Visualizing Omics Data from Spaceflight Samples using the NASA GeneLab Platform

Daniel Berrios[1], Eric Weitz[2], Kirill Grigorev[3], Sylvain V. Costes[4], Samrawit G. Gebre[5], and Afshin Beheshti[5]

[1] USRA, NASA Ames Research Center, Moffett Field, CA, 94035
[2] Data Sciences Platform, Broad Institute of MIT and Harvard, Cambridge, MA, 02142
[3] Institute for Computational Biomedicine, Weill Cornell Medical College, New York, NY, 10065
[4] Space Biosciences Division, NASA Ames Research Center, Moffett Field, CA, 94035
[5] KBRWyle Labs, Space Biosciences Division, NASA Ames Research Center, Moffett Field, CA, 94035

daniel.c.berrios@nasa.gov, eweitz@broadinstitute.org,
kig2007@med.cornell.edu, sylvain.v.costes@nasa.gov,
samrawit.g.gebre@nasa.gov, afshin.beheshti@nasa.gov

Abstract

NASA's GeneLab platform is the first omics database and platform limited to studies related to space flight. This platform provides the scientific community a valuable resource for exploring, evaluating, and discovering hypotheses of space biology. In the platform we offer users multiple tools to explore GeneLab data including: a comprehensive, well-annotated data repository for all space related omics data, omics data analysis tools, space environmental metadata, and differential gene expression visualization tools. Here we focus on the latter, and how these tools can be utilized to explore scientific questions related to space biology. Spaceflight is known to create many health risks in astronauts due to both microgravity and space radiation exposure. Current research is focused on the underlying biological mechanisms behind these health risks and potential countermeasures to mitigate these health risks. One specific health risk is related to increased liver damage during exposure to the space environment (Beheshti et al, Scientific Reports, 2019). Longer exposure to the space environment appears to cause increased lipid accumulation and transcriptomic analysis suggests pathways and genes related to liver disease (i.e. PPARα, Insulin or INS, and Glucagon or GCG) can be activated. We use this scenario to demonstrate how GeneLab's platform and visualization capabilities can be utilized to quickly reveal this process in GeneLab platform data. Lastly, we discuss some issues and best practices for implementing similar, public-platform visualization tools.

1 Introduction

The impact of the space environment on astronauts has been known to increase the chance of many pathologies, including: cardiovascular disease, central nervous system (CNS) disorders, muscle degeneration, bone loss, and cancer [1]. Since upcoming NASA missions will be focused on more long-term and even deep space travel, these health risks will be of even greater interest to NASA in order to more fully understand their biological mechanisms and to develop countermeasures. Recently it has been observed that spaceflight might increase chances of liver disease, specifically non-alcoholic fatty liver disease [2]. This work was done by utilizing capabilities of NASA's GeneLab platform, a data repository and set of analysis tools for space-related omics data [3-5]. The repository contains raw and processed omics space research data (i.e. transcriptomics, proteomics, methylomics, metabolomics, genomics, metagenomics, etc.) for many different species, as well as tools to store and share these or proprietary omics data, a modified version of the Galaxy tool [6] for web-based analysis of omics data, a repository of space environmental data (i.e. accumulated space radiation exposure, environment factors on the International Space Station (ISS), etc.), and, most recently, omics data visualization tools. The GeneLab platform has already been utilized to generate novel and interesting science [3-5, 7, 8]. The omics data visualization tools are a new feature of the platform, including: basic omics visualization features such as principal component analysis (PCA) [9], pair plots, volcano plots [10], heatmaps, and data query tools. There are also more advanced analysis and visualization capabilities including S-plots (a graphical analysis tool at the biochemical pathway level) and chromosome analysis visualization tools referred to as "Ideogram". Below, we discuss the challenges regarding designing and implementing these visualization tools, with specific examples using GeneLab data.

Finally, we describe how visualization tools like these can be utilized to generate new hypotheses related to space biology research, using as example our recent publication describing lipid accumulation after exposure to the space environment [2]. This work utilized four sets of transcriptomic and proteomics data in the GeneLab repository from mice that were flown to space. Our analysis revealed lipid-related pathways, genes, and proteins that were contributing to demonstrable lipid accumulation in liver tissue, with potentially an increased risk of liver disease. Here, we show using the new GeneLab visualization tools how this effect can be observed and studies. Lastly, we demonstrate how Ideogram can be utilized to reveal how certain chromosomes will be impacted by spaceflight.

2 NASA GeneLab Visualization Tools and Capabilities

In this section we discuss GeneLab's visualization tools, beginning with how the tools were developed and implemented, to examples of the application of each tool to visualize GeneLab data and data analysis result.

