



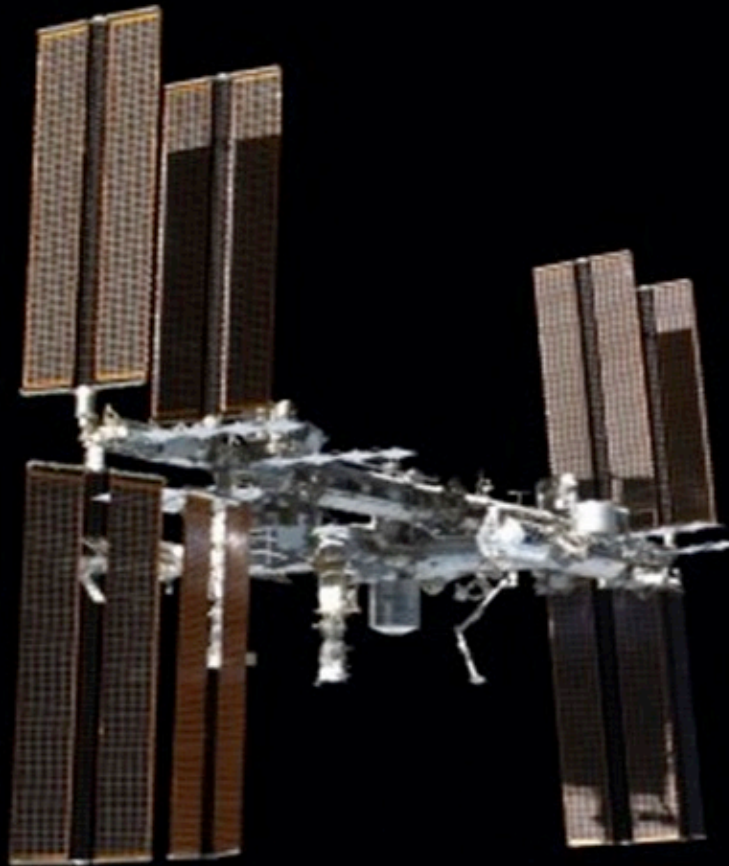
GeneLab: Open Science for Exploration

Jonathan Galazka, Ph.D
Project Scientist

The GeneLab Team



ISS: Our Orbiting Laboratory



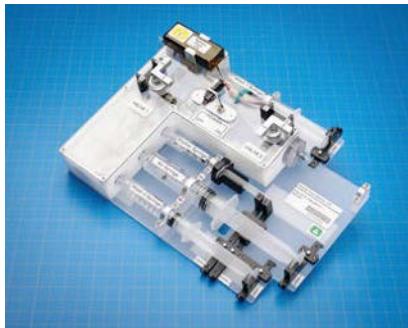
ISS enabling capability for research in cellular and molecular biology includes equipment for *in situ*, on-orbit analysis of biomolecules

Applications of this growing capability range from biomedicine and biotechnology to the growing field of Omics

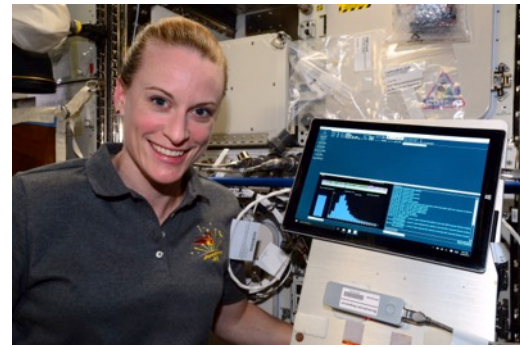
Omics Acquisition in Space is Now a Reality



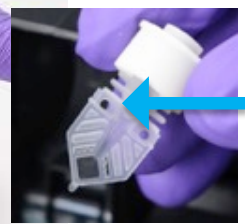
This is truly an exciting time for cellular and molecular biology, omics and biomedicine research on ISS with these amazing additions to the suite of ISS Laboratory capabilities.



Sample Preparation Module



Oxford Nanopore MinION Gene Sequencer



Reaction tube containing lyophilized chemical assay bead (proprietary)

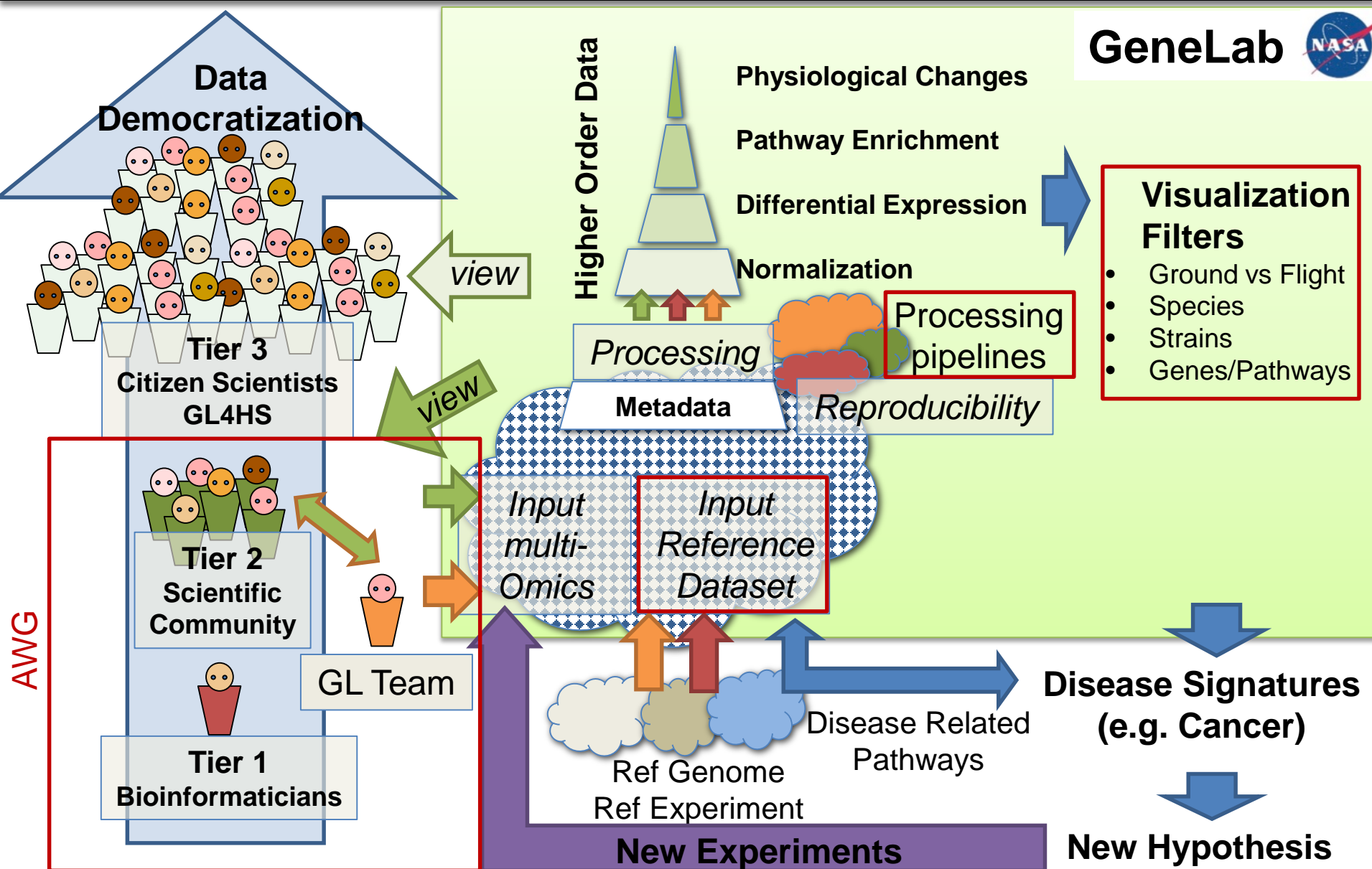
Cepheid Smart Cycler qRT-PCR

- New technologies to produce high-quality Omics data from research missions aboard the ISS
- Limited access and high demand for the ISS platform
- Facilitate systems biology to predict and/or mitigate changes due to microgravity



NASA astronaut Barry "Butch" Wilmore setting up the Rodent Research-1 hardware in the Microgravity Science Glovebox aboard the International Space Station.

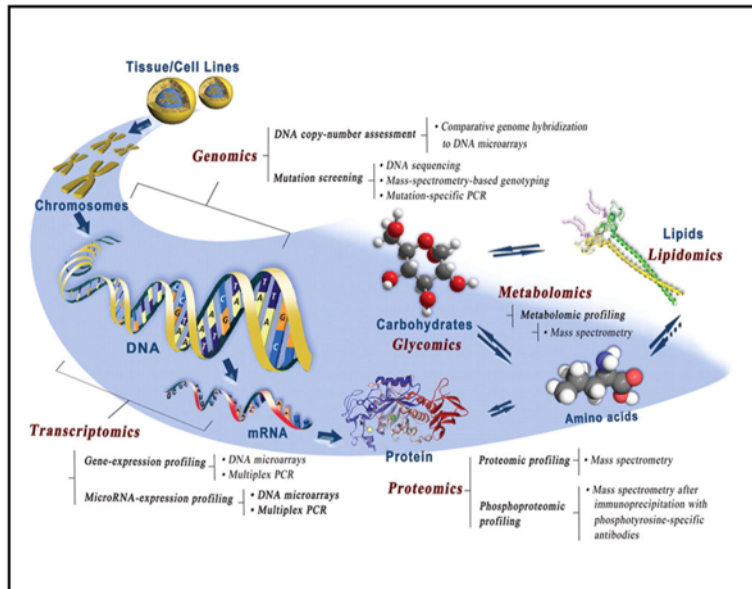
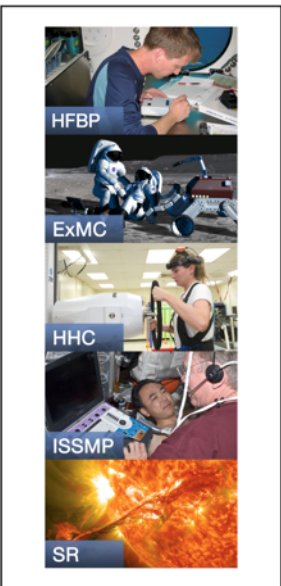
Three-tier Client Strategy to Democratize Data





OMICS

GENE LAB



DATA REPOSITORY 

ANALYSIS TOOLS 

COLLABORATIVE WORKSPACE 

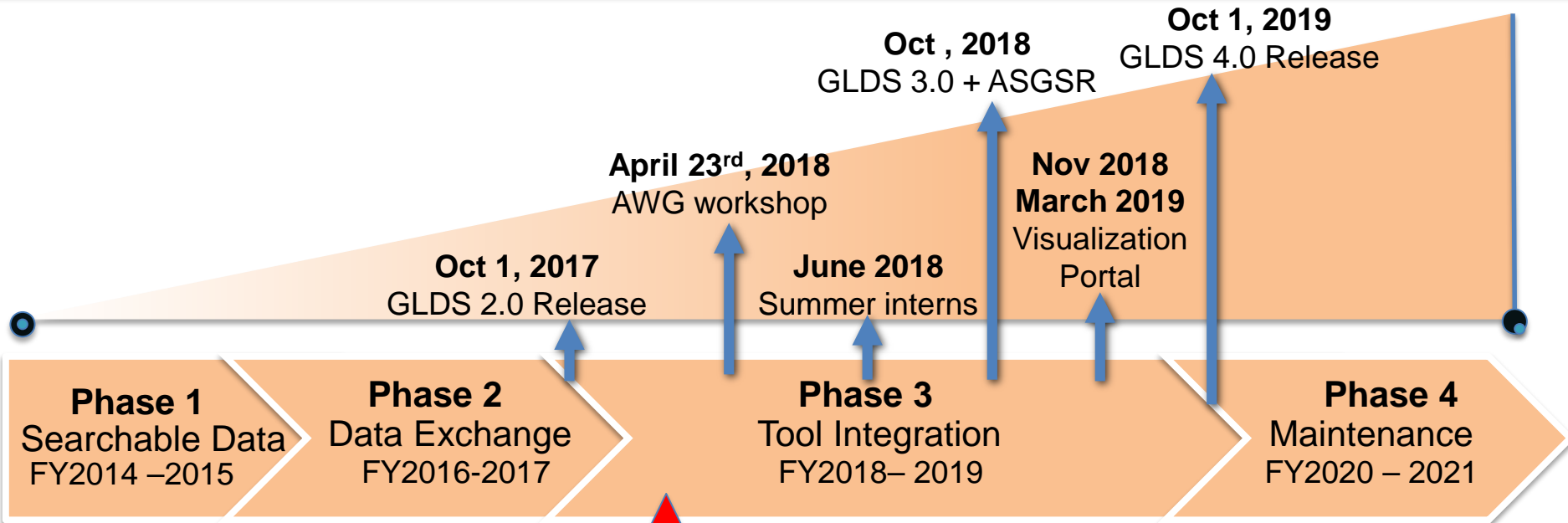
FEDERATED DATABASE





The GeneLab database infrastructure provides a platform for storage, retrieval and analysis of omics datasets – with the ultimately goal to support the missions of HRP

Phased Implementation



Data System

- ✓ Public Website
- ✓ Searchable Data Repository
- ✓ Top Level Requirements
- ✓ New Data and Legacy Data

Data System

- ✓ Link to Public Databases via Data Federation
- ✓ Integrated Search (e.g., data mashup)

Data System

- Integrated Platform across model organisms
- Build Community via AWG
- Provide access to biocomputational tools for omics analysis
- Provide collaboration framework and tools

Open Source Maintenance

- User community becomes primary provider of new tools/knowledge
- Maintain integrity of data, and data system

GLDS Phase 2 (Release 2.0) Google-like Search, Federated Search



Data federation/integration with heterogeneous bioinformatics external databases (GEO, PRIDE, MG-RAST)

The screenshot displays the GeneLab Federated Search interface. At the top, the GeneLab logo and tagline 'Open Science for Exploration' are visible. The main search area shows a search for 'mouse myostatin' with filters for GeneLab, NIH GEO, EBI PRIDE, and ANL MG-RAST. Search results for 'mouse myostatin' are listed, including 'Myostatin inactivation effects on myogenesis in vitro and in vivo' and 'The transcriptomic signature of myostatin inhibitory influence on the differentiation of skeletal muscle C2C12 myotubes'. A 'Search Filters for GeneLab' panel is open, showing filters for Project Type (Ground, Spaceflight), Factors (Age, Anatomical Structure, Antibiotic Concentration, Atmospheric Pressure, Bed Rest, Bleomycin Treatment, cage), Organisms (Mus musculus, Mycobacterium marisnigellii, Oryzias latipes, Pantoea conspicua, Pseudomonas aeruginosa, Rattus norvegicus, Rhodospirillum rubrum, Saccharomyces cerevisiae, Staphylococcus aureus), and Assay Type (deletion pool profiling, DNA methylation profiling, environmental gene survey, genome sequencing, metabolite profiling, protein expression profiling, RNA methylation profiling, transcription profiling). The search results are sorted by Relevance, and 25 results are shown. A total of 3 search results are found for the query 'mouse' with filters 'Age' and 'cage'.

