Gravitational and Space Research

Assessing the effect of spaceflight stress on DNA sequence mutation using RNA-sequencing data

Montana S. Knight[1], Colleen J. Doherty[2], Dahlia M. Nielsen[3,4]*

[1]*Bioinformatics Graduate Program, North Carolina State University, Raleigh, NC, USA;* [2]*Department of Molecular and Structural Biochemistry, North Carolina State University, Raleigh, NC, USA;* [3]*Department of Biological Sciences, North Carolina State University, Raleigh, NC, USA;* [4]*Bioinformatics Research Center, North Carolina State University, Raleigh, NC, USA.*

Abstract

Understanding the effects of space radiation and microgravity on DNA is critical to assessing the impact of long-term spaceflight. While experiments performed in space constitute the most effective means of examining these effects, opportunities are limited and costly. As this bottleneck will likely continue for the foreseeable future, data from past experiments represent a particularly valuable source of information for continuing studies. To this end, NASA created GeneLab, a public Omics database for spaceflight-related data. We used data from GeneLab to examine the effect of spaceflight on DNA mutation rates. Optimally, mutation rates are estimated using DNA sequence data directly. Unfortunately, to date, few DNA-based datasets appear in GeneLab. Transcript data, however, is abundant. Here we used RNA-Seq data to examine DNA sequence variation in Arabidopsis thaliana *seedlings grown aboard the International Space Station (ISS) vs. on the ground. ISS-based samples were grown under two conditions: spaceflight under microgravity, and, by using a specialized onboard centrifuge under induced gravity. This powerful experimental design allowed us to separate the effects of microgravity from non-microgravity spaceflight stress, such as space radiation. More mutations were observed in spaceflight samples than in ground control samples, with transversion mutations being overly represented. Mutation rates identified in samples grown under artificial gravity in space were similar to that of microgravity spaceflight samples, indicating that microgravity exposure played a limited role. This work demonstrates that RNA-Seq data is useful for evaluating DNA damage from spaceflight and provides insight into the types of mutations that occur.*

Introduction

Understanding how spaceflight impacts terrestrial life is essential in ensuring astronauts' health and safety and optimizing future space travel (Choi et al., 2019; Durante, 2014; Kennedy, 2014). Two major environmental stressors during spaceflight, microgravity and space radiation, are known to impact terrestrial life. Space radiation originates from two major sources relevant to biological systems: galactic cosmic rays (GCR) and solar energetic particles (SEPs) generated by solar activity (Bourdarie and Xapsos, 2008; Cucinotta, 2014; Hellweg and Baumstark-Khan, 2007). GCRs consist primarily of high-energy protons, alpha particles (helium nuclei), and heavier charged nuclei (e.g., C, O, Fe) (Bourdarie and Xapsos, 2008; Cucinotta, 2014). Solar flares and coronal mass ejections create bursts of high-energy radiation that include X-rays, gamma rays, and energetic charged particles. These particles, primarily protons, electrons, alpha particles, and heavier nuclei, travel at high velocities. Their speed, combined with their energy and, for heavier nuclei, their high charge, makes them

strongly ionizing when they interact with matter. Although there is a general 11-year cycle of solar activity, there is significant variability within this cycle, and radiation levels can increase dramatically within minutes in response to a single solar event. The sporadic nature of these ionizing radiation events makes it challenging to study their impact on biological systems in low Earth orbit (Bourdarie and Xapsos, 2008).

Although plants can survive and reproduce in environments with chronic ionizing radiation that would be lethal to most animal systems, they often show stronger mutational responses at comparable doses. In a meta-analysis of radiation effects across taxa, Møller and Mousseau (2013) reported that mean effect sizes for plants were larger than those for animals (plants = 0.749 vs. animals = 0.093). Although this difference was not statistically significant, likely due to the limited sample size, subsequent analyses and experimental studies support the idea that plants display elevated mutation rates under chronic irradiation despite their increased survival. For example, work in Chernobyl and Fukushima has shown that plants exposed to chronic ionizing

[†]Corresponding author: Dahlia M. Nielsen

E-mail: dahlia@statgen.ncsu.edu

radiation exhibit elevated mutation rates, reduced reproductive success, and morphological abnormalities (Mousseau and Møller, 2020). Additionally, wheat grown for multiple generations in Chernobyl-contaminated soils exhibited up to a sixfold increase in mutation rates compared to predictions based on a single acute radiation dose (Kovalchuk et al., 2000). The ability to survive as an organism despite their high genetic sensitivity makes plants a particularly powerful model system for studying the effects of space radiation. Plants endure exposures that would be lethal to animals, while still manifesting radiation-induced genetic changes.

However, opportunities to study ionizing radiation on plants are limited because only a few terrestrial environments contain elevated ionizing radiation levels. The best way to understand the effects of microgravity and space radiation on terrestrial life is through experiments in spaceflight. Such space-based experiments have demonstrated that heavy ion radiation impacts DNA. For example, chromosomal aberrations were observed in HeLa cells exposed for 40 days on the Russian MIR space station (Moreno-Villanueva et al., 2017). Damage was also observed in cells exposed for as little as nine days on the Space Shuttle (Moreno-Villanueva et al., 2017). Long-term space radiation exposure increases astronauts' risk of cancer, cataracts, and neurological disorders (Brojakowska et al., 2022; Cucinotta et al., 2001). In plants, spaceflight has been linked to changes in alternative splicing patterns, epigenetic patterns, and germination rates (Chandler et al., 2020; Xu et al., 2018; Ou et al., 2010; Beisel et al., 2019). Microgravity has been linked to muscle atrophy in astronauts, disruptions to DNA repair enzymes, and impacts on cellular proliferation (Moreno-Villanueva et al., 2017; Yuge et al., 2006; Zhang, 2001; Chandler et al., 2020).

While experiments performed in space are the gold standard for observing the effects of spaceflight on terrestrial life, opportunities are limited and expensive (Rinaldi, 2016; Vandenbrink and Kiss, 2016). An alternative is to simulate the spaceflight conditions on Earth (La Tessa et al., 2016; Zhang et al., 2022; Wuest et al., 2015). Microgravity can be simulated with drop towers, 3D clinostats, and random positioning machines (Zhang et al., 2022). Space radiation can be evaluated at facilities like the NASA Space Radiation Laboratory in Upton, NY (La Tessa et al., 2016). Radiation damage can also be evaluated with high-altitude balloon flights. These simulation procedures effectively approximate the impact of spaceflight (Moreno-Villanueva et al., 2017; Wuest et al., 2015). For example, DNA structural variants have been observed in *Arabidopsis thaliana* seeds exposed to space radiation on a high-altitude balloon over Antarctica, demonstrating that such balloons are a practical way to study radiation without entering low earth orbit (Califar et al., 2018). Other types of radiation, such as ultraviolet, are easily studied on Earth. Solar radiation experienced on Earth (which

contains multiple types of UV radiation) has been linked to increased transversion mutations (Kunz and Armstrong, 1998). UVA radiation was shown to significantly increase transversion mutations in mice (Besaratinia et al., 2004). Further, DNA damage from reactive oxygen radicals has been associated with increased G/C transversion mutations (Kunz and Armstrong, 1998). These approximations for both space radiation and microgravity conditions are useful; however, it is important to validate results in actual spaceflight.