2.1 Server and Software Design

2.1.1. Development Framework

For bioscience applications, the use of open-source development frameworks has become widespread, particularly those available for scripting languages like python and R. These languages have seen substantial growth in code libraries available for scientific data analysis and visualization. In order to meet the goals for GeneLab data visualization, we chose to develop our application using the open-source application frameworks Flask (Python), and D3 (javascript). Using these frameworks

allowed us to reduce development time significantly by leveraging numerous pre-built server and client-side libraries for data transformation, plotting, and visualization.

2.1.2. Hosting, Data Storage and Transfer

The GeneLab project includes collaborations with several on-site contracting organizations as well as with numerous faculty and staff from universities and other research institutions across the globe. However, deployment of software into U.S. government-managed production environments can be onerous, as such environments often require tighter security controls than many academic environments, largely because they are more frequently targets for hackers. For this reason, we chose to deploy our application into an environment for education and research, supported by the Universities Space Research Association (USRA), an on-site NASA service contractor. This choice of hosting environment afforded us the opportunity to respond more agilely to end-user requirements, as the time from development to deployment can be more extensively controlled by the GeneLab project, rather than by universally levied government agency-wide edicts.

The GeneLab data repository employs a commercial cloud storage solution for both raw and processed omics data. This choice supports unlimited growth of the GeneLab repository data. Similarly, developers of applications that consume these data, such as the visualization application discussed in this paper, need not be concerned with data storage limitations. Due to this design, the transfer of these data from cloud locations to target applications requires network traversal. Therefore, we developed a local caching mechanism for visualization application data files, so that loading of the data into the application can be expedited, once cached.

For this caching system, the visualization tools access the underlying data using a restful API (https://github.com/LankyCyril/genefab) written in Python and utilizing the Flask framework (https://palletsprojects.com/p/flask/). The API exposes metadata associated with datasets and with assays, and processed assay data in several formats. Datasets metadata contains information about factors and assays associated with the dataset, and is accessed by appending the dataset accession name to the root URL of the API. By default, the information is returned in tab-separated format, but can be represented as JSON by passing the &fmt=json parameter. Data from any file presented in the array metadata can be requested by supplying a regular expression matching the metadata column name and an expression matching the file name to the assay data URL. For the most commonly used data types, short aliases are implemented. Finally, sample and factor information from the metadata can be added to the narrow-form data frame. Usage of the API is outlined at its root URL (http://genelab-visualization.usra.edu/genefab/) and is being iteratively updated and expanded.

2.2 Basic Visualization Tools

We sought to develop a set of visualizations, in order to allow users to quickly be able to see if a given dataset shows substantial variation between samples, and therefore would a good candidate for further analysis. Some of these tools are basic visualization which are widely used for omics analysis. The goal of these tools was to allow these users to quickly answer the question "are they obvious differences in pattern of biological factors between subgroups?" The first visualization is a PCA plot [9] (**Figure 1A**). The plot is interactive, and users can remove groups from the plot by clicking on the group name in the legend. PCA plots are a logical choice as the first method to visualize these data; these plots allow a user to quickly view how gene expression of individual biological samples group and cluster together in inferred component vector space. For example, in **Figure 1A** the user is comparing gene expression from samples flown in space to two types of ground control sample. Users should be able to clearly identify that each group of samples are clustering together indicating that this dataset is a good candidate to study further the specific gene expression differences occurring between
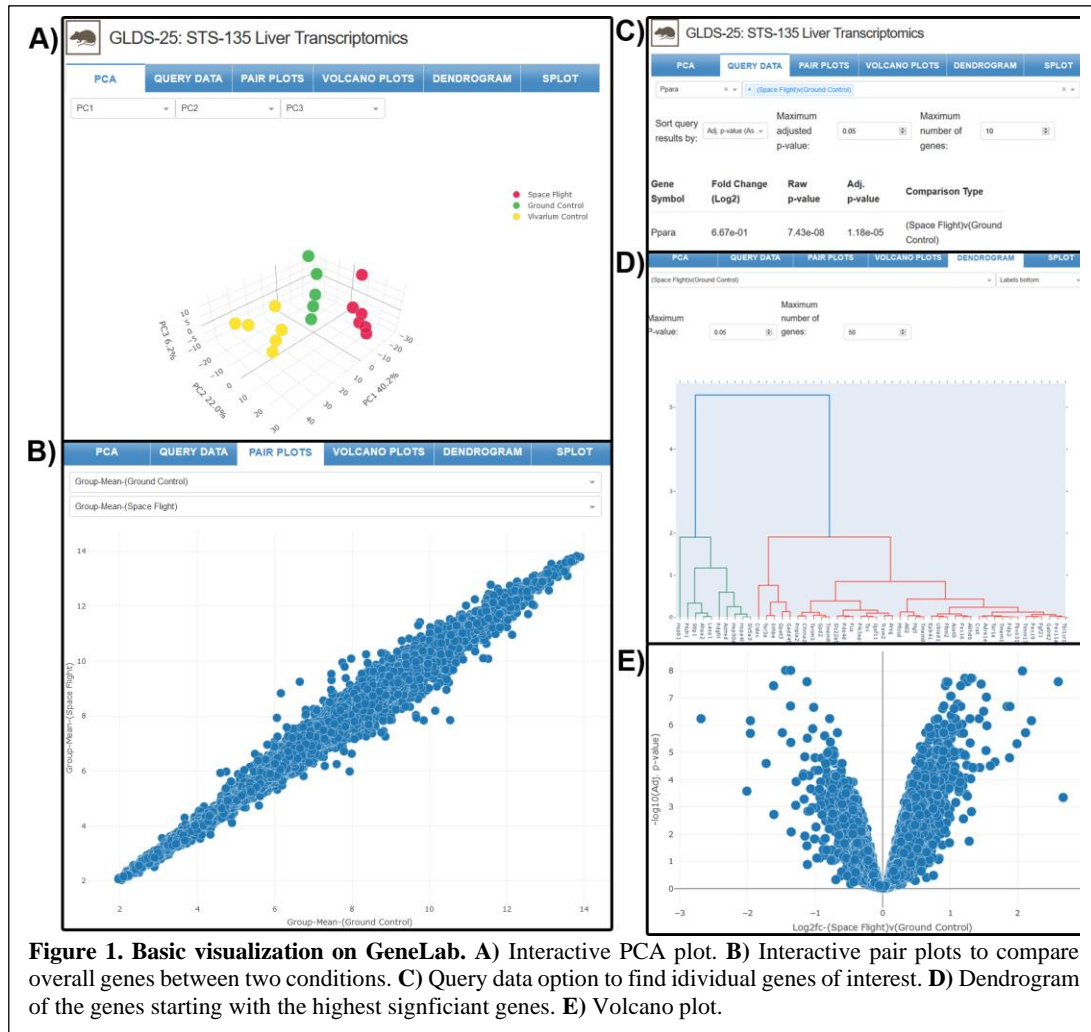
Figure 1. Basic visualization on GeneLab. A) Interactive PCA plot. **B)** Interactive pair plots to compare overall genes between two conditions. **C)** Query data option to find idividual genes of interest. **D)** Dendrogram of the genes starting with the highest signficiant genes. **E)** Volcano plot.