GLDS Phase 2 (Release 2.0) Customized NASA Collaborative Workspace



User Account Mgmt., Access Controls (e.g., Private, Shared, Public Folders)

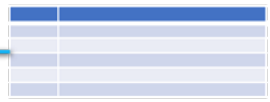
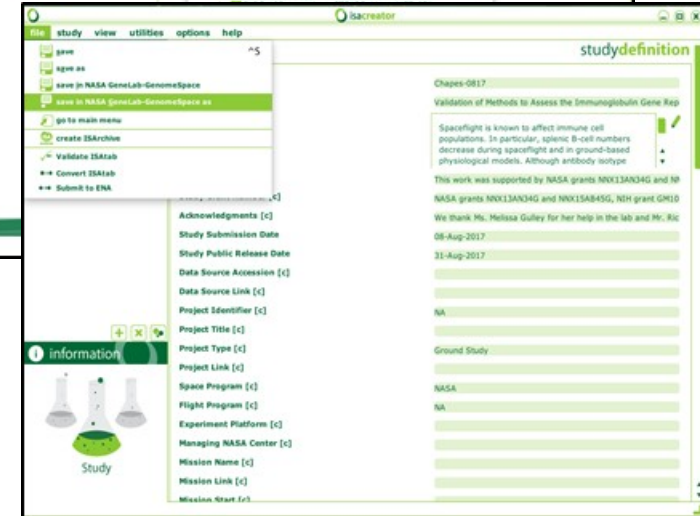
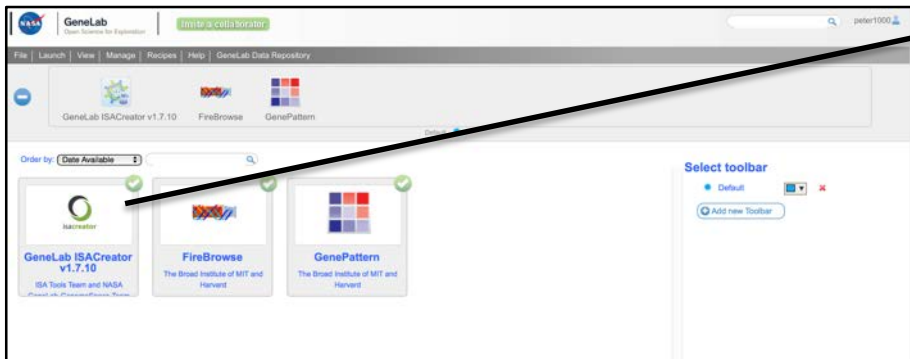
The image displays three overlapping screenshots of the GeneLab web interface:

- Top Left:** A search results page for "Dissecting Low Atmospheric Pressure Stress: Transcriptome Responses to the Components of Hypobaria in Arabidopsis [Experiment 2]". It shows a table with columns for Organisms, Factors, Assay Types, Release Date, and Description. The first entry is for Arabidopsis thaliana under Atmospheric Pressure, with a release date of 11-May-2017.
- Top Right:** A login page titled "NASA GeneLab-GenomeSpace OpenID Login". It includes fields for USERNAME and PASSWORD, "Sign In" and "Cancel" buttons, and a "Register new NASA GeneLab user" link.
- Bottom:** A file browser interface showing a directory structure. The path is "Home > Public > genelab > genelab-data". A table lists files with columns for Filename, Tags, Owner, Size, and Last Modified. The files are named GLDS-1 through GLDS-113, all owned by "genelab".

GLDS Phase 2 (Release 2.0) Metadata Curation via ISACreator Tool



GeneLab-GenomeSpace Integration with ISACreator for Streamlining Data Processing Operations



Metadata Source Mappings

Engaging the Scientific Community



GeneLab **Analysis Working Groups (AWG)** will be tasked with analyzing all data across the GLDS with relevance to a specific domain to generate higher-order data.

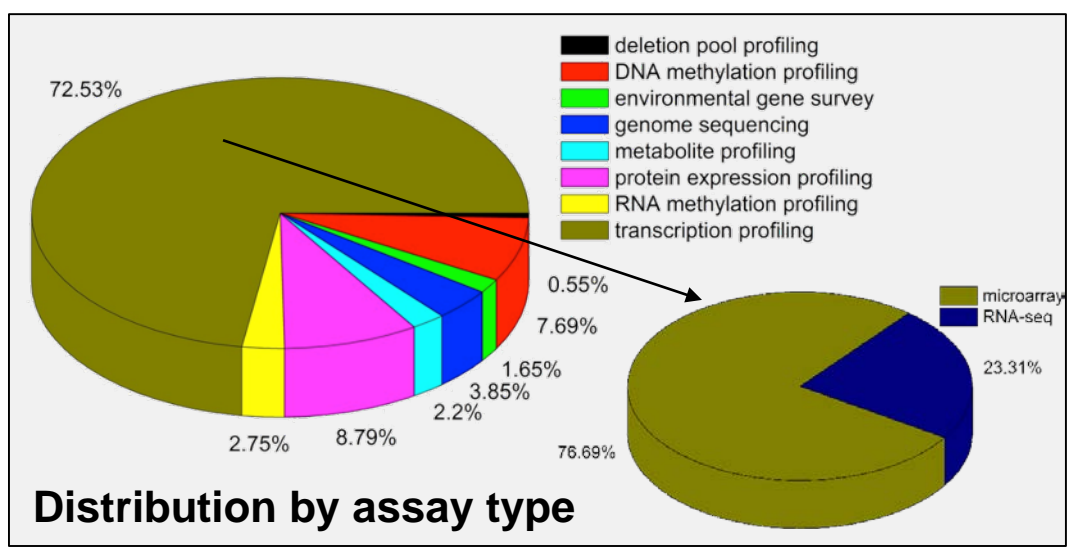
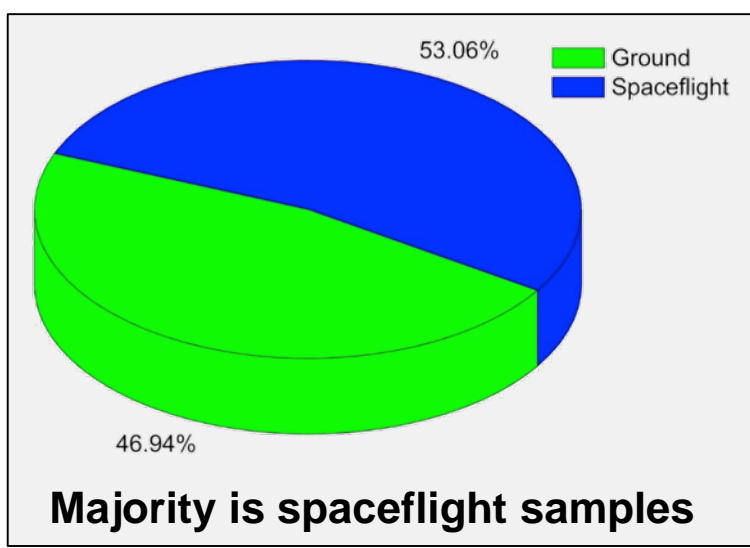
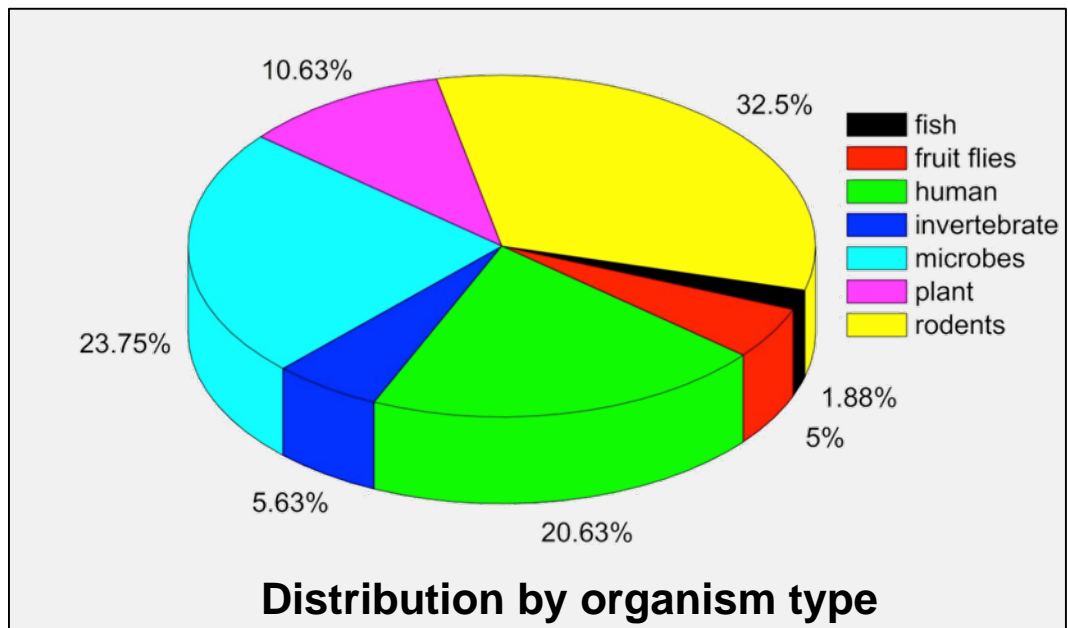
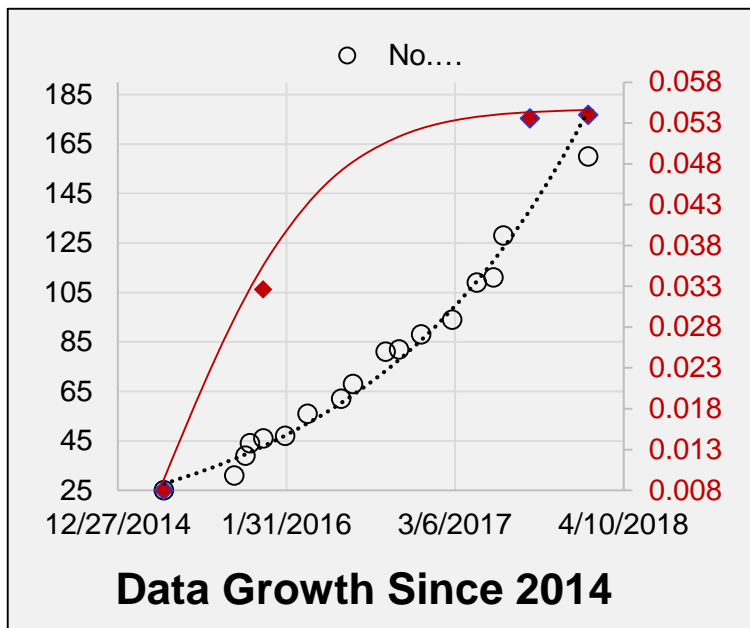
Goals:

1. Peer-reviewed publications describing AWG's comprehensive analysis.
2. Consensus data analysis pipelines relevant to AWG domains to be used on the GLDS will help domains harmonize their analyses.
 - a) Summer interns will process all data based on AWG recommendation
 - b) Processed "higher-order" data relevant to domains will be posted on the GLDS.
 - c) Strategies needed to link metadata to processed data will be put in place for the visualization portal deployment
3. Feedback for the GLDS to be used for improving its utility; test driving passed along to scientific community via the AWG
 - a) Access to galaxy toolshed and Jupyterlab GenePattern notebook within GeneLab provided with CPU and RAM AWS resources
 - b) Integration of GenomeSpace workspace with processing tools
 - c) GLDS 2.0 search query needs to be improved – What should we do different?

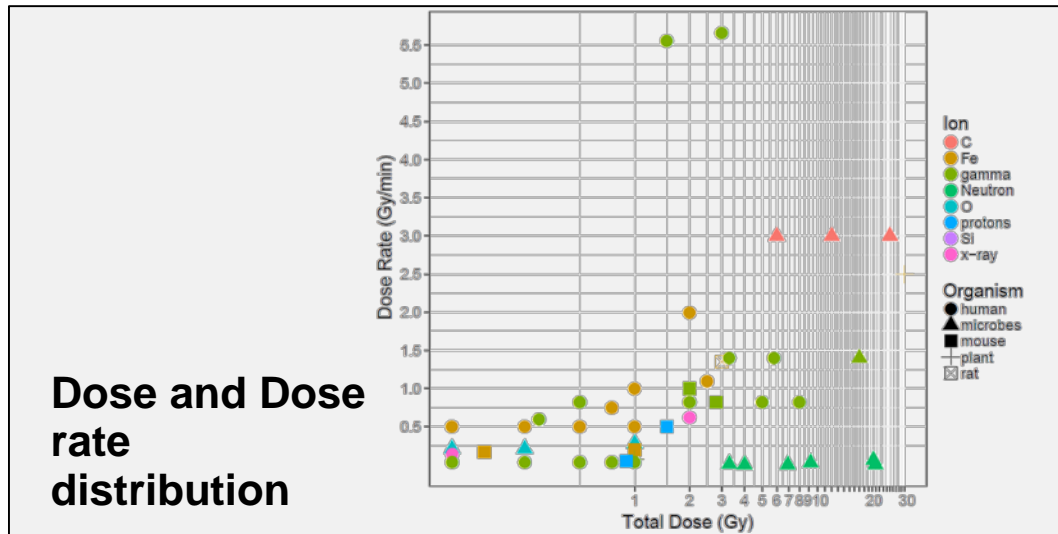
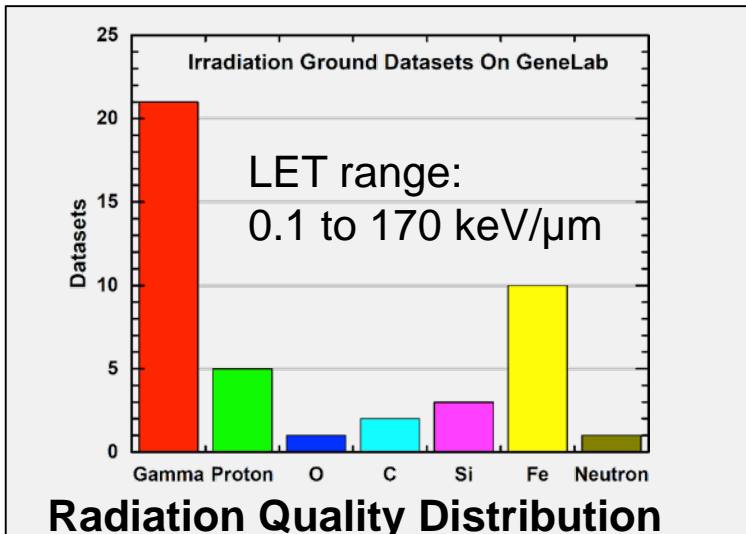
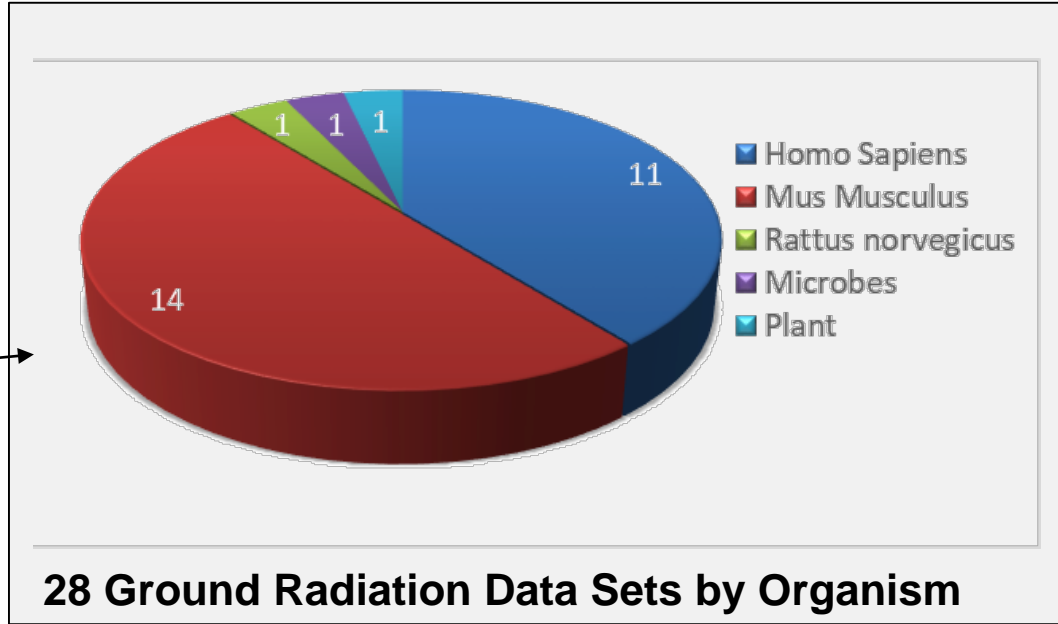
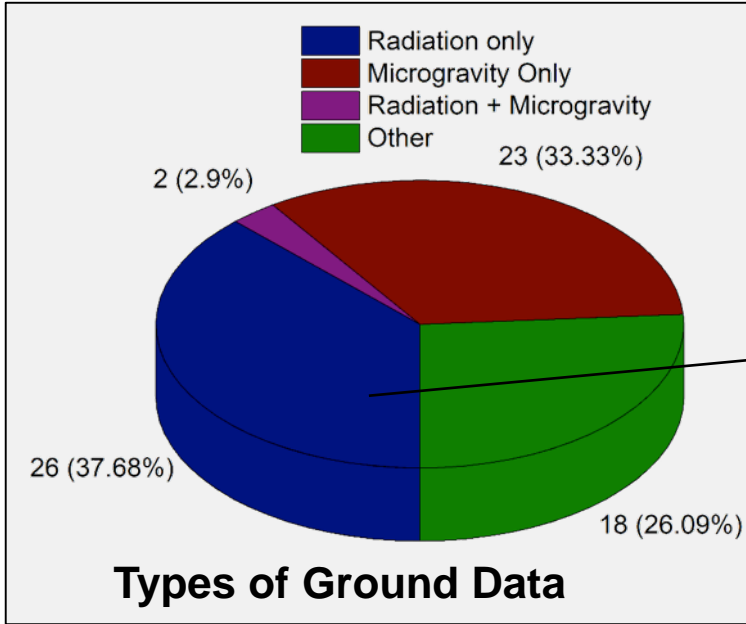
AWGs emphasis:

1. Animal Group
 - a) Mammals
 - b) Non-mammals
2. Plants
3. Microbes
4. Multi-omics/Systems Biology

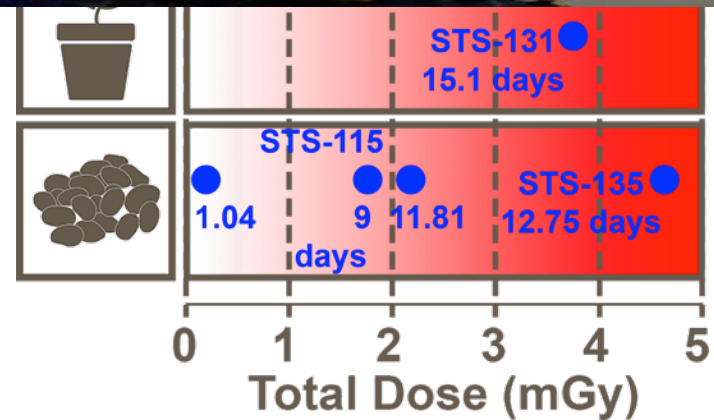
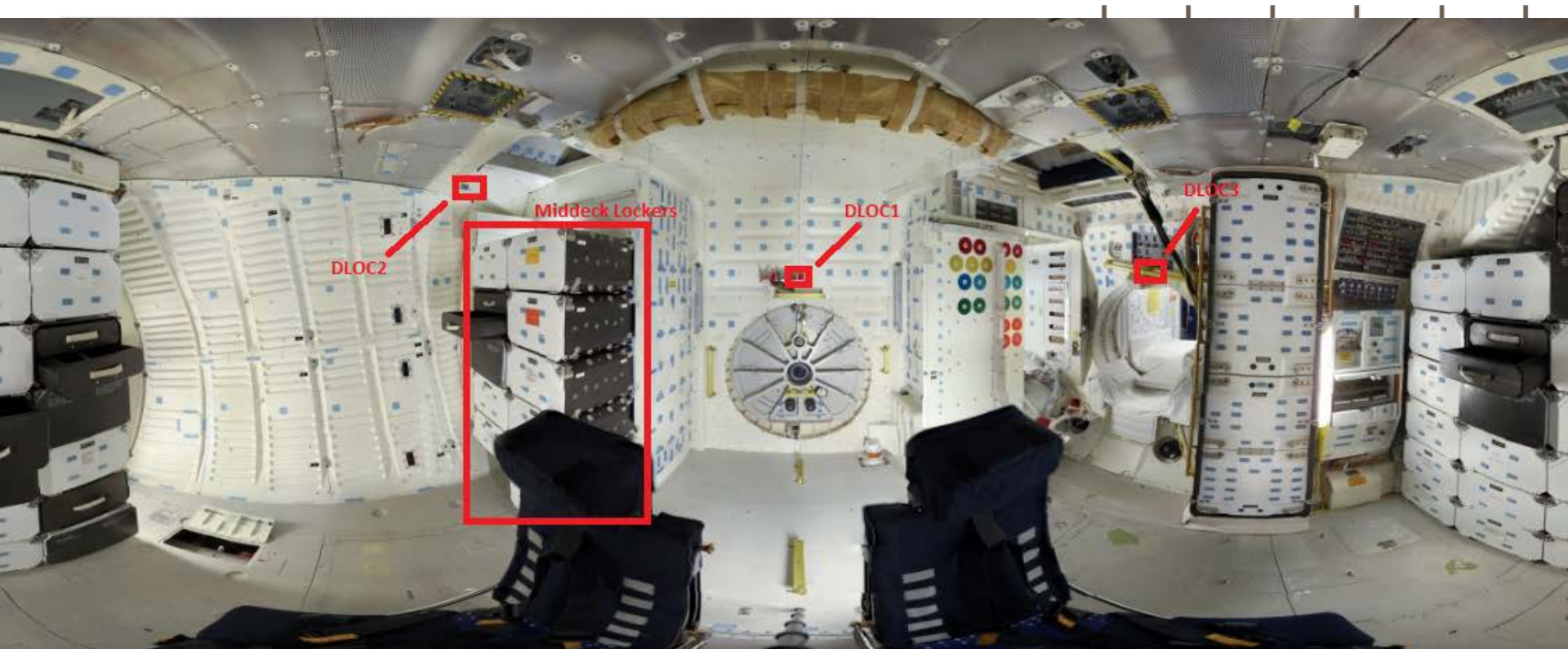
GeneLab Database: 154 data sets



69 Ground Data Sets



STS Samples: Radiation Dosimetry



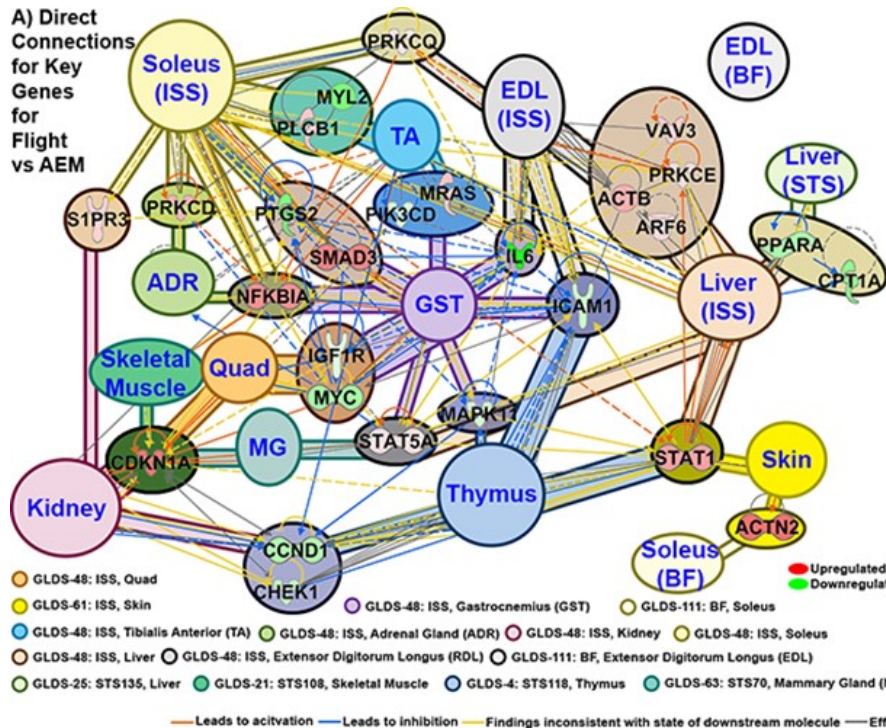


Building a database to support studies on human health and countermeasures

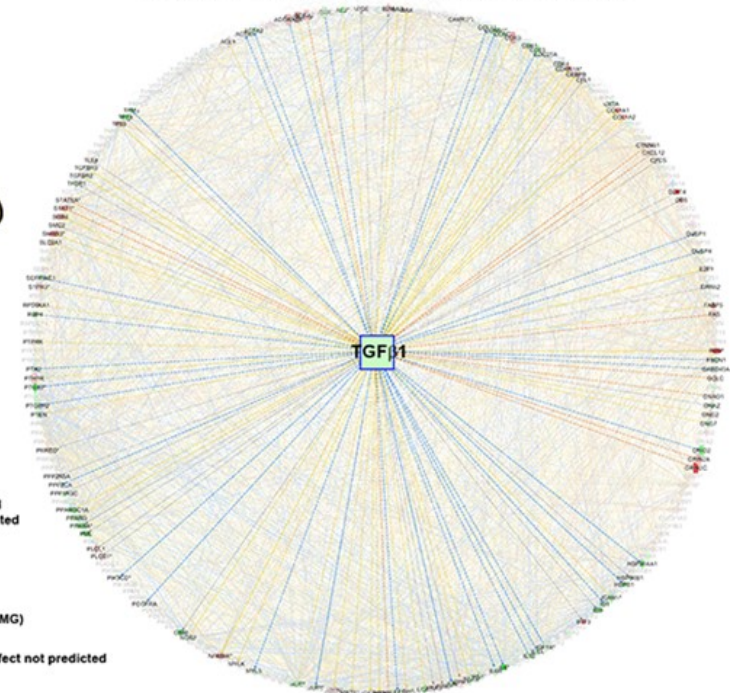
Future analysis capabilities

1. Cohort comparison

- Display the expression of a gene query or its frequency of differential regulation based on sex, species, tissue, or age
 - Example: From a systems biology analysis, TGFβ1 was found to be a master regulator impacting spaceflight



B) Connections Between all Key Genes for all Datasets (Flight vs AEM): Radial Plot with the most Connected Gene in the Middle



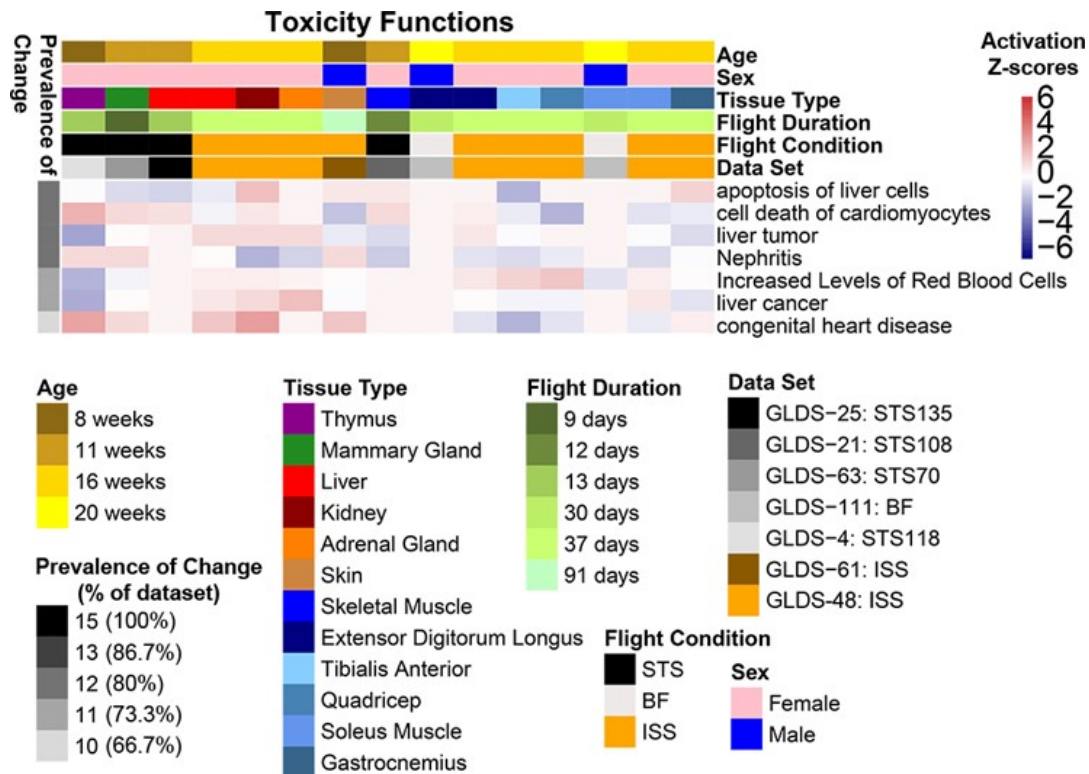


Building a database to support studies on human health and countermeasures

Future analysis capabilities

2. Relevance to human disease

- Display the expression of a query gene or its frequency of differential regulation in disease types
 - Example: Using the GeneLab data we are able to make predictions on impact on health and risk of diseases due to space flight