The prohibitive nature of space flight experiments provides added benefit to secondary analysis of data from prior experiments (Ray et al., 2019). To this end, NASA created a public Omics data repository, GeneLab. Here, we used GeneLab data to examine the effect of space flight on DNA mutation rate. While this repository is heavily biased toward transcriptome datasets, most of which were intended for gene expression analysis, data from RNA-Seq experiments can also be analyzed to gain insights about DNA. We demonstrate that identifying and examining variants in RNA-seq data from *Arabidopsis thaliana* samples grown aboard the ISS provides a proxy for estimating DNA damage from spaceflight.

Experiments performed on board the ISS provide a means of directly studying the effects of exposure to microgravity and high levels of space radiation, however disentangling these effects can be difficult. With an onboard centrifuge, gravity on the ISS can be reintroduced, and the effects of space radiation can be separated from the effects of microgravity and space radiation combined. The experiment chosen for study here leveraged this equipment. It also included a matched set of samples grown on Earth as a ground-based control. By using this dataset, we were able to compare the number of single nucleotide variants identified in ISS-grown *Arabidopsis thaliana* samples and ground control samples. As expected, spaceflight samples resulted in a higher number of variants. Additionally, we examined variant composition in each sample, finding that spaceflight was associated with more transversion mutations. This aligns with previous studies examining the impact of radiation on the nucleotide level (Kunz and Armstrong, 1998; Besaratinia et al., 2004). In addition, this study demonstrates that RNA-Seq data can be used effectively as a proxy to evaluate changes in DNA.

Materials and Methods

Datasets

NASA's GeneLab repository was examined for species and assay type composition on August 29, 2022. As of this date, GeneLab held data from 385 studies. For each of these, we recorded the organism examined, the spaceflight status (e.g., spaceflight, simulated spaceflight, ground control), and the type of assay performed (e.g., RNA-Seq).

Table 1. Number of samples for each treatment combination.

Genotype	Tissue Type	Space Flight μG	Space Flight 0.76G by centrifugation	Ground Control 1G on Earth	Total
Wild Type	Root	4	4	2	10
Wild Type	Shoot	4	3	3	10
InsP 5-ptase Transgenic	Root	4	4	2	10
InsP 5-ptase Transgenic	Shoot	4	3	3	10

To perform the current analysis, we selected GeneLab accession GLDS-223 (Perera et al., 2019). In this study (Sheppard et al., 2021), *Arabidopsis thaliana* seeds were transported to the ISS, where they were germinated in growth chambers and allowed to grow for five days before being placed in a -80°C freezer for storage and transport back to Earth. Two genotypes of Arabidopsis were used for this experiment, wild-type Columbia-0 (Col-0) and a transgenic genotype derived from Col-0 with mammalian type I inositol polyphosphate 5-phosphatase (InsP 5-ptase) (Perera et al., 2006) integrated. The Col-0 genotype is highly inbred, so homozygosity is high and plants are genetically homogeneous.

Seedlings in the ISS were exposed to two gravity conditions, the first of which was the standard microgravity experienced by the station. The second set of plants was grown under simulated gravity (0.76G) by being placed in centrifuge rotors within the growth chambers. Seeds were launched to the ISS on July 8, 2011, and the samples were returned to Earth on March 26, 2013. Another set of seedlings was grown on the Earth's surface under standard 1G gravity, with other chamber settings matching the ISS growth chamber. These constitute the ground control samples. After spaceflight samples were returned to the surface, root and shoot tissues were collected, and RNA was extracted for sequencing. Each RNA-Seq sample consisted of root or shoot tissues from 27 seedlings. There were between two and four biological replicates for each condition (Table 1). For full details on the samples and experimental conditions, see (Perera et al., 2019; Sheppard et al., 2021; Land et al., 2023).

Sequence read processing and alignment
Raw fastq files downloaded from GeneLab were analyzed using fastqc (Andrews, 2010). As a first step in quality control, Fastx_trimmer was used to trim the first thirteen and last five nucleotides of each read in each sequence file (Hannon, 2010); this aids in removing adaptor sequences and low quality nucleotide calls. Fastq_quality_filter was then used to filter out base calls with a phred-quality score lower than 30 (Hannon, 2010). Fastqc was again used to examine the quality of the newly filtered fastq files before aligning the reads to the Arabidopsis genome. Trimmed and filtered fastq reads were mapped to the TAIR10 genome using STAR in 2-pass

mode (Cheng et al., 2017; Dobin et al., 2013; Dobin and Gingeras, 2016). Alignment files were evaluated with Picard's "CollectAlignmentSummaryMetrics." This served as a quality control check of the alignments (Broad Institute, 2016). SplitNCigarReads was then applied to aligned reads. This separates reads that encompass a splice site into the sequence before and after the splice site, and cuts nucleotides mapped to the intronic region (van der Auwera and O'Connor, 2020). Next, base quality score recalibration was performed using BaseRecalibrator and ApplyBQSR (van der Auwera and O'Connor, 2020).

Generating a list of known population variants
As the samples used in this experiment are highly homogeneous, germline variants (inherited sequence variants differing from the reference genome sequence) are expected to be common to all seedlings. Germline variants were identified and used to create the Ultimate Panel of Normals (PoN), constituting the "known list of population variants" for BaseRecalibrator and the "PoN" for Mutect2 (Benjamin et al., 2019).

As RNA-Seq data were used here, sequencing coverage was not uniform across the genome but was instead derived from expressed regions. As gene expression patterns differ across tissue types, and the ability to identify variants depends on sequence coverage, germline variants identified in one tissue might not be found in the other tissues. Therefore, four lists of germline variants were created, one for each combination of tissue type (root and shoot) and genotype (WT and transgenic).

To identify germline variants, we combined results from two approaches. First, HaplotypeCaller (Poplin et al., 2017) was used with a heterozygosity parameter setting of 0.0001 to account for the highly inbred nature of these Arabidopsis genotypes. Next, Mutect2 (Benjamin et al., 2019) was run in "tumor-only" mode, with a max number of haplotypes of two (also to accommodate the low level of heterozygosity). The VCFs produced by Mutect2 were combined using GenomicsDBImport and a PoN was created using CreateSomaticPanelofNormals (van der Auwera and O'Connor, 2020).

The two sets of joint variant calls from HaplotypeCaller (Poplin et al., 2017) and Mutect2 (Benjamin et al., 2019) were then

combined. Variants identified by either tool that were present in at least three of the ten biological replicates per tissue type/genotype combination were considered germline variants. This final list of germline variants, the Ultimate PoN, was used as the "known list of population variants" for BaseRecalibrator and the "PoN" for Mutect2 (Benjamin et al., 2019).

Identification of de novo mutations

To identify *de novo* mutations, Mutect2 in "tumor-only" mode was applied, applying the same parameter settings used to create the PoN. Variants identified were filtered using GATK's VariantFiltration (van der Auwera and O'Connor, 2020), and any variants contained within the Ultimate PoN were also filtered out. Sites with read depths lower than 10 were filtered out. Variants that clustered in groups of three or more within a span of 35 nucleotides were also filtered out. Following this, a deduplication procedure was applied. Deduplication is generally performed after reads are aligned to the genome and before variant calling, to minimize PCR duplication error (Ebbert et al., 2016; Sayols et al., 2016). However, this can result in variants not being called by Mutect2 or being filtered out due to low read depth or support, as deduplication risks eliminating biological duplicate reads without examining them for potential variants (Sayols, 2016). Instead, we performed deduplication after variant calling. Variants were evaluated by the number of unique reads supporting them. If a variant was only found in identical reads aligned to the same location with no mismatches, then it is considered a PCR duplicate and is filtered out. This variant-centric deduplication method retains reads with variants even if they would otherwise be marked as duplicates.