the groups. Users can download the PCA plots as a PNG file, zoom in and out of plots, rotate the plots, and reset plots to default settings.

Using a second visualization tool, the pair plot, users can study how close each biological sample (or replicate) is to another (**Figure 1B**). Pair plots simply plot the gene expression (or other factors) levels of one sample/replicate compared to another; users can alternatively opt to compare the mean expression values of two experimental groups (i.e. the average of all flight data versus the average of all ground data), or any sample/replicate vs. any group mean. Similar interactive capabilities are available as described for the PCA plots above.

To query the expression of specific genes of interest per dataset, we developed a "query tab" visualization in which users can enter a gene of interest (**Figure 1C**). For each gene entered/selected, the following information will appear for each sample or group selected: $\log_2$ fold-change, raw and adjusted p-value.
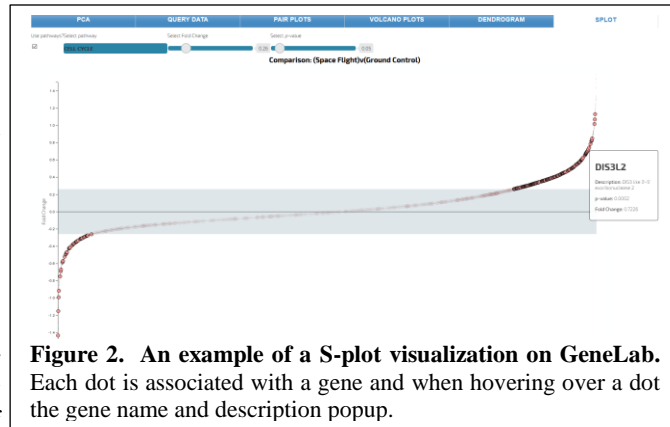
If a user wants to view how the top varying genes cluster together (by Pearson correlation distances), they can use the dendrogram visualization tool (**Figure 1D**). By default, the visualization tool loads the top 100 statistically significantly differentially expressed genes (i.e. ordered by p-value) into the dendrogram. The user has the option to change the number of genes being displayed and/or zoom in

and out of specific regions of the dendrogram. These plots can be utilized to view how groups of genes for each dataset and biological comparison cluster by amount of differential expression.

Lastly, users can also visualize at standard volcano plot of differential gene expression (**Figure 1E**) [10]. These plots can quickly allow the user to identify how many significantly regulated genes exist within each group, and correlate amount of differential expression with statistical significance for each gene. The volcano plots have similar interactive controls as the previously described plots.