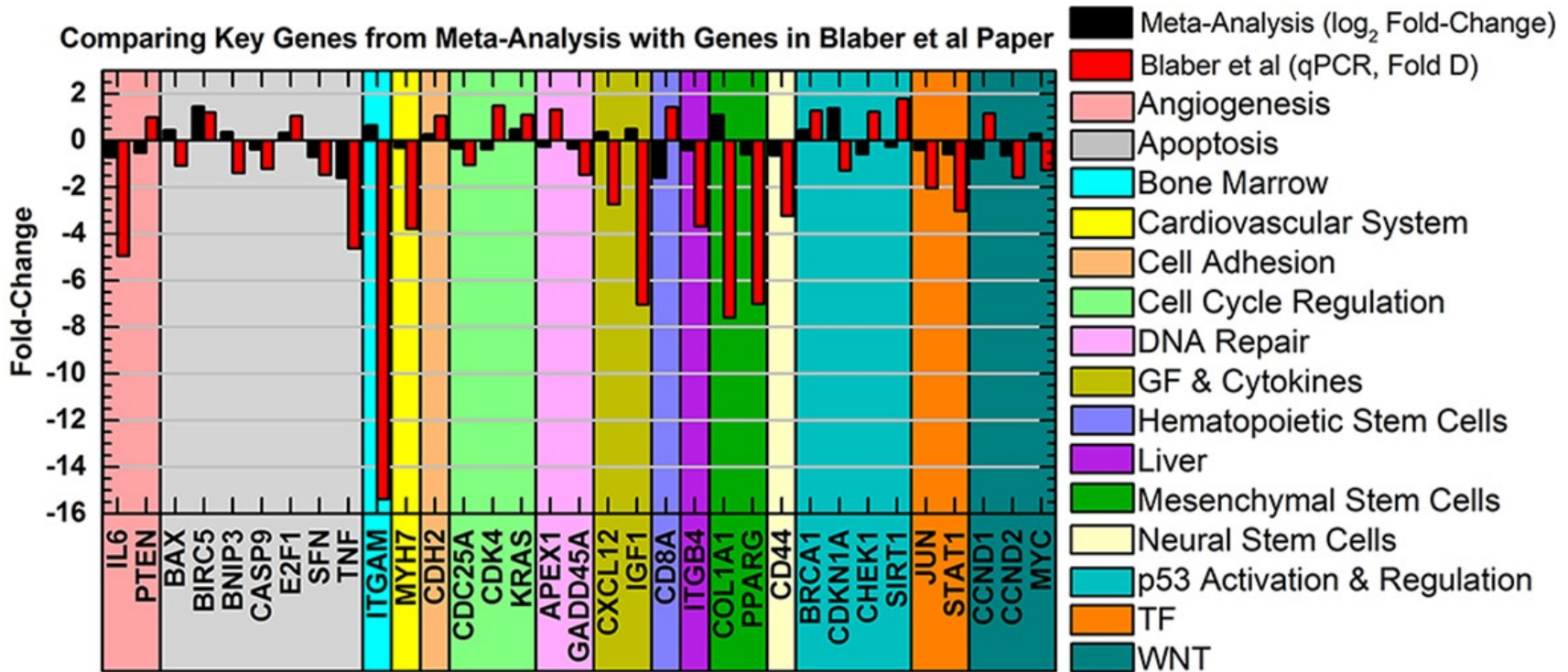
Building a database to support studies on human health and countermeasures



Future analysis capabilities

3. Tissue expression

- Display the expression of a query gene based on cell or tissue type
 - Example: Can make direct comparisons from of key genes to data from the literature.



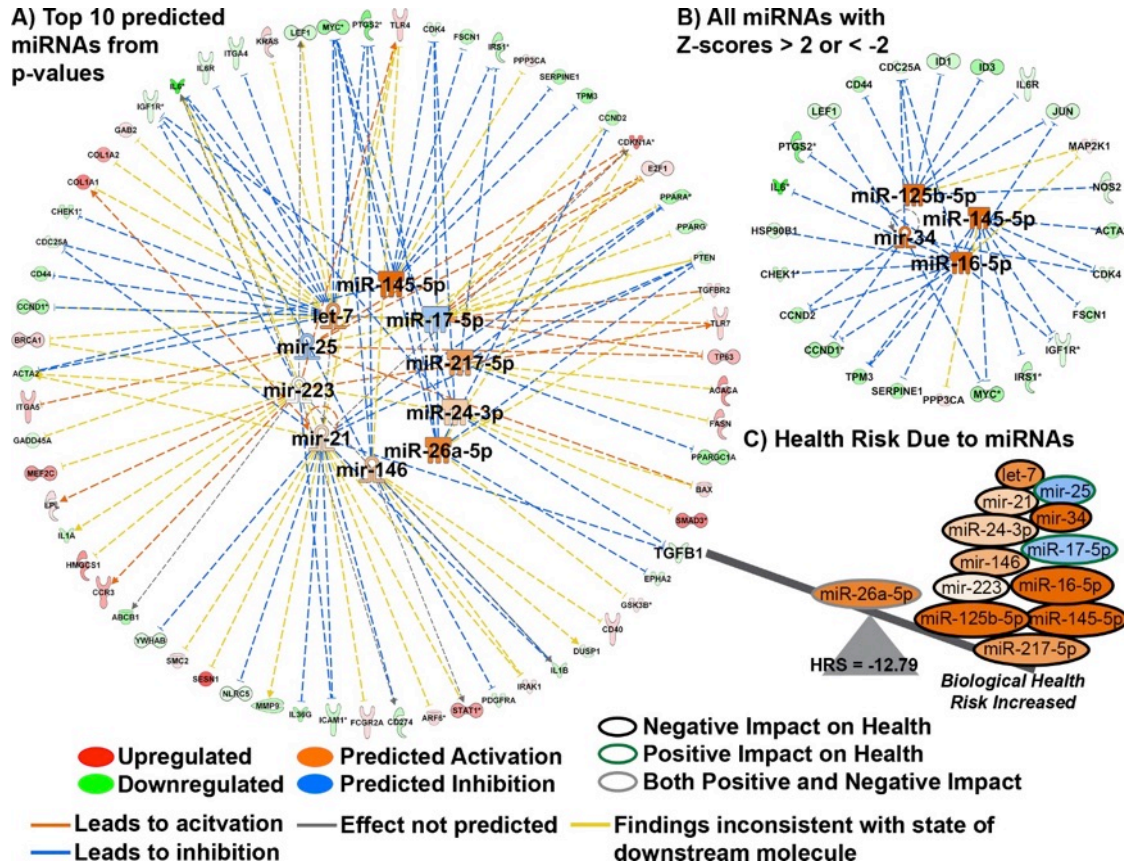


Building a database to support studies on human health and countermeasures

Future analysis capabilities

4. Countermeasure identification

- Display countermeasures reported to impact expression of a gene query
 - Example: Hypothesis generated from GeneLab datasets that miRNAs can be used as countermeasure against spaceflight health risks.

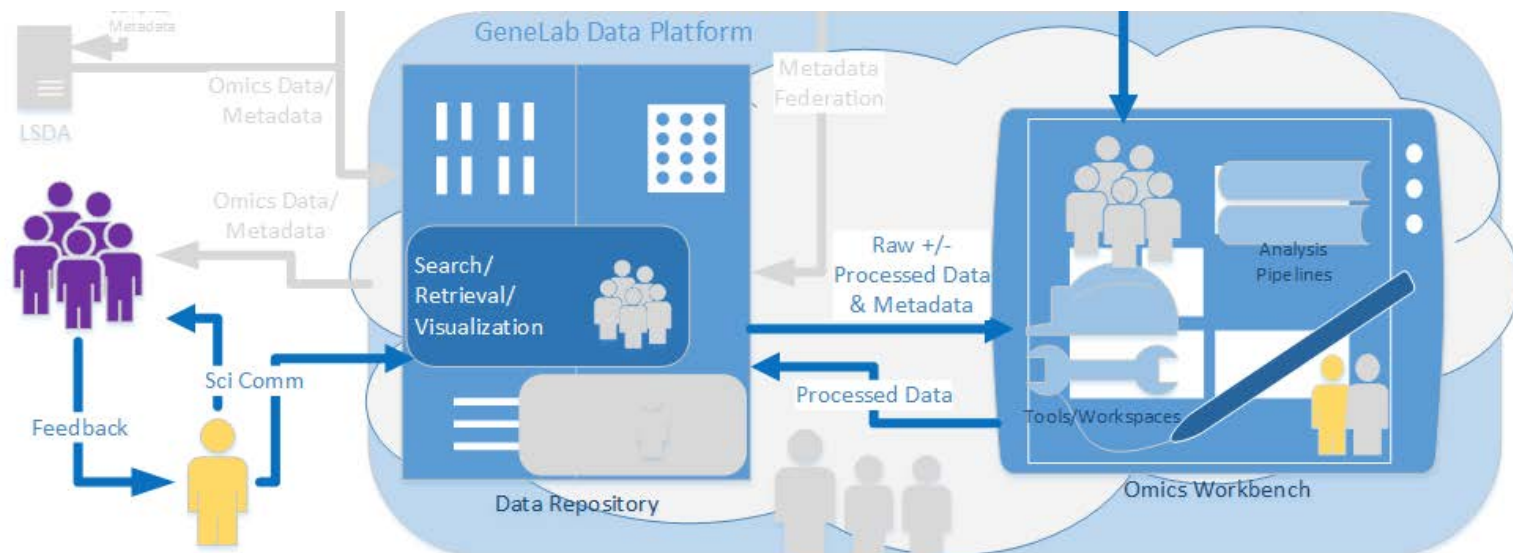


Interaction of GeneLab Science Team with Data System



Beta test GLDS

- Drive system for knowledge
 - Evaluate metadata sufficiency for depth (i.e, experimental design clarity)
 - **Generate critical higher order data in the process**
 - Establish processing pipeline
 - Explore visualization software for data, metadata and higher order data
- Identify tools for future GLDS incorporation
- Lead by example with case studies: Generate and publish higher order data
- Coordinate AWG



The Data Reproducibility Challenge



- Space omics datasets are sparse
 - Need to reduce level of noise
 - Need a method for assay bias identification and correction
- Started a collaboration with NIST (National Institute of Standards and Technology)
 - Implement methods to make the best use of precious flight samples
 - NIST showed high level of variation for RNAseq between 12 different core processing centers in the US



NIST's collaborators
Dr. Munro and Dr. Salit

The image shows a horizontal banner with the Nature Communications logo on the left, which consists of a stylized wave graphic in red and yellow above the text "nature COMMUNICATIONS". To the right of the logo, the text "NIST's collaborators" and "Dr. Munro and Dr. Salit" is displayed in a black, sans-serif font.

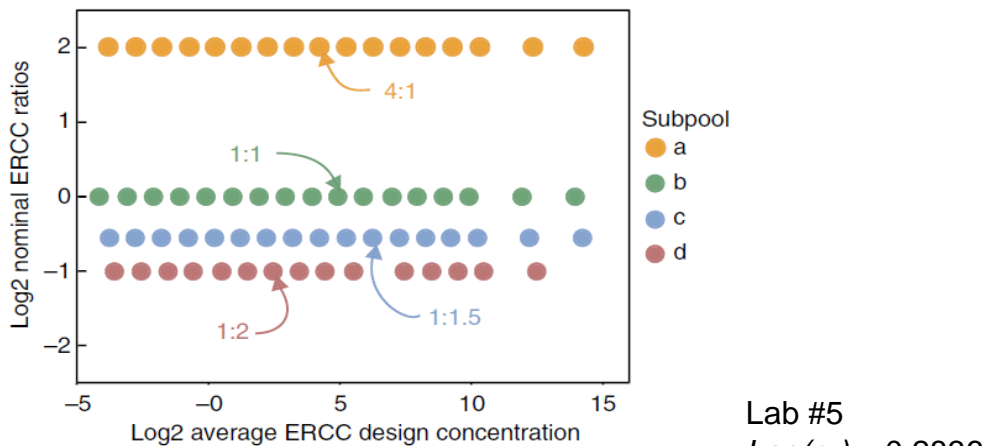
ARTICLE

Received 11 Aug 2014 | Accepted 1 Sep 2014 | Published 25 Sep 2014

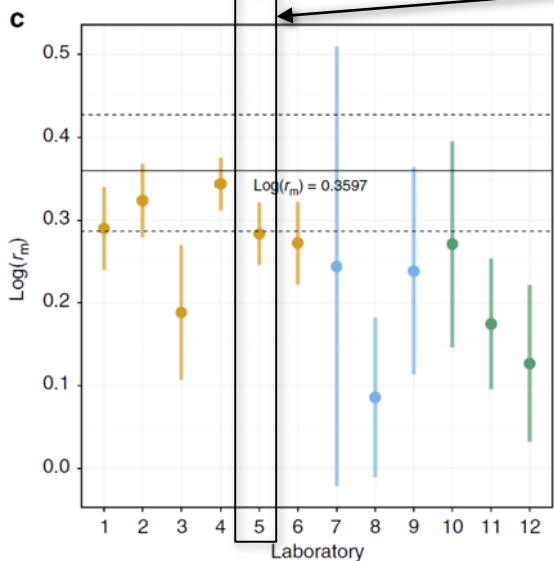
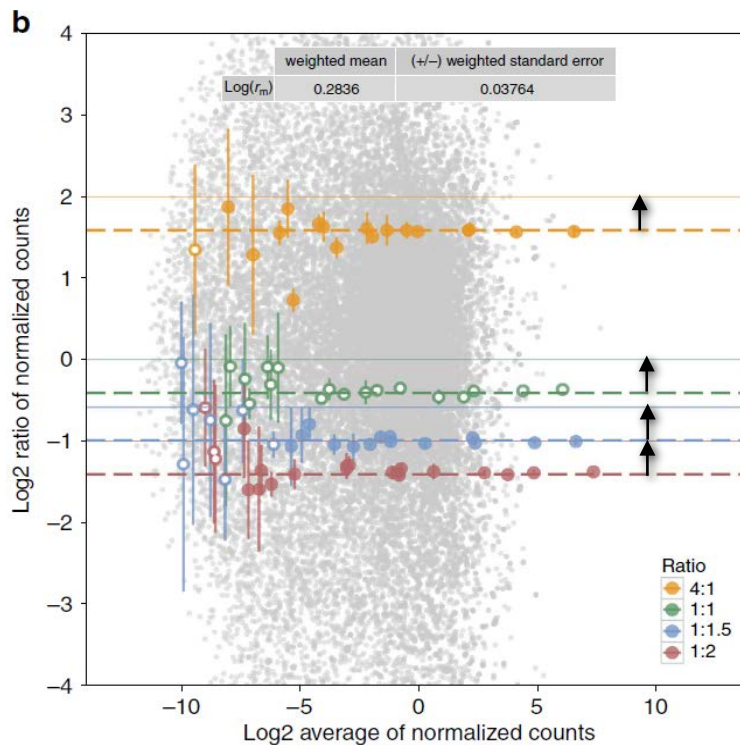
DOI: 10.1038/ncomms6125

Assessing technical performance in differential gene expression experiments with external spike-in RNA control ratio mixtures