Variants were also removed if they were in regions with increased risk of sequencing error, in lowly expressed genes (less than 10 pseudo-counts), or not expressed in all the samples. To ensure expression differences among treatment groups did not influence results, the script only keeps variants from genes expressed in all samples. Regions with an increased risk of sequencing errors were defined as a sequence of length three or more composed of the same nucleotide. Variants called in these repeat sequences (e.g., AA"A") were removed. This is a conservative approach to identifying high-confidence variants. The final output from the variant calling and filtering steps is a variant composition file for each sample.

Variant calling composition

A script was created to analyze each sample's variant call file (VCF) and obtain information on the composition of the mutations. The script outputs the number of:

- Single nucleotide polymorphisms (SNPs)
- Indels (both insertions and deletions)
- Times the alternative SNP was an A, C, G, or T

- Transversion mutations
- Transition mutations
- Times the SNP was an A to C, G, or T variant
- Times the SNP was a C to A, G, or T variant
- Times the SNP was a G to A, C, or T variant
- Times the SNP was a T to A, C, or G variant.

Statistical analysis

Final variant counts were analyzed in R with a generalized linear model. Spaceflight conditions, tissue type, and Arabidopsis genotype were tested for their contribution to the results. Least square estimates were obtained for all treatment effects, including the main effects of potential interactions. The variant composition was also examined using a generalized linear model with a binomial distribution. Proportions for each type of variant were examined, and treatment effects were tested for their influence on the results.

Gene Expression Analysis

To examine the effect of gene expression patterns on variant counts, transcript abundance counts were calculated with HTSeq-count (Anders et al., 2015) using the alignment file from STAR 2-pass and the TAIR10 annotation file (Cheng et al., 2017; Anders et al., 2015). Normalized values were derived using EdgeR (Robinson et al., 2010) and analyzed via Principal Components Analysis in R (R Core Team, n.d.).

Data Availability

This research was conducted using public data from NASA's Omics database, GeneLab (accession number GLDS-223 (Perera et al., 2019)[25]). Scripts used for the analysis pipeline can be found at https://github.com/montana-knight/Calling-Induced-Variants-with-RNA-Seq-Data.

Results

GeneLab contains a high proportion of gene expression data

The purpose of this study was to evaluate if spaceflight-induced mutations could be detected using publicly available transcriptome data downloaded from GeneLab, a repository set up by NASA for data from spaceflight-related experiments. As of August 29, 2022, GeneLab held data from 385 studies. The data were derived from experiments performed under actual and simulated spaceflight conditions, with 54.81% of the datasets containing an actual spaceflight component. We examined the composition of the different data types to assess the types of analysis necessary to achieve our objective.

GeneLab data include many species. *Mus musculus* made up 32% of all GeneLab datasets and 36% of all spaceflight

Table 2. Number of variants identified per genotype/tissue type combination.

	Number of Likely Shared Variants from Mutect2	Number of Joint Genotyped Variants from HaplotypeCaller	Number of Variants in the Ultimate PON
Transgenic Shoot	10933	13291	7469
Transgenic Root	6835	9505	6954
Wild Type Shoot	11408	13902	8021
Wild Type Root	6818	9510	6943

datasets. Plants were another large section of the database, with 14% of all available datasets and 13% of spaceflight-specific datasets focused on plants. Of this plant data, 92% of the data in all GeneLab sets and 89% of data in spaceflight-only sets were from *Arabidopsis thaliana*.

The composition of GeneLab's assay types was also examined. 77% of studies had a transcription profiling component, demonstrating a bias toward gene expression experiments. RNA-Seq data made up 35% of all GeneLab data and 42% of the spaceflight data. There was a total of 133 datasets from RNA-Seq experiments with a variety of species, including 76 sets from *M. musculus*, 23 sets from Arabidopsis, and seven from humans. Eighty-eight datasets were spaceflight-specific RNA-Seq datasets; 13 were from Arabidopsis, 53 were from *M. musculus*, and five were from humans. Less than a third of the spaceflight studies had a DNA component, and over half of the spaceflight experiments with DNA sequence information were from single-cell organisms. No public spaceflight dataset on GeneLab from humans or plants with DNA sequence information is available to assess DNA damage. However, the abundance of RNA-Sequencing data provides an opportunity to analyze transcripts for variants and gain insight into how spaceflight impacts terrestrial life at the nucleotide level.

Germline and de novo variants identified

GLDS-223 (Perera et al., 2019) was selected for analysis. In this study, two sets of Arabidopsis seedlings were grown on the ISS, one in microgravity and another under simulated gravity conditions. A separate set of ground control samples was also grown. Two genotypes of Arabidopsis were used: wild-type Col-0 plants and a transgenic genotype derived from Col-0 (see Materials and Methods for details). Col-0 is a highly inbred line, and thus, plants display high levels of homozygosity and homogeneity. Root and shoot tissues were assayed separately.

To identify mutations caused by spaceflight conditions (*de novo* mutations), sites that deviated from the reference genome but were common to plants in both the spaceflight and ground control conditions (germline variants) were first identified. These variants constituted the four Ultimate PoNs (see Methods for details) and were removed from consideration as *de novo* mutations. The final Ultimate PoNs had an average of 7346.75 variants (Table 2).

Once germline variants were removed, the remaining variants were filtered (see Methods). An average of 31,942.08 variants were initially called by Mutect2. The hard filtering and deduplication steps filtered out the most variants. Hard filters removed 18,349 variants on average, and deduplication filtered an average of 11,899.75 variants. Filtering variants from difficult-to-sequence reads and genes with low read coverage had less impact on final variant counts, with an average of 605.98 and 15.93 variants filtered out, respectively (Supplemental Table S1). All variants passing these filters were considered to be *de novo* mutations. Variant counts for each sample (Supplemental Table S2) were used for further analysis.

Evaluation of the identified variants
Spaceflight influenced the number of variants called in Arabidopsis

To determine if RNA-Seq data from plants grown in space revealed increased mutation levels compared to ground control, we examined differences in the number of mutations identified between spaceflight and ground control samples. Spaceflight samples were expected to be associated with more mutations than ground control samples because of the increased radiation exposure. This hypothesis was based on the expectation that space radiation affects the mutation rate, as previous research has shown that radiation exposure results in DNA damage (Moreno-Villanueva et al., 2017; Kunz and Armstrong, 1998; Besaratinia et al., 2004). To investigate this, we compared samples from the three conditions: ground control, spaceflight in gravity control, and normal spaceflight conditions. For simplicity, we refer to the stress affecting samples grown in the gravity control in space as space radiation. However, we acknowledge that other non-microgravity spaceflight stress conditions could also impact this group (Moreno-Villanueva et al., 2017; Beheshti et al., 2018; Thirsk et al., 2009).

Final variant counts were analyzed via a generalized linear model to examine the impact of each spaceflight condition, tissue type (root and shoot), genotype (WT and transgenic), and any possible interactions (Supplemental Tables S3, S4). Interaction effects were investigated first to see if relationships between tissue type, genotype, or spaceflight condition (space radiation and gravity factor) impacted the number of *de novo* mutations identified. Each two-, three-, and four-way

Table 3. ANOVA results, testing for differences in numbers of *de novo* mutations.