2.3 Pathway Visualization Tool

There are many different ways to visualize biological pathways for omics analysis. Our pathway visualization tool utilizes an S-plot of

Figure 2. An example of a S-plot visualization on GeneLab. Each dot is associated with a gene and when hovering over a dot the gene name and description popup.

statistical significance vs. fold change of differential gene expression an overlay of biochemical pathway association (**Figure 2**). The S plot allows a user to determine how differentially expressed genes in each dataset are impacting pathways through each gene's with one or more Gene Ontology (GO) pathways. The interactive plot allows a user to adjust both significance level and fold-change cut-off values. The S-plot loads with no pathways indicated on the plot. Once the user selects a pathway of interest, which they can search for or select using the pathway input selector, all significantly (by fold change and statistical significance) differentially expressed genes within the dataset will be highlighted with a larger circle. For example, in F**igure 2** we have plotted the genes for the data set "GLDS-25" with a p-value cut-off of < 0.05 highlighting all genes related to the "Cell Cycle" pathway. The gray bar covers the genes with a $\log_2$ fold-change < 1.2 or > -1.2. This grey bar can be adjusted to include more fold-change values with the top slider bar or by entering a value directly in the slider box. The user can also adjust the p-values to include more genes with less statistical significance or highlight fewer genes with a higher statistical cutoff. Lastly, if a user hovers over any of the highlighted genes, a popup window appears with the gene name, a description of the gene, and exact p-value, and $\log_2$ fold-change value.

2.4 Ideogram: Chromosome Visualizations

Ideogram.js (https://github.com/eweitz/ideogram) is a JavaScript library for chromosome visualization. Ideogram supports drawing and animating genome-wide datasets for human, mouse, and many other eukaryotes. The Ideogram API for annotations supports histograms, heatmaps, overlays, and points of arbitrary shape and color layered in tracks. Ideogram can be embedded as a reusable component in any web page or application. Examples at https://eweitz.github.io/ideogram show how to use Ideogram to research and report findings on cancer, clinical variants, gene expression, evolution, agriculture, and more. GeneLab has developed a novel genome visualization using Ideogram to enable users to explore graphically how space flight affects gene expression on chromosomes (**Figure 3**). When a user first displays Ideogram for a dataset, all the chromosomes associated with the specific species for that datasets appear with a distribution and location of all the $\log_2$ fold-change values for the genes mapped on the chromosomes (F**igure 3**). The visualization is interactive; the user clicks on each chromosome to zoom in on the specific chromosome for further details. In addition, the user can adjust both the $\log_2$ fold-change cut-off values and the adjusted p-values . The user can also select which group comparison in that dataset will be mapped on the chromosomes, for datasets with more than two
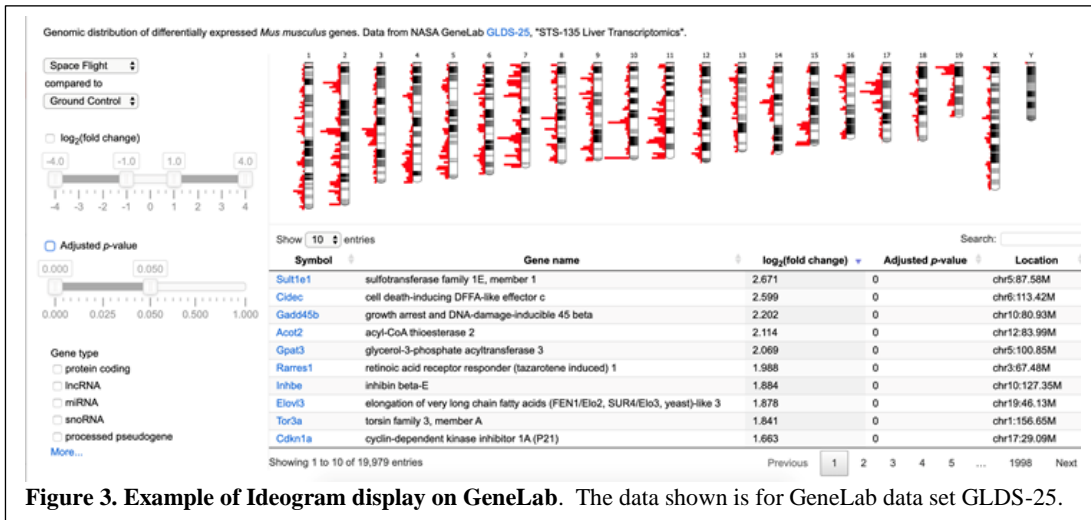
Figure 3. Example of Ideogram display on GeneLab. The data shown is for GeneLab data set GLDS-25.

subgroups. Lastly, a user can filter the type of genes (i.e. protein coding RNAs, lncRNAs, miRNAs, snoRNAs, etc.). There is an additional gene query function under the chromosomes with the $\log_2$ fold-change values, adjusted p-value and the chromosome location, and the user can search for specific gene(s) of interest.