#1 Risk: Data Reproducibility



Lab #5
 $\text{Log}(r_m) = 0.2836$



r_m = Weighted mean estimates of mRNA fraction differences for the sample set with error bars representing weighted standard errors

$$R_S = r_m \left(\frac{E_1}{E_2} \right)_S$$

$$\text{Log}(R_S) = \text{Log}(r_m) + \text{Log}(E_1) - \text{Log}(E_2)$$

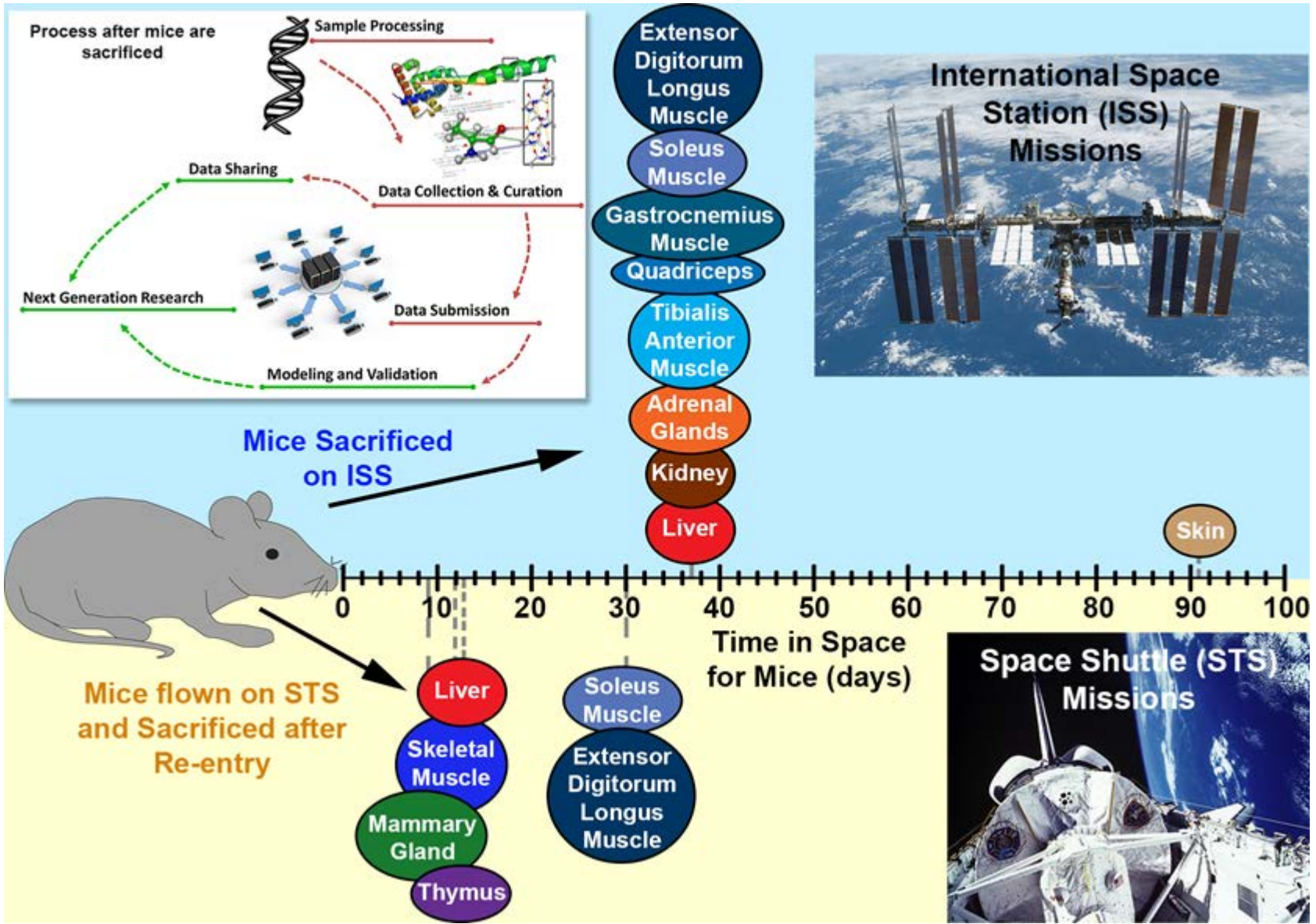
Sample Processing Laboratory (SPL)



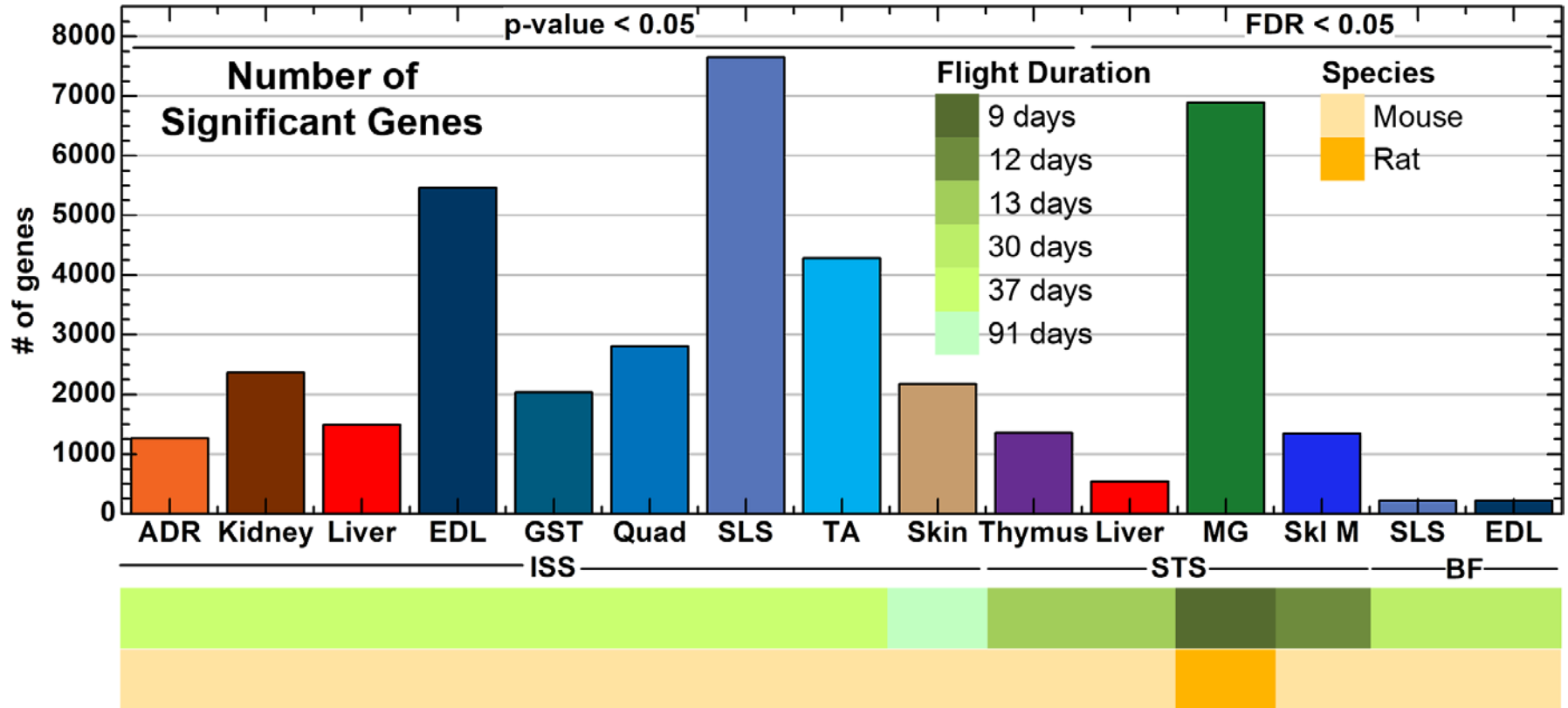
- Expertise:
 - DNA/RNA/protein extraction
 - Cell culture
 - Animal work
- Develop standards for sample processing (species dependent)



General Overview of GeneLab Mice Data



Number of Significant Genes from Multiple Datasets



Predicted Master Regulators



Metabolic and Immune System related functions are regulated for in most tissues due to microgravity

This dysregulation in overall immune and metabolic functions can cause an impact on health of astronauts due to microgravity

Signaling by Interleukins
Immune System
Signal Transduction
TGF β Signaling

Immune System
Glycogen Catabolic Process

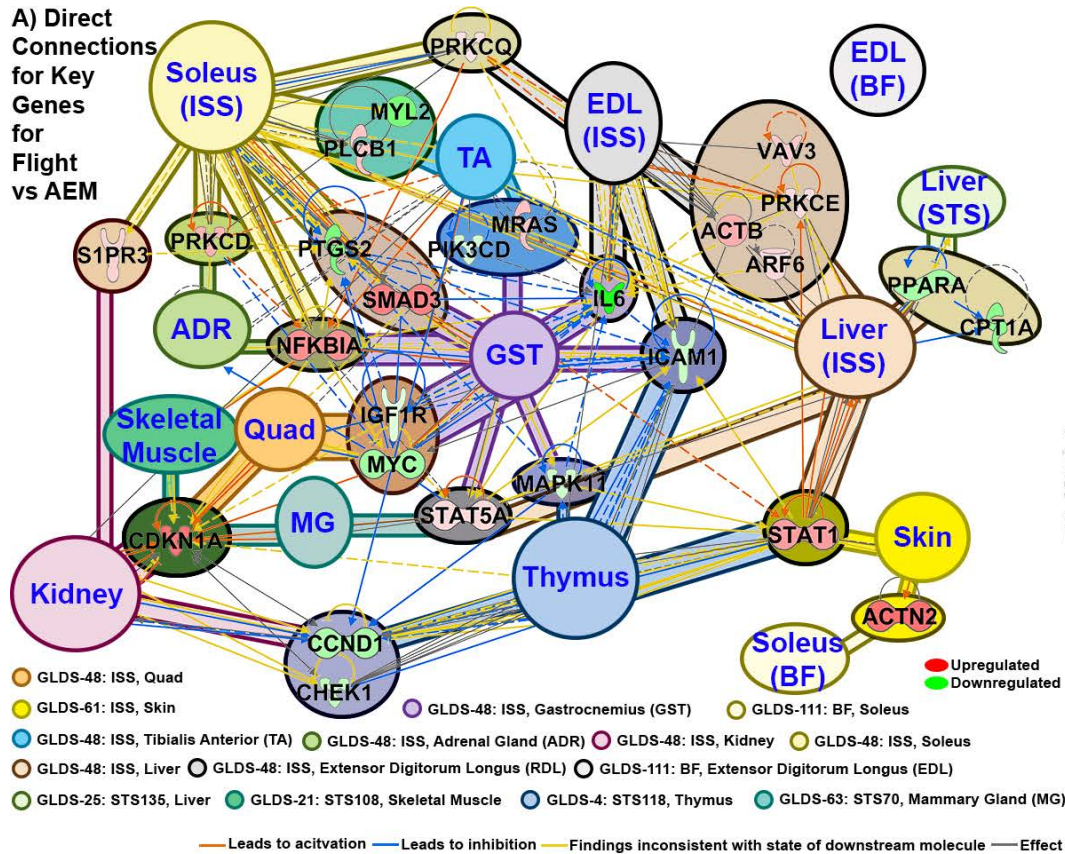
TP53

- **p53** common in all tissue and conditions when comparing tissue from mice Flight vs Controls.
- **p53 known to be involved in:** tumor suppressor, conserving stability by preventing genome mutation, DNA repair, cell cycle, apoptosis

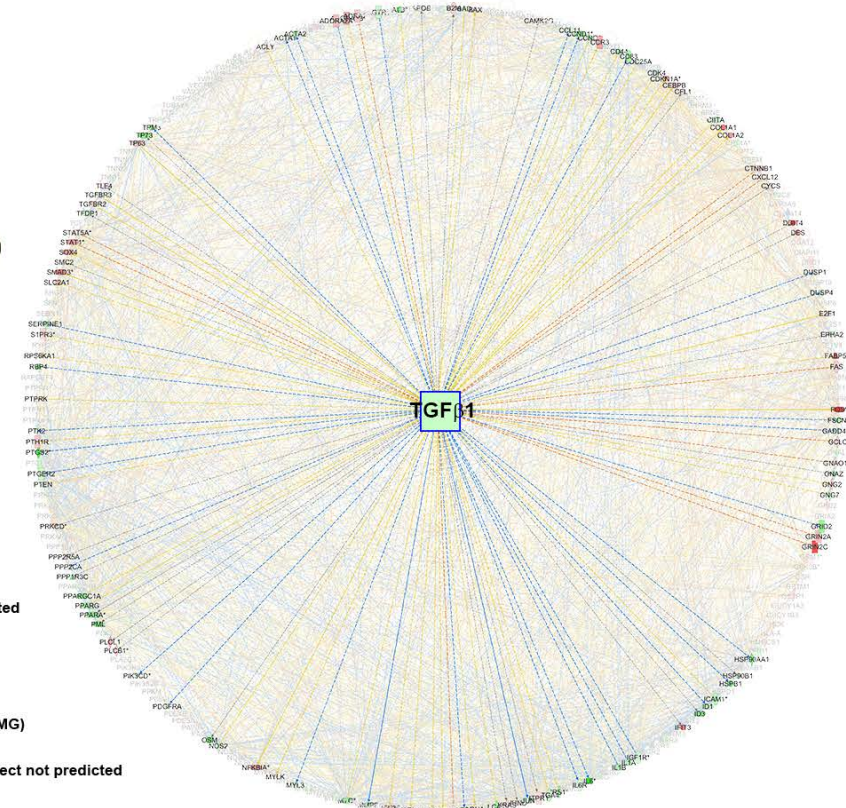
Key Genes and the Connections: Flight vs Ground (AEM – Rodent Habitat)



A) Direct Connections for Key Genes for Flight vs AEM



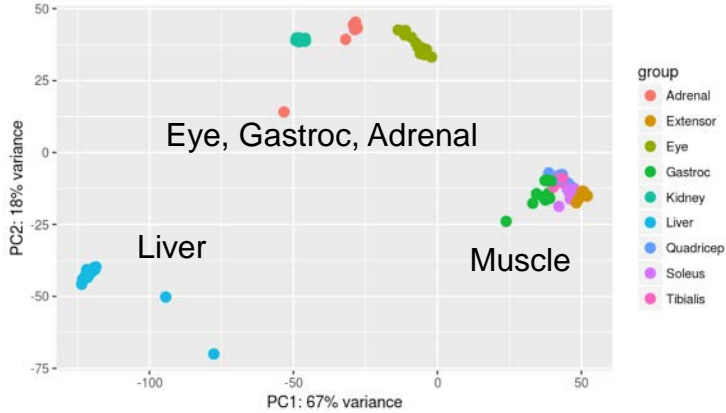
B) Connections Between all Key Genes for all Datasets (Flight vs AEM): Radial Plot with the most Connected Gene in the Middle



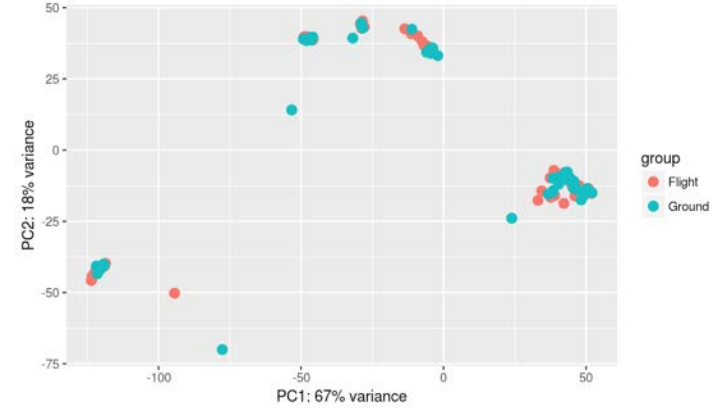
Mission-specific analysis: RR-1 Transcriptomics