Comparison	Difference Estimate	SE	P-value
Shoot vs. root tissue	990.306	71.523	1.34E-43
Interaction effect b/w tissue and gravity condition	-1708.917	572.959	0.003
Interaction b/w tissue type and non-microgravity spaceflight stress	3327.833	651.432	3.25E-07
Interaction b/w gravity and non-microgravity spaceflight stress	-12248.417	753.981	2.42E-59
Interaction b/w tissue type, gravity condition, and non-microgravity spaceflight stress	-4322.917	753.981	9.84E-09
Gravity vs. microgravity	148.302	71.62	0.038
Spaceflight vs. ground control	225.354	81.429	0.006
Spaceflight with gravity control vs. ground control	201.604	92.235	0.029
Spaceflight with gravity control vs. ground control in shoot tissue	613.833	126.545	1.23E-06
Spaceflight vs. ground control in shoot tissue	641.333	107.284	2.26E-09

interaction was tested. Four interaction effects significantly affected the number of *de novo* mutations (Table 3). The first was a three-way interaction between tissue type, gravity, and radiation. The other significant interactions were two-way interactions nested between these three conditions.

The genotype of Arabidopsis did not impact the number of variants identified (estimate: 19.39, p-value = 0.7863) (Fig. 1, Supplemental Table S4). This was not surprising, as the two genotypes of Arabidopsis used in this study were wild-type Col-0 and a transgenic genotype derived from Col-0 overexpressing the mammalian type I InsP 5-ptase. Since InsP 5-ptase is not an Arabidopsis gene (Perera et al., 2006), it did not map to the Arabidopsis reference genome during the read alignment step. Therefore, the downstream analysis did not consider any reads corresponding to that gene. Additionally, the expression of this gene does not appear to have affected the variant calling procedure for other genes (this potentially could occur if the expression of the transgene affects the expression of native Arabidopsis genes, as transcript abundance has a bearing on variant calling).

Tissue type emerged as a significant factor in the analysis (Fig. 1). Without taking individual spaceflight conditions into account, root tissue samples had far fewer *de novo* mutations compared to shoot tissue samples (estimate: 990.3 variants, p-value = 1.34e-43). The two tissue types were analyzed separately due to the three-way significant interaction effect found between tissue type, gravity, and non-microgravity spaceflight stress. There were no significant differences between the three spaceflight/ground conditions for root tissue (root samples in spaceflight with simulated gravity vs. root samples in normal spaceflight conditions, estimate: 39.99, p-value=0.72; root samples in spaceflight with simulated gravity vs. root samples grown on Earth, estimate: 210.63, p-value = 0.12) (Fig. 1, Supplemental Table S4).

The number of *de novo* mutations found in the shoot tissue, on the other hand, was significantly different between the
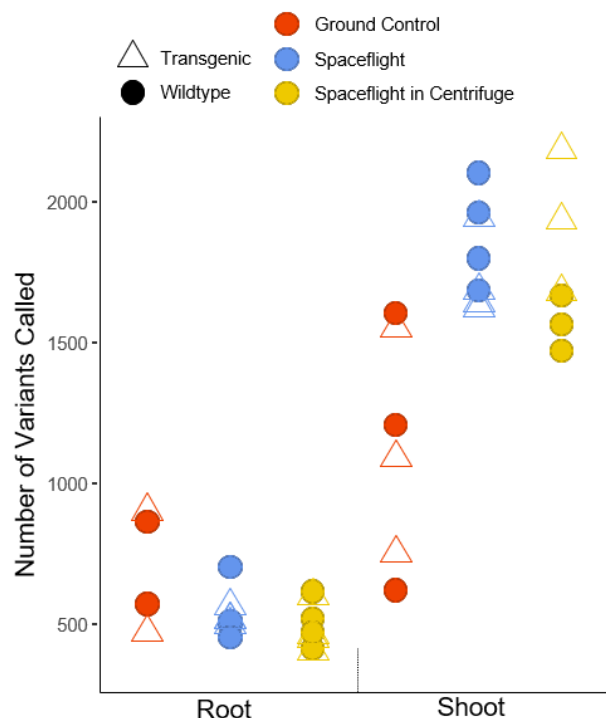
Figure 1. Number of *de novo* variants identified. Each point represents one RNA-Seq sample.

ground and spaceflight conditions (Fig. 1). Overall, shoot tissue samples grown aboard the ISS had a higher number of mutations compared to the ground control samples (estimate: 641.33, p-value = 2.26e-09). Shoot tissue samples grown in simulated gravity aboard the ISS had more mutations than the ground control samples (estimate: 613.83, p-value = 1.23e-06). This comparison points to space radiation, rather than microgravity, as the factor that induces more DNA damage. The impact of microgravity was investigated by comparing the two groups of samples grown on the ISS. These ISS-grown

Table 4. ANOVA results, testing for the proportion of transversion mutations.

Comparison	Difference Estimate	SE	P-value
Shoot vs. root tissue	0.0624	0.007	2.65E-19
Interaction effect b/w tissue and gravity condition	-0.1324	0.0558	0.018
Interaction b/w tissue type and non-microgravity spaceflight stress	0.1461	0.0634	0.021
Interaction b/w gravity and non-microgravity spaceflight stress	-3.8392	0.0735	<0.001
Interaction b/w tissue type, gravity condition, and non-microgravity spaceflight stress	-0.3608	0.0735	9.03E-07
Spaceflight vs. ground control	0.0163	0.0079	0.04
Spaceflight with gravity control vs. ground control	0.0177	0.0089	0.048
Spaceflight with gravity control vs. ground control in shoot tissue	0.031	0.0078	7.81E-05
Spaceflight vs. ground control in shoot tissue	0.0345	0.0069	4.72E-07

samples were all exposed to space radiation but were grown in either microgravity or simulated gravity conditions. The impact of microgravity was not significant in the shoot tissue samples grown aboard the ISS (estimate: 55.00, p-value = 0.642). This indicates that microgravity did not significantly contribute to the accumulation of variants observed, supporting the hypothesis that space radiation was the primary driver of mutations.

Spaceflight influenced the types of variants identified in Arabidopsis seedlings

Solar radiation has been linked to a higher number of transversion mutations (Kunz and Armstrong, 1998; Besaratinia et al., 2004), motivating us to examine whether a higher number of transversion mutations could be observed in the spaceflight samples selected for this study. Variant calls for each sample were analyzed to examine the types of mutations identified. The ratio of transversion mutations to the total number of SNPs was analyzed using a generalized linear model with a binomial distribution, with the ratio as a proportion. The model's predictor variables were spaceflight conditions, Arabidopsis genotype, and tissue type (Fig. 2). Once again, the Arabidopsis genotype did not significantly impact the proportion of transversion mutations (estimate: 0.008, p-value: 0.212). Interactions between tissue type, gravity, and space radiation level and all nested two-way interactions were significant (Table 4). Thus, all three conditions were incorporated into the model to investigate the impact of spaceflight.