3   Analysis with GeneLab Visualization: Lipid Accumulation During Spaceflight Leads to Liver Disease.

To demonstrate the visualization capabilities on GeneLab we will provide a brief example of how a user can interact with the visualization tools to produce novel hypotheses related to biological changes that can occur during spaceflight. We will focus on potential increases in lipid accumulation in the liver, and the key biological functions and pathways driving this response [2]. For this analysis we will focus on the GLDS-25 liver dataset on GeneLab. These data are from an experiment using C57BL/6 wild-type female mice which were flown for 13 days in space on a U.S. "Space Shuttle" mission at 9 weeks of age [11, 12]. During this experiment the mice were flown back to Earth and sacrificed within 24 hours after return. These experiments also contained two ground controls: 1) a habitat ground control (also previously referred to as an Animal Enclosure Module, AEM) and 2) a vivarium ground control. The mice in the habitat ground control (or AEM control) experienced the same environmental conditions (temperature, humidity, etc.) as the mice flown in space, except for microgravity and space radiation components [4]. The mice in the vivarium controls were housed in "normal" animal housing and conditions for all regular experiments performed on Earth. The livers of the mice were harvested and the samples were run on an Affymetrix microarray platform to obtain the transcriptomic data. Additional experimental details are available on GeneLab for this dataset (https://genelab-data.ndc.nasa.gov/genelab/accession/GLDS-25).

3.1   Differences Between Flight and Ground Samples

To examine if this dataset has any changes between spaceflight and ground control samples, we first observe its PCA plot (**Figure 1A**): a clear separation occurs between the group of biological replicates for the spaceflight samples and the two ground controls on this plots). This suggests that spaceflight induces a key biological global response specifically related to the conditions in space in liver tissue.

Interestingly, global biological differences may also be present between the two ground controls. We have reported that key biological differences can occur in mice depending on the housing conditions utilized on Earth [7]. We discovered that these biological differences occur due to the ambient CO_2 differences for mice housed in the ground control habitat versus in the ground control vivarium. The habitats have CO_2 levels which are approximately 10-fold higher than the vivarium controls (or
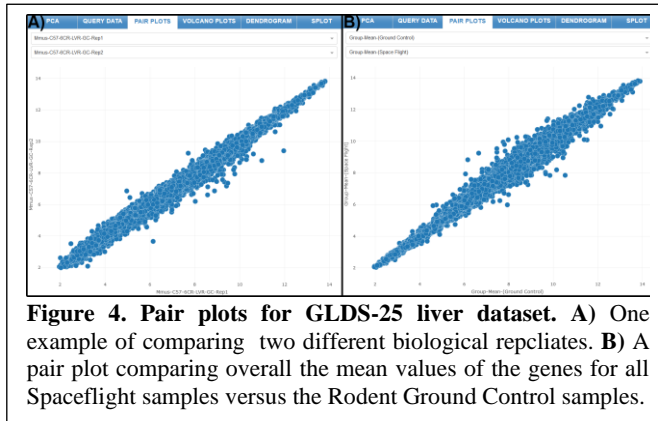
Figure 4. Pair plots for GLDS-25 liver dataset. A) One example of comparing two different biological repcliates. **B)** A pair plot comparing overall the mean values of the genes for all Spaceflight samples versus the Rodent Ground Control samples.