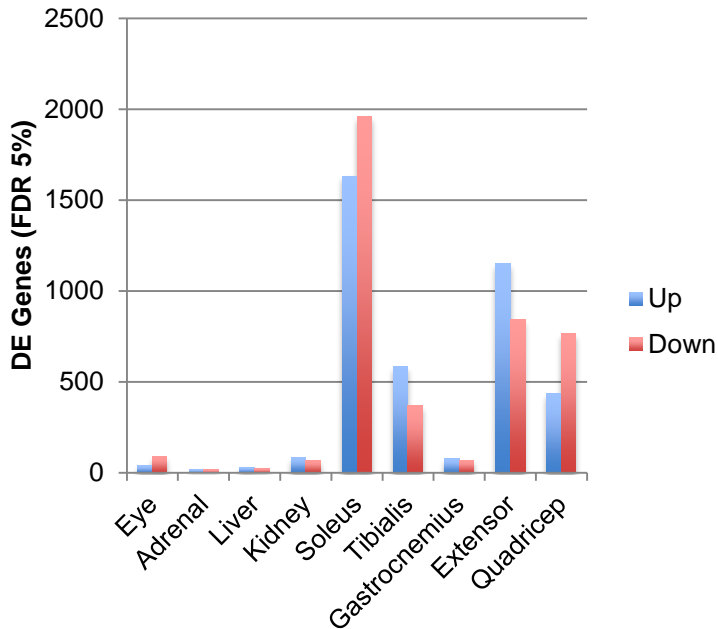
Samples cluster by tissue type



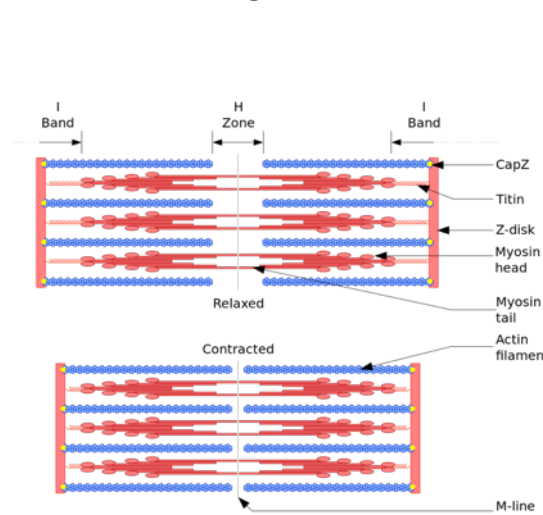
Samples do not cluster by flight/ground



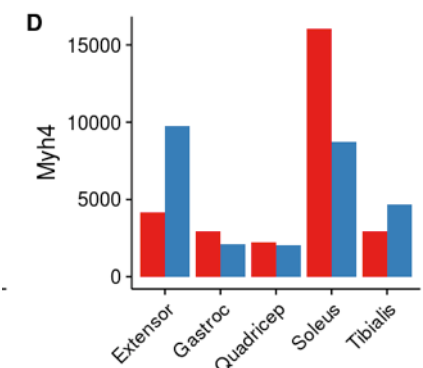
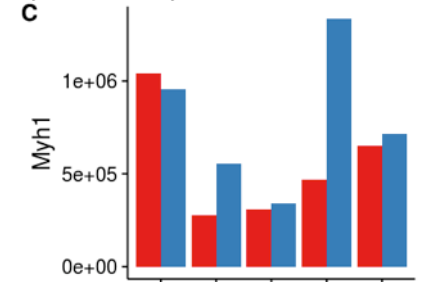
Many changes in muscles



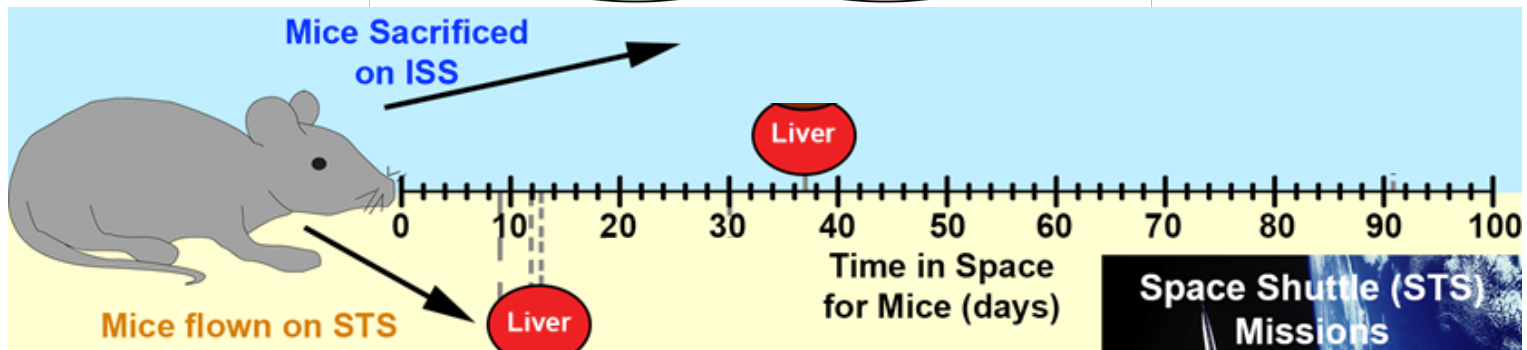
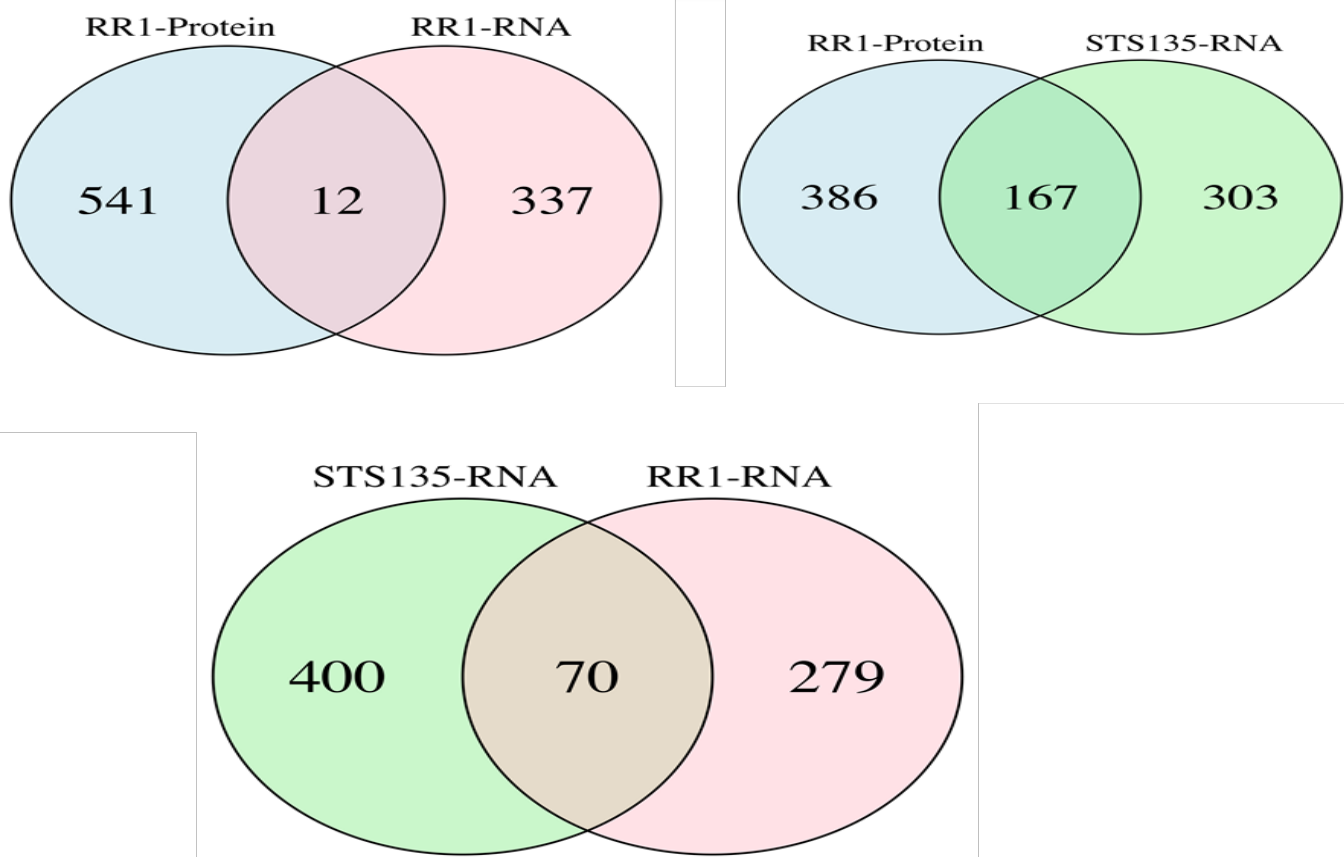
Changes to muscle myosin types



Muscle structure



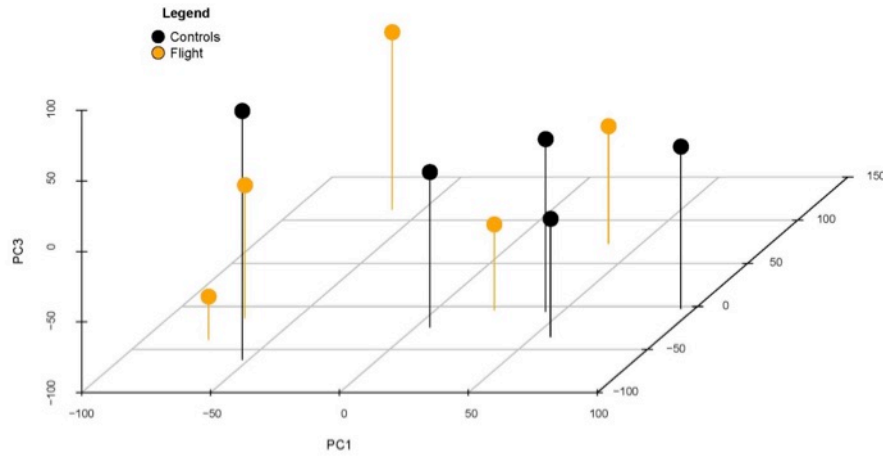
Impact of Microgravity on Liver Tissue: STS135 & RR1 Intersect Venn Diagram Analysis



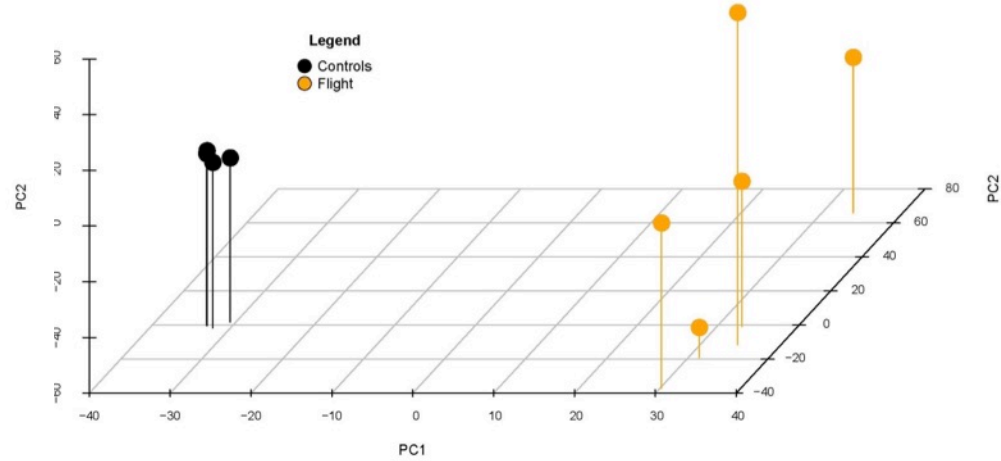
Impact of Microgravity on Liver Tissue: STS135 & RR1 Principle Component Analysis