Spaceflight shoot tissue samples had a higher proportion of transversion mutations compared to the ground control shoot tissue samples (estimate: 0.035, p-value = 4.72e-07) (Table 4, Fig. 2). This difference remains significant when samples experiencing simulated gravity in spaceflight are compared to ground control samples, indicating that the effect is not due to microgravity (estimate: 0.031, p-value 7.81e-05) (Table 4, Fig. 2) and that instead, as seen in previous studies (Kunz and Armstrong, 1998; Besaratinia et al., 2004), space radiation is
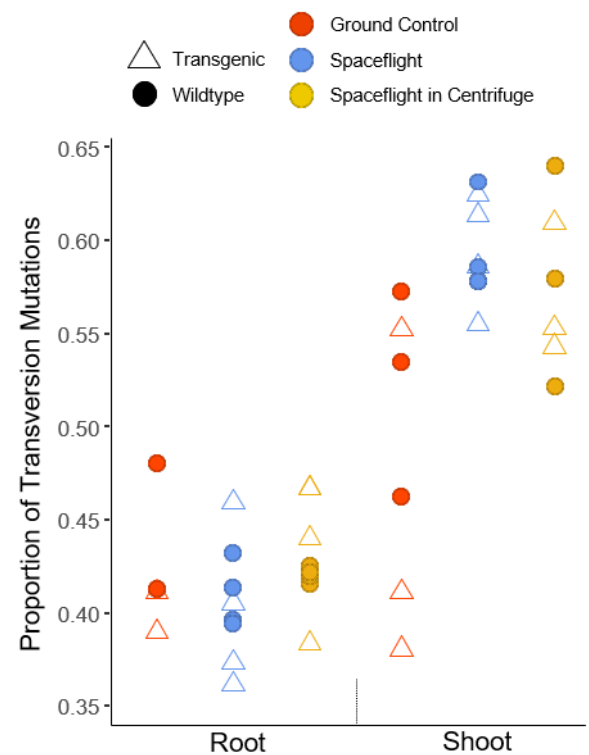
Figure.2. Proportion of transversion mutations. For each sample, the proportion of *de novo* variants identified that are transversions is shown. Each point represents one RNA-Seq sample.

associated with a higher proportion of transversion mutations. Root tissue samples did not show significant differences in the proportion of transversion mutations among either spaceflight condition, consistent with the reduced number of all single nucleotide variants identified in root tissue.

We also examined the types of transversion point mutations enriched in spaceflight. G→C/C→G mutations and G→T/C→A mutations were significantly enriched in the spaceflight shoot samples grown in simulated gravity (p<0.001). The

other two types of transversion mutations, A→C/T→G and A→T/T→A, were not enriched in the spaceflight samples.

Discussion

Space travel presents new challenges to terrestrial life. Experiments performed on board the ISS provide a means of studying these effects through both direct experimentation and secondary analysis of data collected from these experiments and submitted to NASA's GeneLab online public repository. In the latter case, the analysis options are constrained by the experimental designs of the original experiments, including the original datatypes collected. A secondary issue is that it is difficult to separate the effects of combined forces, such as microgravity and space radiation. Here, we investigate the effects of spaceflight on DNA mutation rates using data extracted from GeneLab. The majority of genomic-centered datasets contained in this repository are focused on gene expression. We examined whether expression data generated using RNA-Seq technology could be used as an indirect proxy for identifying DNA sequence changes. We leveraged a dataset from an experiment in which Arabidopsis seedlings were grown on the ISS both under microgravity and under simulated gravity, and on Earth under normal gravity. This allowed us to examine the effects of microgravity and space radiation on mutation rates separately.

Our results indicate that for shoot tissue, though not for root tissue, spaceflight stress caused a higher number of *de novo* mutations in spaceflight samples than found in ground control samples (estimate: 614, p-value = 1.23e-06). We did not observe a difference between spaceflight samples grown in microgravity vs. spaceflight samples grown in artificial gravity in either tissue type (root estimate: 40, p-value = 0.72; shoot estimate: 55, p-value =0.64), indicating that microgravity did not contribute significantly to sequence changes. This result was inconsistent with previous reports that microgravity affects relevant pathways for mitigating damage, including effects on DNA repair enzymes and impacts on cellular proliferation observed in astronauts (Moreno-Villanueva et al., 2017; Yuge et al., 2006; Zhang, 2001; Chandler et al., 2020). One possibility for this inconsistency is the amount of time the samples for the experiment we selected spent in space. If the effect microgravity has on DNA occurs at a slower rate, changes may not have been detectable within the timeframe of this experiment.

The results for shoot tissue differed from those for root tissue, which was unexpected. This difference may be due to technical factors related to the experimental design and/ or the assays. For instance, transcript abundance patterns across the genome varied greatly between root and shoot tissues (Supplemental Figure S1). Shoot tissue samples had more highly expressed genes compared to the root tissue samples, which may have affected the relative ability to detect variants in these tissues. This highlights a challenge of calling variants from RNA-seq data: The process depends on transcript abundance, which depends on the sample type and the experimental conditions. Another technical factor affecting the analyses of these data is the sample size, as there were only two root tissue samples in the ground control set for both Arabidopsis genotypes (Table 1). Small sample sizes such as this directly affect the power to detect differences between groups. Determining if the tissue-specific differences we observe are due to actual biological differences will likely require larger sample sizes and/or a direct analysis of DNA sequence data.

The types of mutations detected were also examined to investigate the effect of spaceflight at the nucleotide level. The results showed transversion mutations were more common in spaceflight than ground control samples, with space radiation believed to be the contributing treatment. Specifically, we found that the proportion of G→T/C→A transversions was significantly higher in shoot tissue samples in the simulated gravity that were exposed to space radiation vs. the ground control samples that were not exposed to space radiation. Additionally, the proportion of C→G/G→C transversions increased when examining the effect of space radiation on shoot tissue samples grown in simulated gravity. Transversion mutations have previously been shown to be more common than transition mutations in the presence of radiation (Kunz and Armstrong, 1998; Besaratinia et al., 2004). However, a higher number of transition mutations were found when examining blood taken from astronauts who had been aboard the ISS for a median of 12 days (Brojakowska et al., 2022; Cucinotta et al., 2001). In this study, DNA sequence data for 37 genes were examined. The contrary results from this study and ours could be due to the smaller number of genes they examined, differences between human and Arabidopsis sensitivity to DNA damage and repair, or differences in comparing mutations identified from RNA vs. directly from DNA. Despite these differences, the results of both studies illustrate that more research is needed to understand DNA damage caused by spaceflight conditions. This includes examining factors such as duration of exposure and whether the effects are different for different tissues or species.

While the experimental design of this study offered an excellent opportunity for examining the effect of spaceflight while controlling for the effect of microgravity, there were limitations. One issue with examining RNA-Seq data rather than DNA sequence data is that the identified variants could result from spaceflight-specific RNA-editing events and not DNA damage. This is not something that can be tested without collecting both DNA and RNA samples. Additionally, each RNA-Seq sample was derived from a pool of 27 plants in this

study. Using pooled samples affects the ability to examine the effects of spaceflight on individual plants. For example, an interesting avenue of exploration would be to test whether mutations in particular genes (such as DNA repair genes) within an individual were associated with higher mutation rates. This type of test is not feasible for this dataset. Data based on pooled samples also affected the sensitivity for identifying *de novo* mutations, as it is necessary to identify variants that exist within a single plant in a mixture of 27 plants. This issue is compounded for mutations that occurred later in development, as the number of cells within the plant that carry the mutation is smaller. The variant identification tool used for this analysis, Mutect2 (Benjamin et al., 2019), has been benchmarked as having high sensitivity and specificity in calling variants at low frequencies (Maruzani et al., 2024). However, many *de novo* mutations were likely missed in this analysis. Assaying individual samples rather than pools of samples would increase the ability to identify mutations and provide opportunities for more in-depth testing of their effects. Worth noting is that while plants were only grown for five days on the ISS before plant tissue was frozen, the seeds (containing the plant embryos) experienced longer exposures to space flight conditions. Mutations that occur earlier in development are easier to identify, as the number of daughter cells derived from the cell that experienced the original mutation is larger.