equivalent to the normal CO_2 in our atmosphere). The habitat ground controls have a higher CO_2 level that match the higher CO_2 levels which are present on the International Space Station and Space Shuttle missions. This example demonstrates that a user can easily notice on a biological global scale these key intriguing changes occurring and start exploring potential explanations for these differences.
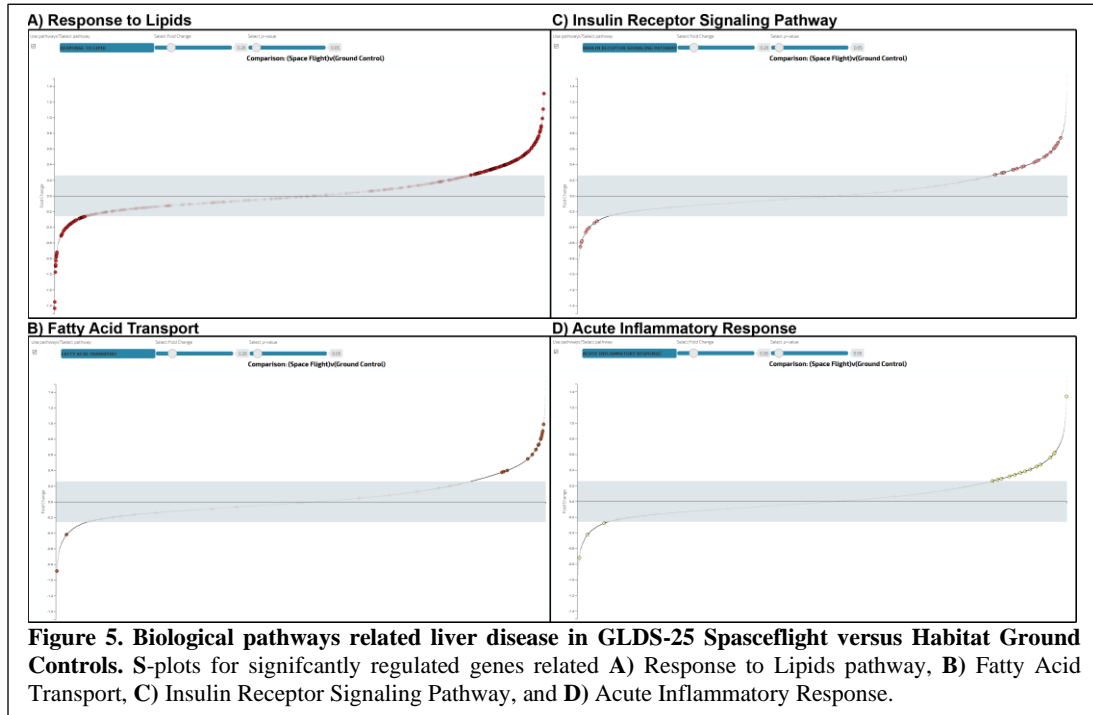
To determine the variance that might occur in the biological replicates and the overall amount of significantly regulated genes, we utilized both the pair plots and the volcano plots. When comparing the individual biological replicates for GLDS-25 within each experimental group, we observe little variance occurring between the biological replicates (**Figure 4A**). We have shown one example of this in **Figure 4A** displaying the comparison between two different biological replicates for the habitat ground controls. This example shows how tightly the genes are correlated between the biological replicates which is good indication that there is little variance between the two biological replicates.

Next, we want to look for differences in gene expression occurring between the spaceflight and ground control sample group. By selecting to compare the mean expression values in each group, we observe that there appears to be many differentially expressed genes (**Figure 4B**). The volcano plot for this comparison (**Figure 1E**) also indicates that there are many differentially regulated genes by both adjusted p-value < 0.05 and overall fold-change values > 1.2. By quickly utilizing the GeneLab visualization tools we are able to determine that GLDS-25 has global differences occurring in the liver tissue samples due to spaceflight justifying further exploration of these data.

3.2 Key Lipid Related Pathways Being Dysregulated During Spaceflight

Finally, we want to examine if the observed differential gene expression changes we see occurring in the liver tissue samples are lipid pathway related, as previously published [2, 11, 13]. The changes previously reported in the liver has been known to be associated with non-alcoholic fatty liver disease also known as NAFLD [14]. To examine specific biological pathways that are associated with NAFLD, we utilize the S-plot. It is known that lipid and fatty acid pathways can have drastic impact on the development of liver disease [15, 16]. Since we also know from previously reported data that lipid accumulation occurs in livers during spaceflight [2, 11], we will utilize the S-plots to investigate lipid and fatty acid pathways changes occurring in the GLDS-25 dataset. The S-plots revealed that the majority of genes related to lipid (**Figure 5A**) and fatty acid pathways (**Figure 5B**) are up-regulated in the space environments, indicating that the pathways are most likely activated. It has also been previously reported that increased insulin dysregulation are heavily involved with risk of developing NAFLD [17, 18]. Again, using the S-plot, we observe that the majority of genes involved with "Insulin Receptor Signaling" pathway are also up-regulated in space environments, indicating that this pathway is also likely being activated (**Figure 5C**). Lastly, it is known that inflammation is a key player with the progression of liver disease and NAFLD [19]. The S-plot with the pathway "Acute Inflammatory
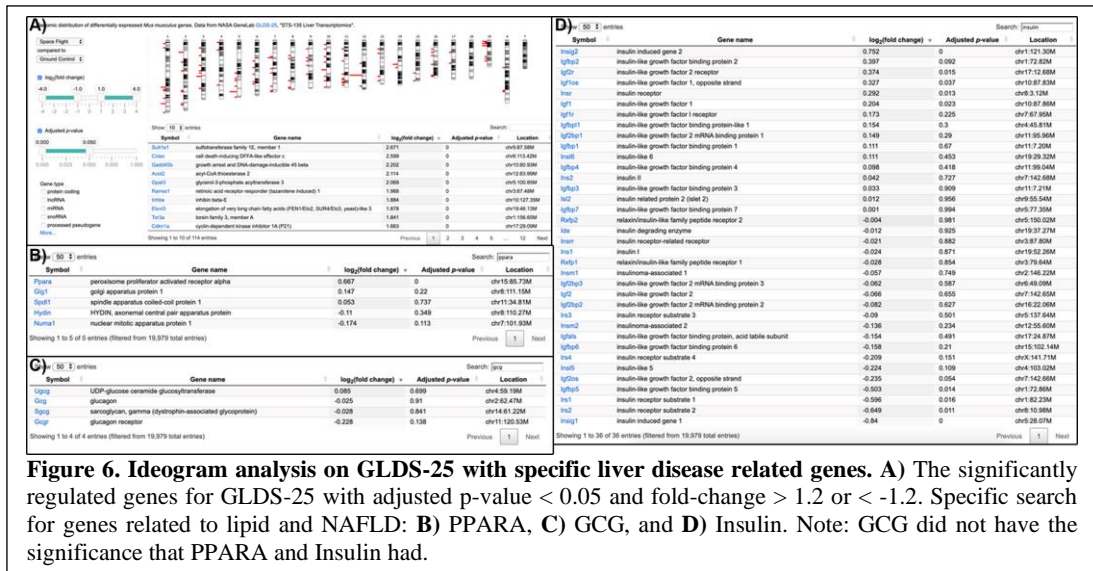
Figure 5. Biological pathways related liver disease in GLDS-25 Spasceflight versus Habitat Ground Controls. S-plots for signifcantly regulated genes related **A)** Response to Lipids pathway, **B)** Fatty Acid Transport, **C)** Insulin Receptor Signaling Pathway, and **D)** Acute Inflammatory Response.