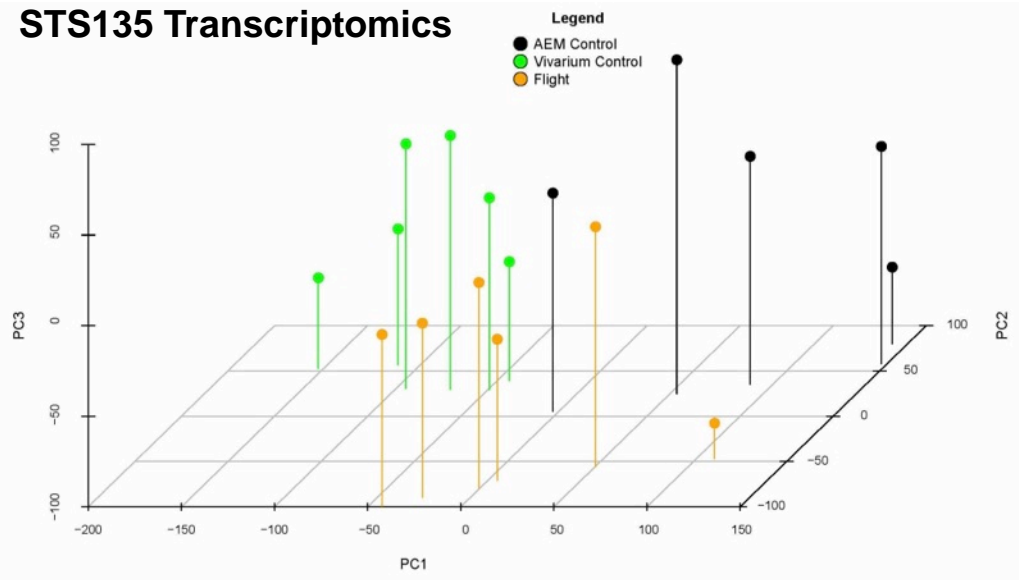
RR1 Transcriptomics



RR1 Proteomics



STS135 Transcriptomics

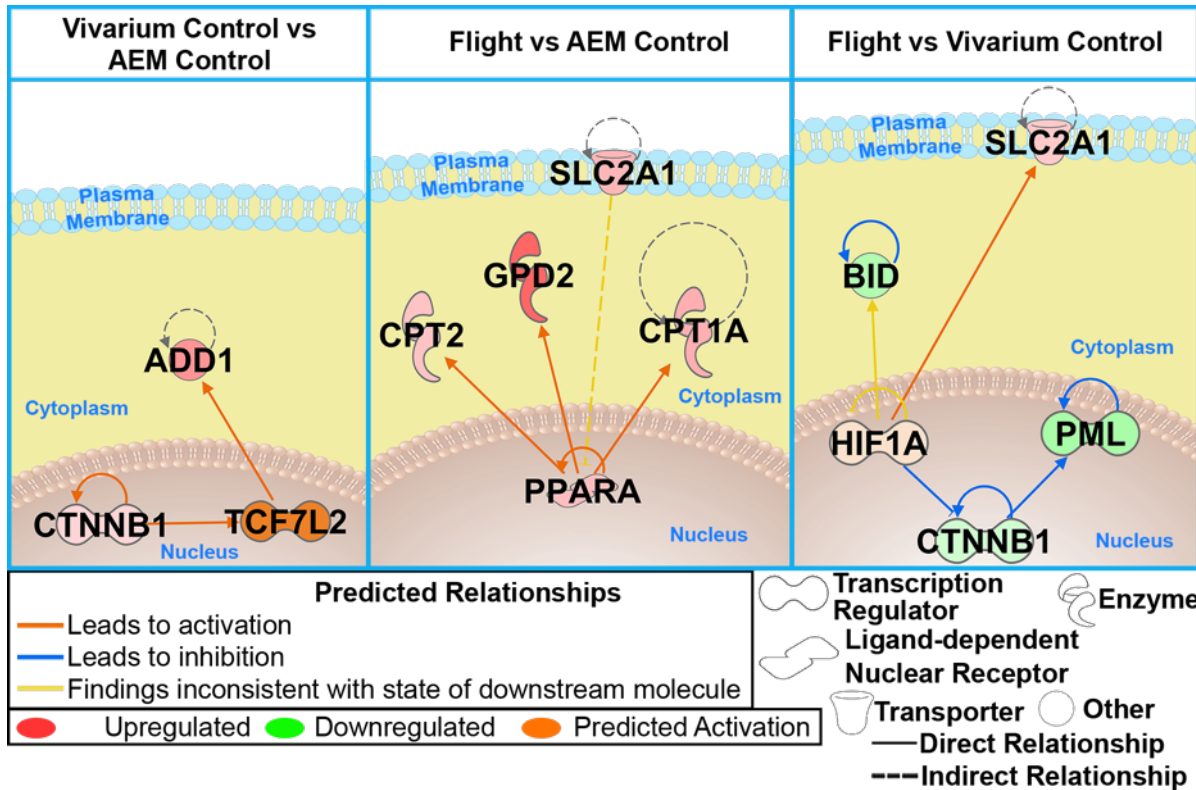


RR1 & STS135 Mice Liver KEGG Enrichment Pathways

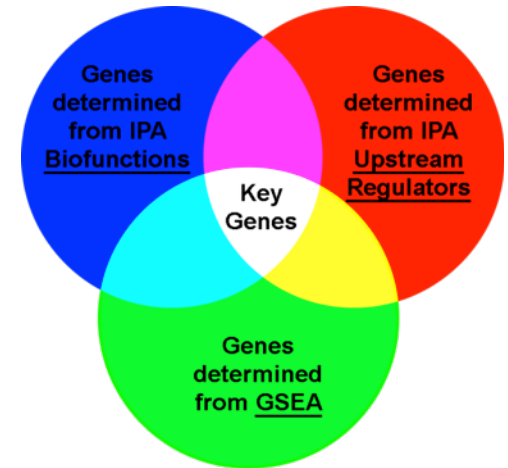


Pathway	STS-135 RNA	RR1 Protein	RR1 RNA
mmu01100:Metabolic pathways	****	****	****
mmu01200:Carbon metabolism	****	****	**
mmu01130:Biosynthesis of antibiotics	****	****	**
mmu01212:Fatty acid metabolism	****	****	**
mmu00640:Propanoate metabolism	****	****	*
mmu00062:Fatty acid elongation	***	*	*
mmu00620:Pyruvate metabolism	**	*	*
mmu00380:Tryptophan metabolism	**	***	**
mmu00520:Amino sugar and nucleotide sugar metabolism	*	*	*
mmu00190:Oxidative phosphorylation	****	**	*
mmu00280:Valine, leucine and isoleucine degradation	****	****	NS
mmu04146:Peroxisome	****	**	NS
mmu04141:Protein processing in endoplasmic reticulum	****	*	NS
mmu00020:Citrate cycle (TCA cycle)	****	****	NS
mmu03013:RNA transport	****	**	NS
mmu03010:Ribosome	****	****	NS
mmu00071:Fatty acid degradation	****	****	NS
mmu00650:Butanoate metabolism	****	****	NS
mmu01210:2-Oxocarboxylic acid metabolism	***	***	NS
mmu00630:Glyoxylate and dicarboxylate metabolism	***	****	NS
mmu01230:Biosynthesis of amino acids	**	***	NS
mmu00970:Aminoacyl-tRNA biosynthesis	**	**	NS
mmu05010:Alzheimer's disease	**	**	NS
mmu00310:Lysine degradation	**	**	NS
mmu05012:Parkinson's disease	**	**	NS
mmu03050:Proteasome	**	**	NS
mmu00410:beta-Alanine metabolism	**	**	NS
mmu00920:Sulfur metabolism	**	**	NS
mmu00270:Cysteine and methionine metabolism	**	*	NS
mmu00010:Glycolysis / Gluconeogenesis	**	*	NS
mmu05016:Huntington's disease	*	**	NS
mmu00072:Synthesis and degradation of ketone bodies	*	**	NS
mmu00250:Alanine, aspartate and glutamate metabolism	*	**	NS
mmu00860:Porphyrin and chlorophyll metabolism	*	*	NS
mmu04932:Non-alcoholic fatty liver disease (NAFLD)	*	*	NS
mmu01040:Biosynthesis of unsaturated fatty acids	**	NS	*
mmu04922:Glucagon signaling pathway	**	NS	*
mmu00061:Fatty acid biosynthesis	**	NS	*
mmu04710:Circadian rhythm	*	NS	*

Key Genes Affected by Microgravity in Liver

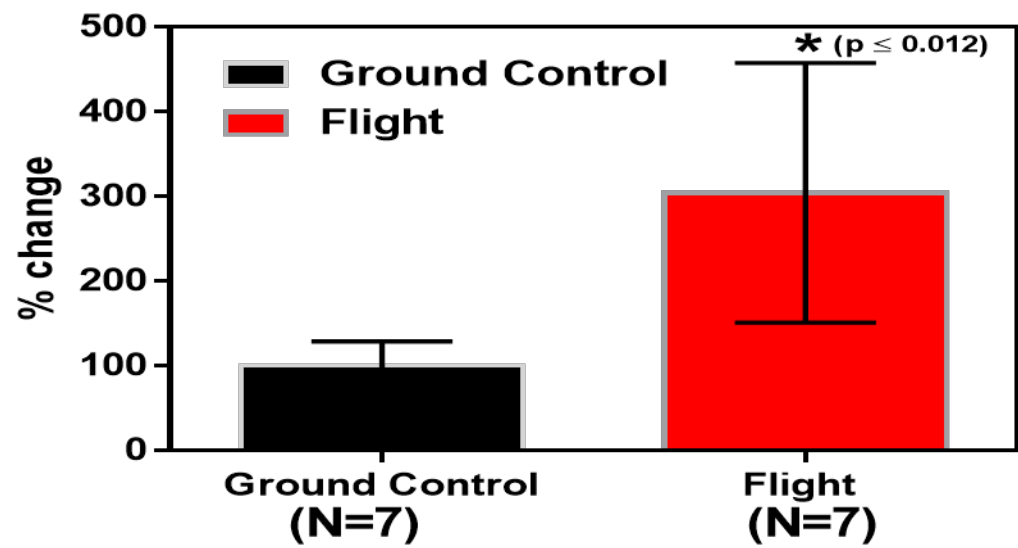
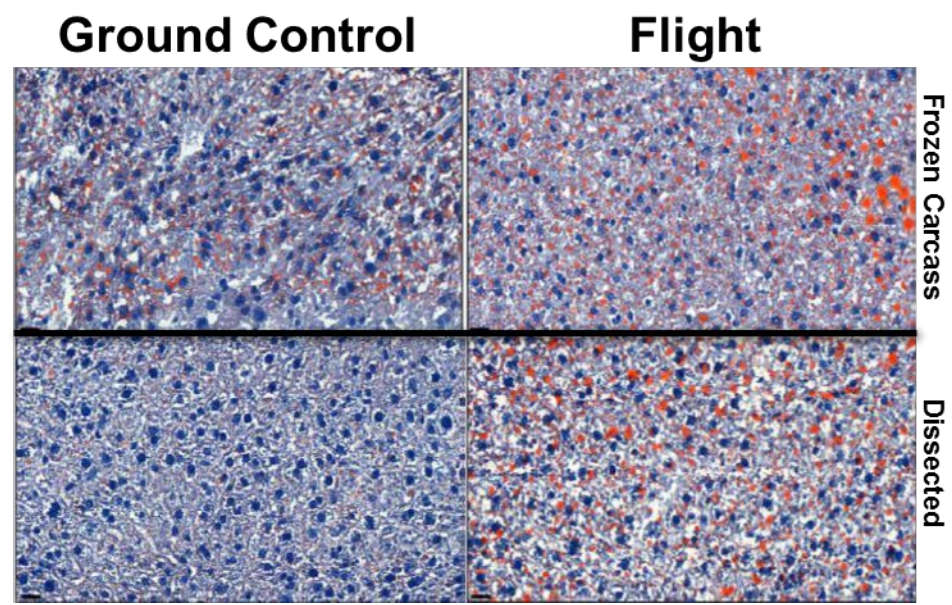


Key genes determined by the following:



- Common theme shows **PPARA** being putative key regulator in the liver
- Disruption of PPARA pathways is typically a precursor to liver disease
- Leads to hypothesis generation of possible mechanism occurring in the liver that is impacted by space radiation and microgravity.

Histopathology Confirms Liver Disease



Confounding Factor 1: Cage Effects

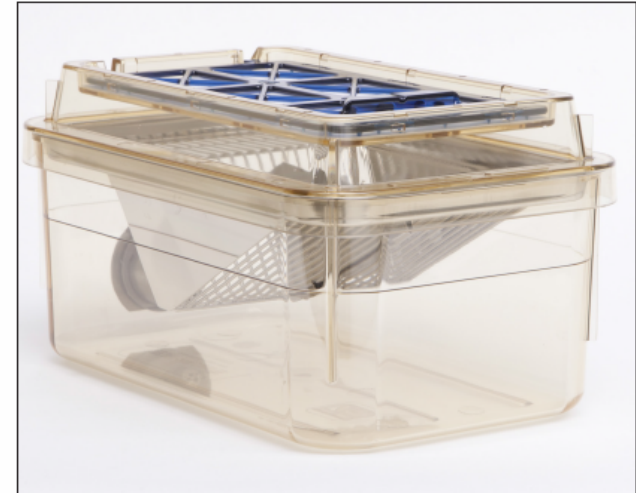
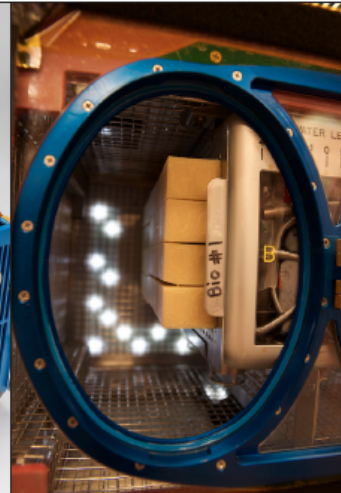


Vivarium vs Rodent Habitat control (AEM) across 5 different rat/mice studies, (no flight samples – CO2 level matches flight info)

Cage Types

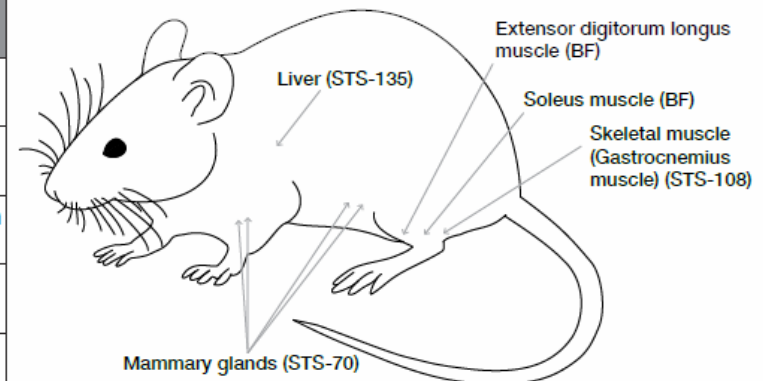


Animal Enclosure Module (AEM)



Sample vivarium cage

GeneLab study	Mission	Species	CO ₂ (ppm)	Tissue type
GLDS-21	STS-108	mouse	~3000	skeletal muscle (gastrocnemius)
GLDS-111	BF	mouse	~600	soleus muscle
GLDS-111	BF	mouse	~600	extensor digitorum longus muscle
GLDS-25	STS-135	mouse	~3000	liver
GLDS-63	STS-70	rat	~3000 (est.)	mammary gland



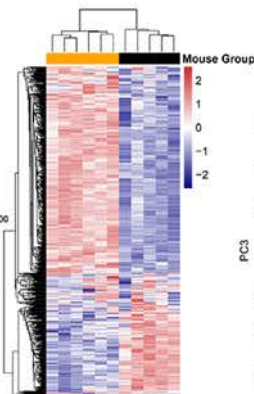
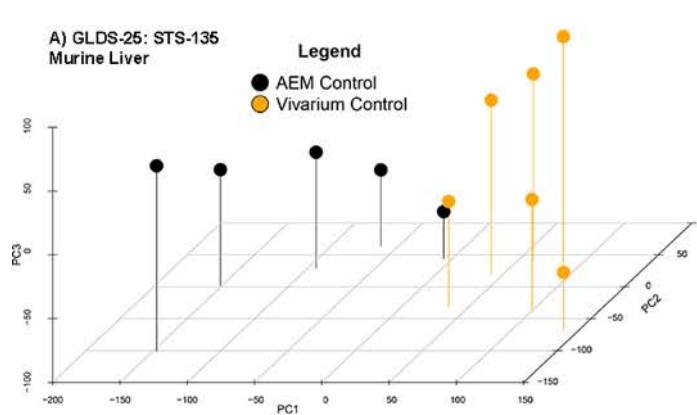
PCA Plots Suggest Strong Cage Effect



A) GLDS-25: STS-135
Murine Liver

Legend

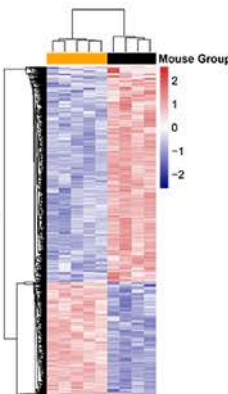
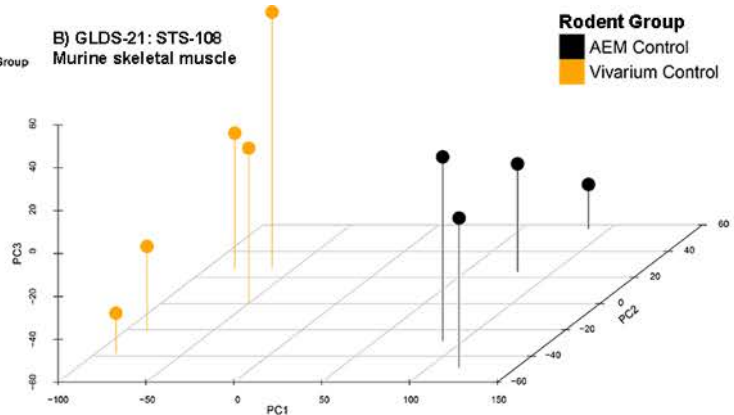
- AEM Control
- Vivarium Control



B) GLDS-21: STS-108
Murine skeletal muscle

Rodent Group

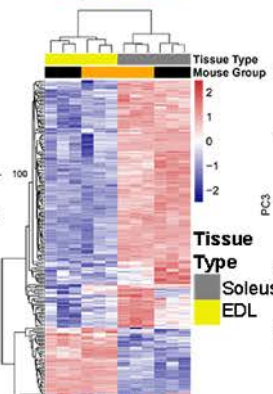
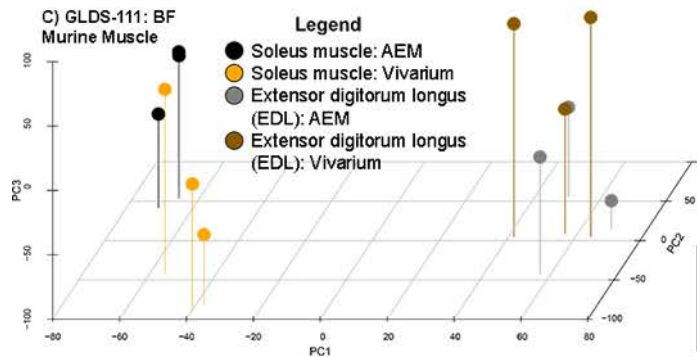
- AEM Control
- Vivarium Control



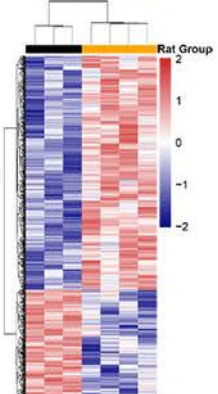
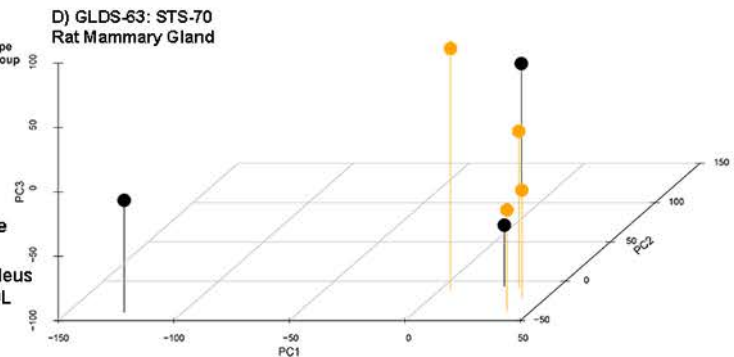
C) GLDS-111: BF
Murine Muscle

Legend

- Soleus muscle: AEM
- Soleus muscle: Vivarium
- Extensor digitorum longus (EDL): AEM
- Extensor digitorum longus (EDL): Vivarium



D) GLDS-63: STS-70
Rat Mammary Gland



Differential Gene Expression: Cage or CO2 Effect?