There will always be limitations when using data derived from an experiment designed for purposes other than the questions asked in a secondary analysis. However, the knowledge gained from the secondary analyses can still be substantial. An experimental design in which DNA was collected on separate samples (rather than pooled samples) would provide more in-depth results regarding space-flight mutation rates; however, these studies do not currently exist, and are not likely to be performed in the foreseeable future. The experimental design for the study used here did incorporate several ideal components for this investigation, including matched ground control samples, and space flight samples in both microgravity and induced gravity. Thus, in spite of the inherent limitations of the data, our results were able to demonstrate the use of past experiments as a rich resource, overcoming the bottleneck that limited access to spaceflight creates.

Acknowledgments

Author Disclosure Statement

The authors declare no competing financial or non-financial interests.

Author Contributions

C.J.D. came up with the original idea. All three authors contributed to developing the analysis strategy. M.S.K. developed a pipeline and analyzed data with the support of C.J.D. and D.M.N. C.J.D. and D.M.N. provided continuous feedback and encouragement during the study. M.S.K. and D.M.N. were responsible for the majority of the writing.

References

Anders S, Paul Theodor P, Wolfgang H (2015) HTSeq--a Python framework to work with high-throughput sequencing data. *Bioinformatics* **31**: 166-69.

Andrews S (2010) FastQC: A quality control tool for high throughput sequence data. *Bioinformatics*. http://www.bioinformatics.babraham.ac.uk/projects/fastqc/.

Auwera G, O'Connor BD (2020) *Genomics in the Cloud*, Sebastopol, CA: O'Reilly Media.

Beheshti A, Cekanaviciute E, Smith DJ, Costes SV (2018) Global transcriptomic analysis suggests carbon dioxide as an environmental stressor in spaceflight: A systems biology GeneLab case study. *Scientific Reports* **8**: 4191.

Beisel NS, Noble J, Barbazuk WB, Anna-Lisa P, Ferl RJ (2019) Spaceflight-induced alternative splicing during seedling development in *Arabidopsis Thaliana*. *NPJ Microgravity* **5**: 9.

Benjamin D, Sato T, Cibulskis K, Getz G, Stewart C, Lichtenstein L (2019) Calling somatic SNVs and indels with Mutect2. *bioRxiv*. https://doi.org/10.1101/861054.

Besaratinia A, Synold TW, Xi B, Pfeifer GP (2004) G-to-T transversions and small tandem base deletions are the hallmark of mutations induced by ultraviolet a radiation in mammalian cells. *Biochemistry* **43**: 8169-77.

Bourdarie S, Xapsos M (2008) The near-Earth space radiation environment. *IEEE Transactions on Nuclear Science* **55**: 1810-32.

Broad Institute (2016) *Picard*, Broad Institute. https://github.com/broadinstitute/picard.

Brojakowska A, Kour A, Thel MC, Park E, Bisserier M, Garikipati VNS, Hadri L, Mills PJ, Walsh K, Goukassian DA (2022) Retrospective

analysis of somatic mutations and clonal hematopoiesis in astronauts. *Communications Biology* **5**: 828.

Califar B, Tucker R, Cromie J, Sng N, Schmitz RA, Callaham JA, Barbazuk B, Paul AL, Ferl RJ (2018) Approaches for surveying cosmic radiation damage in large populations of *Arabidopsis Thaliana* Seeds – antarctic balloons and particle beams. *Gravitational and Space Research: Publication of the American Society for Gravitational and Space Research* **6**: 54-73.

Chandler JO, Haas FB, Khan S, Bowden L, Ignatz M, Enfissi EMA, Gawthrop F, Griffiths A, Fraser PD, Rensing SA, Leubner-Metzger G (2020) Rocket science: The effect of spaceflight on germination physiology, ageing, and transcriptome of Eruca Sativa Seeds. *Life (Basel, Switzerland)* **10**: 49.

Cheng CY, Krishnakumar V, Chan AP, Thibaud-Nissen F, Schobel S, Town CD (2017) Araport11: A complete reannotation of the *Arabidopsis Thaliana* reference genome. *The Plant Journal: For Cell and Molecular Biology* **89**: 789-804.

Choi WG, Barker RJ, Kim SH, Swanson SJ, Gilroy S (2019) Variation in the transcriptome of different ecotypes of *Arabidopsis Thaliana* reveals signatures of oxidative stress in plant responses to spaceflight. *American Journal of Botany* **106**: 123-36.

Cucinotta FA, Schimmerling W, Wilson JW, Peterson LE, Badhwar GD, Saganti PB, Dicello JF (2001) Space radiation cancer risks and uncertainties for Mars missions. *Radiation Research* **156**: 682-88.

Cucinotta FA (2014) Space radiation risks for astronauts on multiple International Space Station missions. *PloS One* **9**: e96099.

Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, Batut P, Chaisson M, Gingeras TR (2013) STAR: Ultrafast universal RNA-Seq aligner. *Bioinformatics (Oxford, England)* **29**: 15-21.

Dobin A, Gingeras TR (2016) Optimizing RNA-Seq mapping with STAR. *Methods in Molecular Biology* **1415**: 245-62.

Durante M (2014) Space radiation protection: Destination Mars. *Life Sciences in Space Research* **1**: 2-9.

Ebbert MTW, Wadsworth ME, Staley LA, Hoyt KL, Pickett B, Miller J, Duce J (2016) Alzheimer's disease neuroimaging initiative. In *Evaluating the Necessity of PCR Duplicate Removal from next-Generation Sequencing Data and a Comparison of* Approaches, J.S.K. Kauwe, and P.G. Ridge (ed), p. 238. BMC Bioinformatics.

Hannon GJ (2010) *FASTX-Toolkit*, http://hannonlab.cshl.edu/fastx_toolkit.

Hellweg CE, Baumstark-Khan C (2007) Getting ready for the manned mission to Mars: The astronauts' risk from space radiation. *The Science of Nature* **94**: 517-26.

Kennedy AR (2014) Biological effects of space radiation and development of effective countermeasures. *Life Sciences in Space Research* **1**: 10-43.

Kovalchuk O, Dubrova YE, Arkhipov A, Hohn B, Kovalchuk I (2000) Wheat mutation rate after Chernobyl. *Nature* **407**: 583-84.

Kunz BA, Armstrong JD (1998) Differences in the mutational specificities of sunlight and UVB radiation suggest a role for transversion-inducing DNA damage in solar photocarcinogenesis. *Mutation Research* **422**: 77-83.

Land ES, Sheppard J, Doherty CJ, Perera IY (2023) Conserved plant transcriptional responses to microgravity from two consecutive spaceflight experiments. *Frontiers in Plant Science* **14**: 1308713.

La Tessa C, Sivertz M, Chiang IH, Lowenstein D, Rusek A (2016) Overview of the NASA space radiation laboratory. *Life Sciences in Space Research* **11**: 18-23.

Maruzani R, Brierley L, Jorgensen A, Fowler A (2024) Benchmarking UMI-aware and standard variant callers for low frequency ctDNA variant detection. *BMC Genomics* **25**: 827.

Møller AP, Mousseau TA (2013) The effects of natural variation in background radioactivity on humans, animals and other organisms: Evolution and natural variation in radioactivity. *Biological Reviews of the Cambridge Philosophical Society* **88**: 226-54.

Moreno-Villanueva M, Wong M, Lu T, Zhang Y, Wu H (2017) Interplay of space radiation and microgravity in DNA damage and DNA damage response. *NPJ Microgravity* **3**: 14.

Mousseau TA, Møller AP (2020) Plants in the light of ionizing radiation: What have we learned from Chernobyl, Fukushima, and other 'hot' places. *Frontiers in Plant Science* **11**: 552.