Response" selected again reveals a majority of significantly up-regulated genes in the liver, indicating activation of this pathway (**Figure 5D**). Overall these analyses using the GeneLab S-plot reveals that known pathways related to increase risk of liver disease are being activated by space flight.

3.3 Spaceflight Lipid Dysregulation Impacting Specific Chromosomes

It is known that chromosome location of genes can provide important information on the function of the genes and overall impact of the biological mechanisms of diseases. The impact of spaceflight on

Figure 6. Ideogram analysis on GLDS-25 with specific liver disease related genes. A) The significantly regulated genes for GLDS-25 with adjusted p-value < 0.05 and fold-change > 1.2 or < -1.2. Specific search for genes related to lipid and NAFLD: **B)** PPARA, **C)** GCG, and **D)** Insulin. Note: GCG did not have the significance that PPARA and Insulin had.

causing chromosomes abnormalities is also of great concern as to how it can alter and increase potential health risks. For liver disease specifically, it has been reported that the location and dysregulation of genes on specific chromosome regions can increase risk for both lipid accumulate and NAFLD [20, 21]. We utilized the GeneLab ideogram visualization to start exploring how the genes in specific chromosomal regions can contribute to liver disease health risk due to spaceflight. Again, using GLDS-25 data, we observed the distribution of significantly regulated genes (adjusted p-value < 0.05) with an overall fold change of 2 on the chromosomes (**Figure 6A**). We noticed several chromosomes on whom the majority of the genes being dysregulated: 1, 2, 3, 5, 7, 8, 9, 10, 11, 12, 15, and 19. It has been reported in a large GWAS study on genetic variants associated with NAFLD that chromosome 8 is heavily involved with progression to NAFLD, chromosome 7 is associated with fibrosis in the liver, chromosomes 10, 11, and 12 are involved with lobular inflammation, and chromosomes 1, 4, and 12 are involved with elevated aspartate aminotransferase level associated with NAFLD [20]. In another publication, the loss of chromosome 8p has been associated with altered lipid metabolism and associated with liver tumor progression [21].

Based on our previous publication [2] and the genes known to be associated with NAFLD, we did a data query using the GeneLab Ideogram visualization for the following genes: PPARα, GCG (or glucagon), and Insulin. PPARα is significantly upregulated with a large fold-change and is found to be on chromosome 15 (**Figure 6B**). GCG is found on chromosome 2 (**Figure 6C**) and the majority of the insulin related genes are found on chromosomes 1, 2 8, and 10 (**Figure 6D**). The visualization through Ideogram of the location of these genes on the chromosomes related with NAFLD and found to be heavily regulated during spaceflight in the liver brings additional insight on how spaceflight will induce changes in the liver at the chromosomal-genetic level.

