html/org.At.tair.db.html
org.EcK12.eg.db	https://bioconductor.org/packages/release/data/annotation/html/org.EcK12.eg.db.html
org.Sc.sgd.db	https://bioconductor.org/packages/release/data/annotation/html/org.Sc.sgd.db.html

Supplemental Table 5: Sample Names

GLDS-168
Mmus_C57-6J_LVR_RR1_FLT_wERCC_Rep1_M25
Mmus_C57-6J_LVR_RR1_FLT_wERCC_Rep2_M26
Mmus_C57-6J_LVR_RR1_FLT_wERCC_Rep3_M28
Mmus_C57-6J_LVR_RR1_FLT_wERCC_Rep4_M29
Mmus_C57-6J_LVR_RR1_FLT_wERCC_Rep5_M30
Mmus_C57-6J_LVR_RR1_GC_wERCC_Rep1_M36
Mmus_C57-6J_LVR_RR1_GC_wERCC_Rep2_M37
Mmus_C57-6J_LVR_RR1_GC_wERCC_Rep3_M38
Mmus_C57-6J_LVR_RR1_GC_wERCC_Rep4_M39
Mmus_C57-6J_LVR_RR1_GC_wERCC_Rep5_M40
GLDS-245
Mmus_C57-6T_LVR_FLT_ISS-T_Rep1_F5
Mmus_C57-6T_LVR_FLT_ISS-T_Rep2_F8
Mmus_C57-6T_LVR_FLT_ISS-T_Rep3_F9
Mmus_C57-6T_LVR_FLT_ISS-T_Rep4_F10
Mmus_C57-6T_LVR_FLT_ISS-T_Rep5_F7
Mmus_C57-6T_LVR_FLT_ISS-T_Rep6_F1
Mmus_C57-6T_LVR_FLT_ISS-T_Rep7_F2
Mmus_C57-6T_LVR_FLT_ISS-T_Rep8_F3
Mmus_C57-6T_LVR_FLT_ISS-T_Rep9_F4
Mmus_C57-6T_LVR_FLT_ISS-T_Rep10_F6
Mmus_C57-6T_LVR_GC_ISS-T_Rep1_G4
Mmus_C57-6T_LVR_GC_ISS-T_Rep2_G9
Mmus_C57-6T_LVR_GC_ISS-T_Rep3_G10
Mmus_C57-6T_LVR_GC_ISS-T_Rep4_G7
Mmus_C57-6T_LVR_GC_ISS-T_Rep5_G8
Mmus_C57-6T_LVR_GC_ISS-T_Rep6_G1
Mmus_C57-6T_LVR_GC_ISS-T_Rep7_G2
Mmus_C57-6T_LVR_GC_ISS-T_Rep8_G3
Mmus_C57-6T_LVR_GC_ISS-T_Rep9_G5
Mmus_C57-6T_LVR_GC_ISS-T_Rep10_G6