Ou X, Long L, Wu Y, Yu Y, Lin X, Qi X, Liu B (2010) Spaceflight-induced genetic and epigenetic changes in the rice (Oryza Sativa L.) genome are independent of each other. *Genome* **53**: 524-32.

Perera I, Sheppard JV, Land E (2019) The effect of spaceflight on transgenic Arabidopsis Plants with compromised signaling. *NASA GeneLab*. https://doi.org/10.26030/WDTZ-V612.

Perera IY, Hung CY, Brady S, Muday GK, Boss WF (2006) A universal role for inositol 1,4,5-trisphosphate-mediated signaling in plant gravitropism. *Plant Physiology* **140**: 746-60.

Poplin R, Ruano-Rubio V, DePristo MA, Fennell TJ, Carneiro MO, Van der Auwera GA, Kling DE, Gauthier LD, Levy-Moonshine A, Roazen D, Shakir K, Thibault J, Chandran S, Whelan C, Lek M, Gabriel S, Daly MJ, Neale B, MacArthur DG, Banks E (2017) Scaling accurate genetic variant discovery to tens of thousands of samples. *bioRxiv*. https://doi.org/10.1101/201178.

Ray S, Gebre S, Fogle H, Berrios DC, Tran PB, Galazka JM, Costes SV (2019) GeneLab: Omics database for spaceflight experiments. *Bioinformatics (Oxford, England)* **35**: 1753-59.

R Core Team. n.d. *R: A Language and Environment for Statistical Computing*, https://www.R-project.org/.

Rinaldi A (2016) Research in space: In search of meaning: Life science research aboard the International Space Station has come under scrutiny for its costs and apparent lack of returns. *EMBO Reports* **17**: 1098-1102.

Robinson MD, McCarthy DJ, Smyth GK (2010) edgeR: A bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics (Oxford, England)* **26**: 139-40.

Sayols S, Scherzinger D, Klein H (2016) dupRadar: A bioconductor package for the assessment of PCR artifacts in RNA-Seq data. *BMC Bioinformatics* **17**: 428.

Sheppard J, Land ES, Toennisson TA, Doherty CJ, Perera IY (2021) Uncovering transcriptional responses to fractional gravity in Arabidopsis Roots. *Life (Basel, Switzerland)* **11**: 1010.

Thirsk R, Kuipers A, Mukai C, Williams D (2009) The space-flight environment: The International Space Station and beyond. *Journal de l'Association Medicale Canadienne [Canadian Medical Association Journal]* **180**: 1216-20.

Vandenbrink JP, Kiss JZ (2016) Space, the final frontier: A critical review of recent experiments performed in microgravity. *Plant Science: An International Journal of Experimental Plant Biology* **243**: 115-19.

Wuest SL, Richard S, Kopp S, Grimm D, Egli M (2015) Simulated microgravity: Critical review on the use of random positioning machines for mammalian cell culture. *BioMed Research International* **2015**: 971474.

Xu P, Chen H, Jin J, Cai W (2018) Single-base resolution methylome analysis shows epigenetic changes in Arabidopsis seedlings exposed to microgravity spaceflight conditions on board the SJ-10 recoverable satellite. *NPJ Microgravity* **4**: 12.

Yuge L, Kajiume T, Tahara H, Kawahara Y, Umeda C, Yoshimoto R, Wu SL, Yamaoka K, Asashima M, Kataoka K, Ide T (2006) Microgravity potentiates stem cell proliferation while sustaining the capability of differentiation. *Stem Cells and Development* **15**: 921-29.

Zhang LF (2001) Vascular adaptation to microgravity: What have we learned. *Journal of Applied Physiology (Bethesda, Md.: 1985)* **91**: 2415-30.

Zhang Y, Richards JT, Hellein JL, Johnson CM, Woodall J, Sorenson T, Neelam S, Ruby AMJ, Levine HG (2022) NASA's ground-based microgravity simulation facility. *Methods in Molecular Biology* **2368**: 281-99.

Supplemental Table S1. Number of variants remaining after each filtering step.

	Called by Mutect2	Hard Filtered	Deduplicated	Difficult to Sequence Reads Removed	Filtered by Gene Expression
Transgenic Shoot	48037	26220.1	3097.6	1639.7	1611.7
Transgenic Root	14638.1	1473.3	721.8	549	539.8
Wild-type Shoot	49944.7	25162.5	2208.2	1589.2	1571
Wild-type Root	15148.5	1516.4	745.7	571.5	563.2

Supplemental Table S2. Final numbers of variants for each sample.

| *Sample*	Line	Tissue Type	Treatment	Response
FLS18	Wild-Type	Shoot	Spaceflight with Gravity Control	1568
FLS28	Wild-Type	Shoot	Normal Spaceflight	1802
FLS38	Wild-Type	Shoot	Normal Spaceflight	1688
FLS48	Wild-Type	Shoot	Spaceflight with Gravity Control	1668
FLS58	Wild-Type	Shoot	Spaceflight with Gravity Control	1474
FLS68	Wild-Type	Shoot	Normal Spaceflight	2106
FLS78	Wild-Type	Shoot	Normal Spaceflight	1966
GCS1	Wild-Type	Shoot	Ground Control	1209
GCS14	Wild-Type	Shoot	Ground Control	1607
GCS18	Wild-Type	Shoot	Ground Control	622
FL08	Wild-Type	Root	Spaceflight with Gravity Control	617
FL18	Wild-Type	Root	Spaceflight with Gravity Control	522
FL28	Wild-Type	Root	Normal Spaceflight	500
FL38	Wild-Type	Root	Normal Spaceflight	511
FL48	Wild-Type	Root	Spaceflight with Gravity Control	414
FL58	Wild-Type	Root	Spaceflight with Gravity Control	471
FL68	Wild-Type	Root	Normal Spaceflight	454
FL78	Wild-Type	Root	Normal Spaceflight	704
GC14	Wild-Type	Root	Ground Control	573
GC16	Wild-Type	Root	Ground Control	866
GCS13	Transgenic	Shoot	Ground Control	1552
GCS17	Transgenic	Shoot	Ground Control	754
GCS2	Transgenic	Shoot	Ground Control	1094
FLS23	Transgenic	Shoot	Normal Spaceflight	1628
FLS33	Transgenic	Shoot	Normal Spaceflight	1644
FLS63	Transgenic	Shoot	Normal Spaceflight	1688
FLS73	Transgenic	Shoot	Normal Spaceflight	1946
FLS13	Transgenic	Shoot	Spaceflight with Gravity Control	1939
FLS43	Transgenic	Shoot	Spaceflight with Gravity Control	1683
FLS53	Transgenic	Shoot	Spaceflight with Gravity Control	2189
GC01	Transgenic	Root	Ground Control	904
GC13	Transgenic	Root	Ground Control	473
FL23	Transgenic	Root	Normal Spaceflight	518
FL33	Transgenic	Root	Normal Spaceflight	515
FL63	Transgenic	Root	Normal Spaceflight	567
FL73	Transgenic	Root	Normal Spaceflight	498
FL03	Transgenic	Root	Spaceflight with Gravity Control	404
FL13	Transgenic	Root	Spaceflight with Gravity Control	603
FL43	Transgenic	Root	Spaceflight with Gravity Control	452
FL53	Transgenic	Root	Spaceflight with Gravity Control	464

Supplemental Table S3. Least-squared means from Analysis of Variance.