4 Conclusions

NASA's GeneLab is a platform that allows users with various backgrounds, from advanced bioinformaticians to neophyte scientists with little background in genomics to easily navigate and visualize the GeneLab omics data sets. Here, we have provided a description of the GeneLab visualization tools and provided a scientific example of how an individual can use them to quickly visualize differential gene expression data. Specifically, we have shown that by utilizing these visualization tools, users can quickly generate and test hypotheses, leading to further experimentation and validation of existing data. The basic visualization tools which include PCA plots, volcano plots, dendrograms are common tools that will allow users a quick guide of the data. The new and unique tools which include the S-plots and Ideogram allow users to explore the space biology omics data with a new approach. Specifically, Idoegram allows users to map genes onto chromosome which is not commonly explored with omics data involving space biology. The visualization tools on GeneLab can continue to evolve rapidly since we chose an open-source development framework. In addition, we continue to seek collaborations from the community to add more visualization tools to help the scientific community analyze GeneLab data.

5 Acknowledgements

References

[1]   Health Standards for Long Duration and Exploration Spaceflight: Ethics Principles, Responsibilities, and Decision Framework, ed. J. Kahn, C.T. Liverman, and M.A. McCoy. 2014, Washington (DC).

[2]   Beheshti, A., et al., *Multi-omics analysis of multiple missions to space reveal a theme of lipid dysregulation in mouse liver.* Sci Rep, 2019. **9**(1): p. 19195.

[3]   Beheshti, A., et al., *NASA GeneLab Project: Bridging Space Radiation Omics with Ground Studies.* Radiat Res, 2018. **189**(6): p. 553-559.

[4]   Beheshti, A., et al., *Exploring the Effects of Spaceflight on Mouse Physiology using the Open Access NASA GeneLab Platform.* J Vis Exp, 2019(143).

[5]   Ray, S., et al., *GeneLab: Omics database for spaceflight experiments.* Bioinformatics, 2019. **35**(10): p. 1753-1759.

[6]   Afgan, E., et al., *The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2018 update.* Nucleic Acids Res, 2018. **46**(W1): p. W537-W544.

[7]   Beheshti, A., et al., *Global transcriptomic analysis suggests carbon dioxide as an environmental stressor in spaceflight: A systems biology GeneLab case study.* Sci Rep, 2018. **8**(1): p. 4191.

[8]   Beheshti, A., et al., *GeneLab Database Analyses Suggest Long-Term Impact of Space Radiation on the Cardiovascular System by the Activation of FYN Through Reactive Oxygen Species.* Int J Mol Sci, 2019. **20**(3).

[9]   Lever, J., M. Krzywinski, and N. Altman, *Principal component analysis.* Nature Methods, 2017. **14**(7): p. 641-642.

[10]   Li, W., *Volcano plots in analyzing differential expressions with mRNA microarrays.* J Bioinform Comput Biol, 2012. **10**(6): p. 1231003.

[11]   Jonscher, K.R., et al., *Spaceflight activates lipotoxic pathways in mouse liver.* PloS one, 2016. **11**(4): p. e0152877.

[12]   Pecaut, M., et al., *STS-135 Liver Transcriptomics.* 2015, NASA GENELAB.

[13]   Blaber, E.A., M.J. Pecaut, and K.R. Jonscher, *Spaceflight Activates Autophagy Programs and the Proteasome in Mouse Liver.* Int J Mol Sci, 2017. **18**(10).

[14]   Cobbina, E. and F. Akhlaghi, *Non-alcoholic fatty liver disease (NAFLD) - pathogenesis, classification, and effect on drug metabolizing enzymes and transporters.* Drug Metab Rev, 2017. **49**(2): p. 197-211.

[15]   Musso, G., R. Gambino, and M. Cassader, *Recent insights into hepatic lipid metabolism in non-alcoholic fatty liver disease (NAFLD).* Progress in lipid research, 2009. **48**(1): p. 1-26.

[16]   Nguyen, P., et al., *Liver lipid metabolism.* J Anim Physiol Anim Nutr (Berl), 2008. **92**(3): p. 272-83.

[17]   Henkel, J., et al., *Oncostatin M produced in Kupffer cells in response to PGE 2: possible contributor to hepatic insulin resistance and steatosis.* Laboratory investigation, 2011. **91**(7): p. 1107.

[18]   Kantartzis, K., et al., *The DGAT2 gene is a candidate for the dissociation between fatty liver and insulin resistance in humans.* Clinical Science, 2009. **116**(6): p. 531-537.

[19]   Arrese, M., et al., *Innate immunity and inflammation in NAFLD/NASH.* Digestive diseases and sciences, 2016. **61**(5): p. 1294-1303.

[20]   Chalasani, N., et al., *Genome-wide association study identifies variants associated with histologic features of nonalcoholic Fatty liver disease.* Gastroenterology, 2010. **139**(5): p. 1567-76, 1576 e1-6.

[21]   Cai, Y., et al., *Loss of Chromosome 8p Governs Tumor Progression and Drug Response by Altering Lipid Metabolism.* Cancer Cell, 2016. **29**(5): p. 751-766.