Genotype	Tissue Type	Treatment	LSMean Estimate	Standard Error
Wild-Type	Shoot	Ground Control	1146	126.55
Wild-Type	Shoot	Spaceflight with Gravity Control	1570	126.55
Wild-Type	Shoot	Spaceflight	1890.5	109.59
Wild-Type	Root	Ground Control	719.5	154.99
Wild-Type	Root	Spaceflight with Gravity Control	506	109.59
Wild-Type	Root	Spaceflight	542.25	109.59
Transgenic	Shoot	Ground Control	1133.33	126.55
Transgenic	Shoot	Spaceflight with Gravity Control	1937	126.55
Transgenic	Shoot	Spaceflight	1726.5	109.59
Transgenic	Root	Ground Control	688.5	154.99
Transgenic	Root	Spaceflight with Gravity Control	480.75	109.59
Transgenic	Root	Spaceflight	524.5	109.59

Supplemental Table S4. Analysis of Variance results.

Comparison	Difference Estimate	SE	P-value
Shoot vs. root tissue	-990.306	71.523	1.34E-43
Wild-type vs. transgenic genotype	19.389	71.523	0.786
Interaction effect b/w tissue and genotype	-264.333	429.136	0.538
Interaction effect b/w tissue and gravity condition	-1708.917	572.959	0.003
Interaction b/w tissue type and non-microgravity spaceflight stress	3327.833	651.432	3.25E-07
Interaction b/w A. thaliana genotype and gravity condition	-661.583	572.959	0.248
Interaction b/w A. thaliana genotype and non-microgravity spaceflight stress	-247.333	651.432	0.704
Interaction b/w gravity and non-microgravity spaceflight stress	-12248.417	753.981	2.42E-59
Interaction b/w tissue type, A. thaliana genotype, and gravity condition	-703.083	572.959	0.22
Interaction b/w tissue type, A. thaliana genotype, and non-microgravity spaceflight stress	-209.333	651.432	0.748
Interaction b/w tissue type, gravity condition, and non-microgravity spaceflight stress	-4322.917	753.981	9.84E-09
Interaction b/w A. thaliana genotype, gravity condition, and non-microgravity spaceflight stress	-792.583	753.981	0.293
Interaction b/w tissue type, A. thaliana genotype, gravity condition, and non-microgravity spaceflight stress	-648.083	753.981	0.39
Gravity vs. microgravity	148.302	71.62	0.038
Gravity vs. microgravity in space	47.5	80.657	0.556
Gravity vs. microgravity in space in shoot tissue	55	118.372	0.642
Gravity vs. microgravity in space in root tissue	40	109.592	0.715
Spaceflight vs. ground control	225.354	81.429	0.006
Spaceflight with gravity control vs. ground control	-201.604	92.235	0.029
Spaceflight with gravity control vs. ground control in shoot tissue	-613.833	126.545	1.23E-06
Spaceflight with gravity control vs. ground control in root tissue	210.625	134.222	0.117
Spaceflight vs. ground control in shoot tissue	641.333	107.284	2.26E-09
Spaceflight vs. ground control in root tissue	-190.625	122.527	0.12
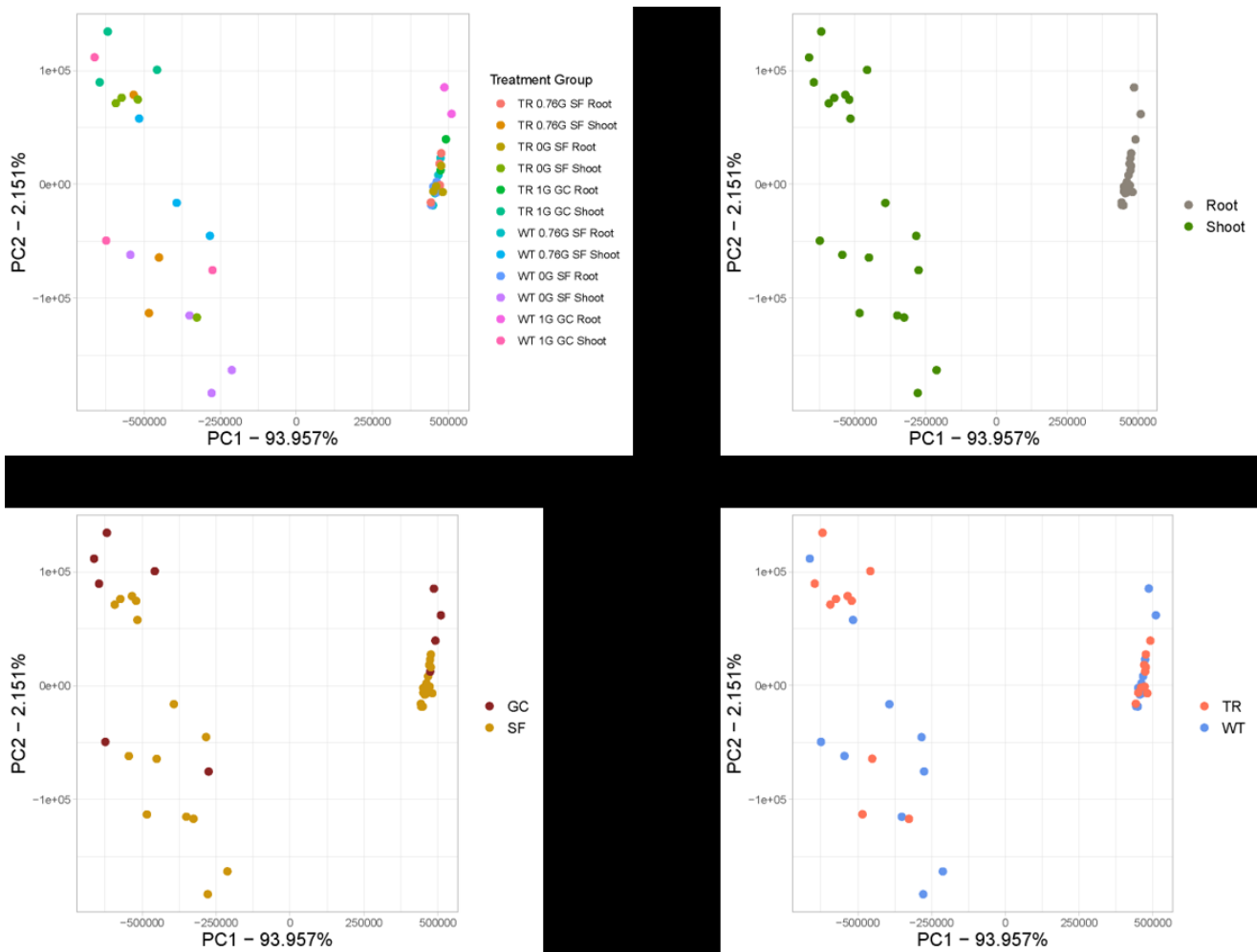
Supplemental Figure S1. Principal Component Analysis on normalized gene expression counts. The x-axis in each plot reflects variance across Principal Component 1 (PC1), which explained 93.96% of the variance. The y-axis in each reflects Principal Component 2 (PC2), which explained 2.15% of the variance. (A) The PCA plot colored by treatment group: TR=transgenic, WT=wildtype, SF=spaceflight, and GC=ground control. (B) The same PCA plot colored by tissue type. Tissue type appears to account for the majority of the variance in the first principal component. (C) The PCA plot colored by spaceflight condition, which does not appear to explain the variance on these two PCs. (D) The PCA plot colored by genotype, which also does not appear to explain the variance on these two PCs.