A) Venn Diagram of all significant genes

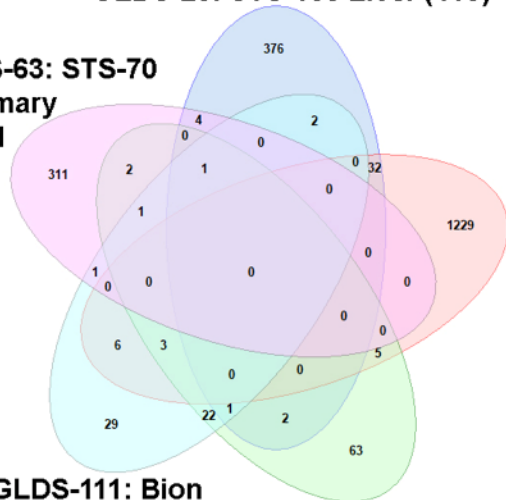
GLDS-25: STS-135 Liver (418)

GLDS-63: STS-70
Mammary
Gland
(348)

GLDS-21: STS-108
Skeletal Muscle
(1303)

GLDS-111: Bion
Extensor Digitorum
Longus (66)

GLDS-111: Bion
Soleus Muscle
(100)



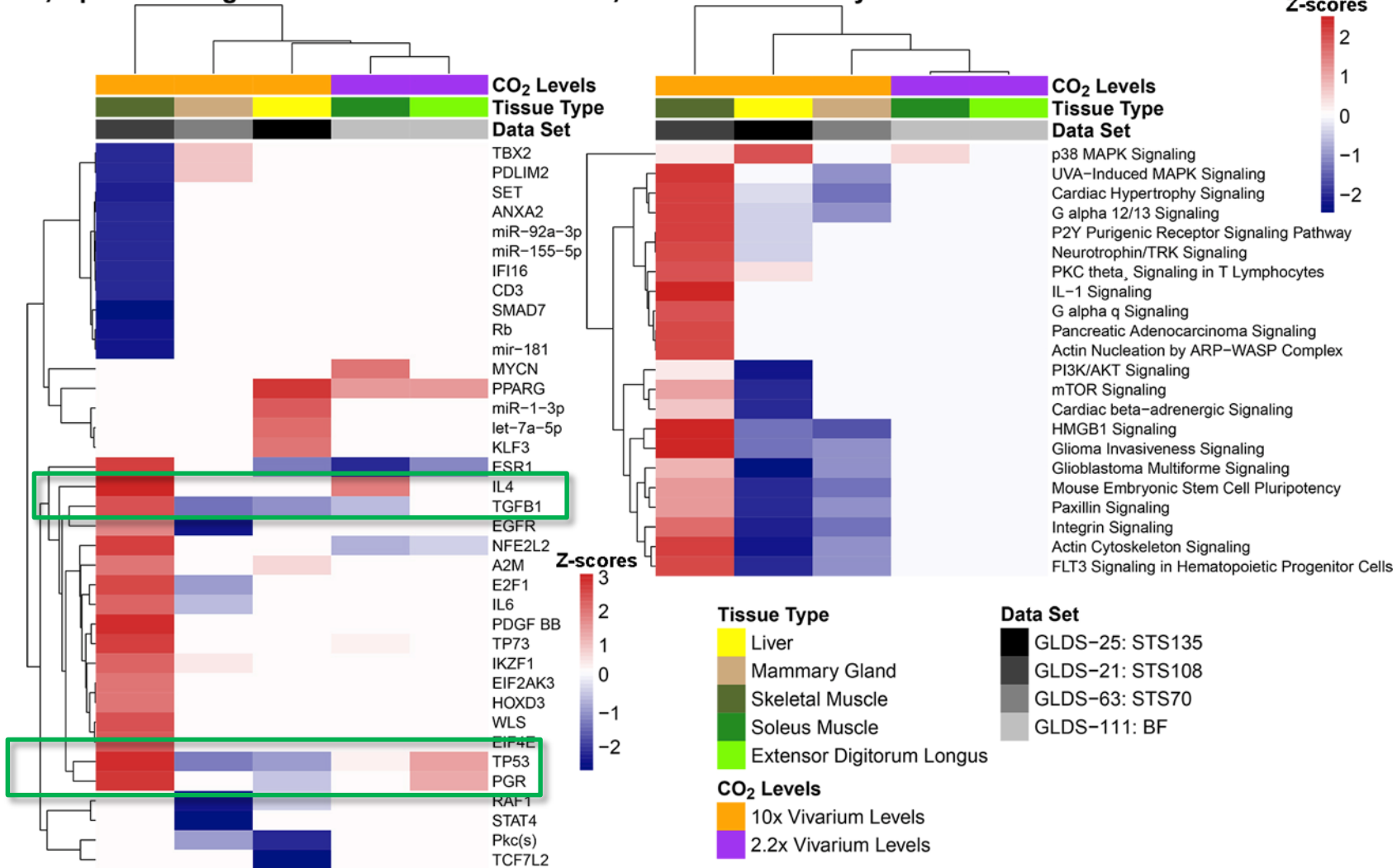
An increase in aldosterone is associated with metabolic syndrome, which is characterized by chronic inflammation; aldosterone secretion can be triggered by hypoxia.

Upstream regulators and canonical pathways show response is tissue specific and highest for high CO₂



A) Upstream Regulators: AEM vs Vivarium

B) Canonical Pathways: AEM vs Vivarium

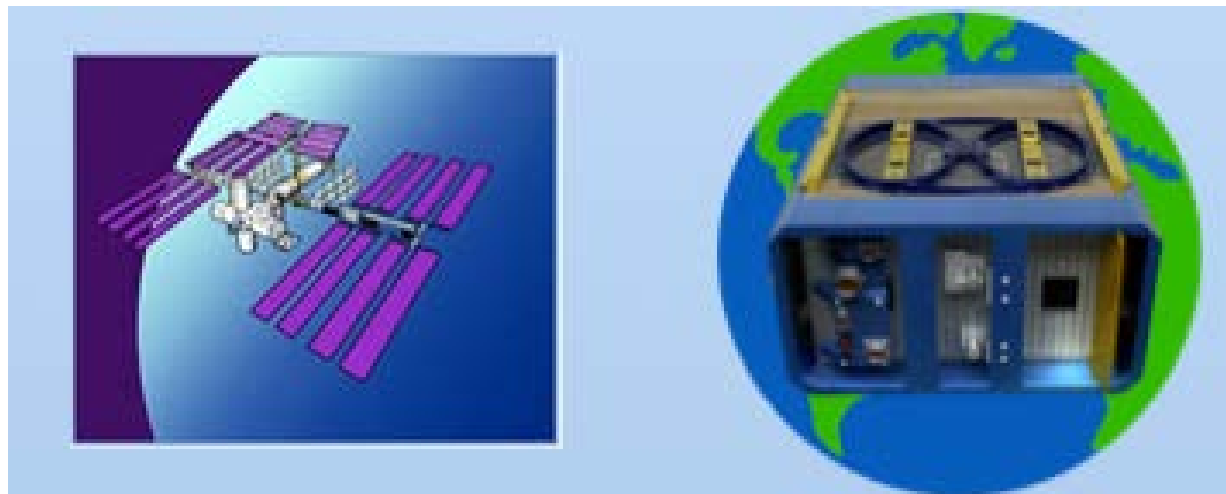


Mild chronic hypoxia due to increased CO₂ levels could explain both the increase in immune responses and a reduction in metabolism – **Need to confirm with AEM experiments at ambient CO₂ levels.**

Confounding Factor 2: Preservation Methods in Space



Liver collection for RR1



N=2

N=5

N=2

N=5

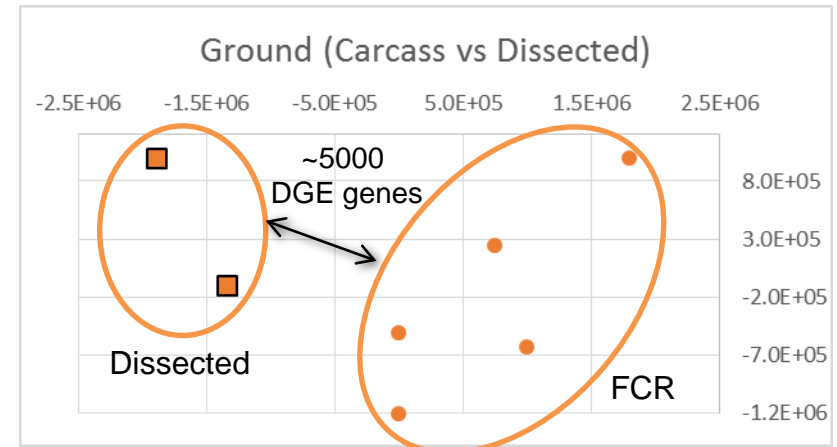
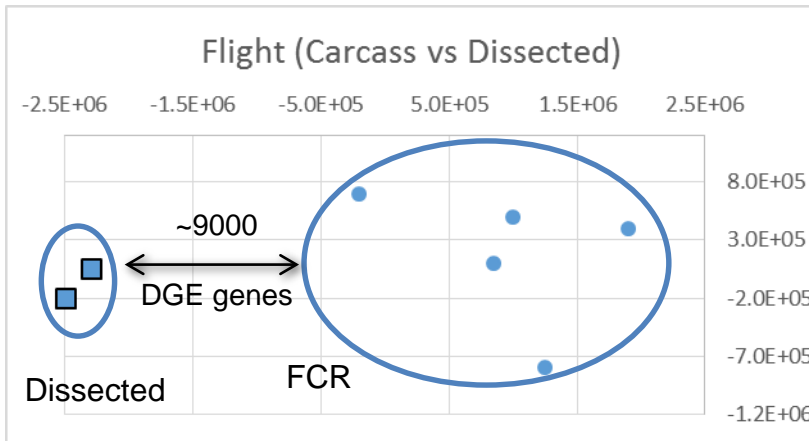
Dissected
then
Frozen

Full Carcass
frozen

Dissected
then
Frozen

Full Carcass
frozen


Principal Component Analysis of On-Orbit Dissected vs Frozen Carcass Livers

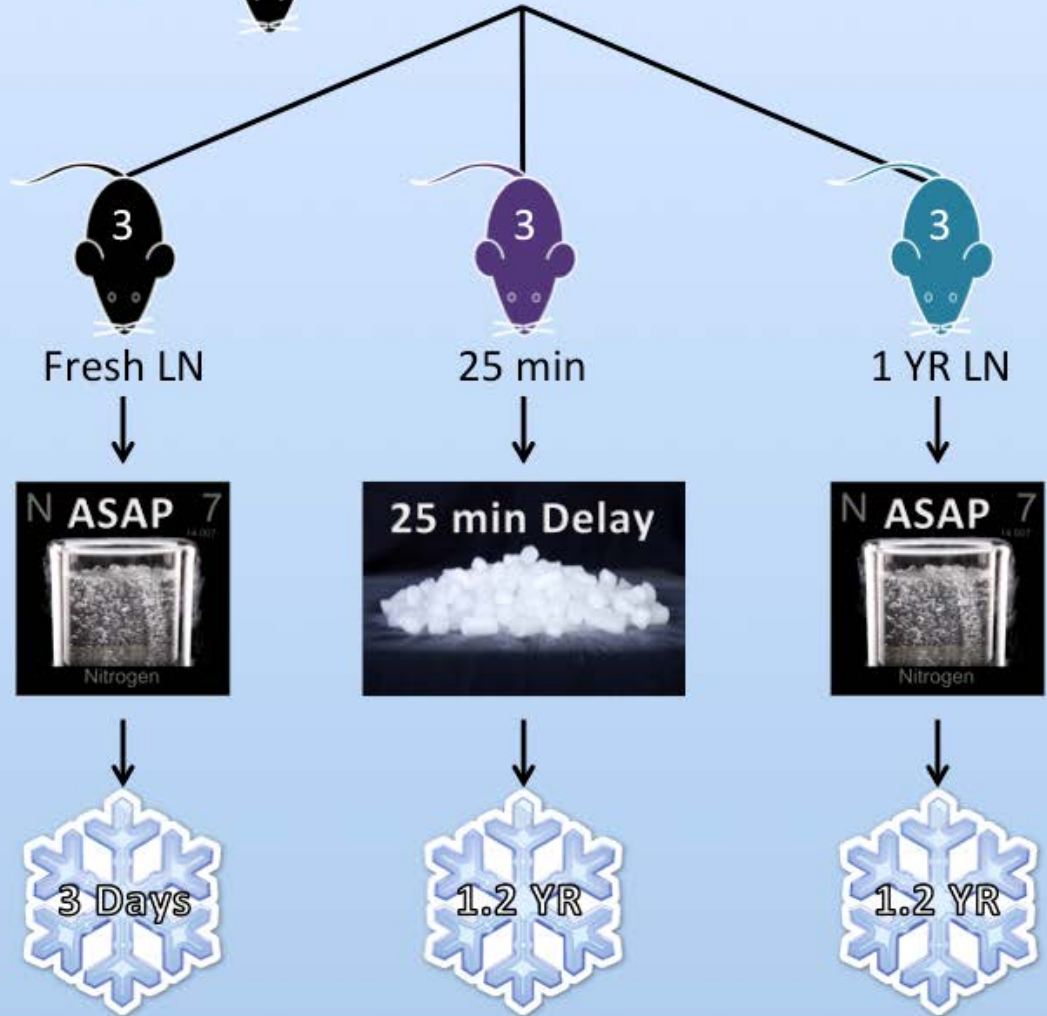


➔ Strong separation of differentially expressed genes between FCR and frozen tissue, either in space or on the ground (worst in space) – 4000 genes in common, principally linked to catabolic pathways (i.e. tissue degradation).

Transcriptomics Data: Pre-validation Experiment



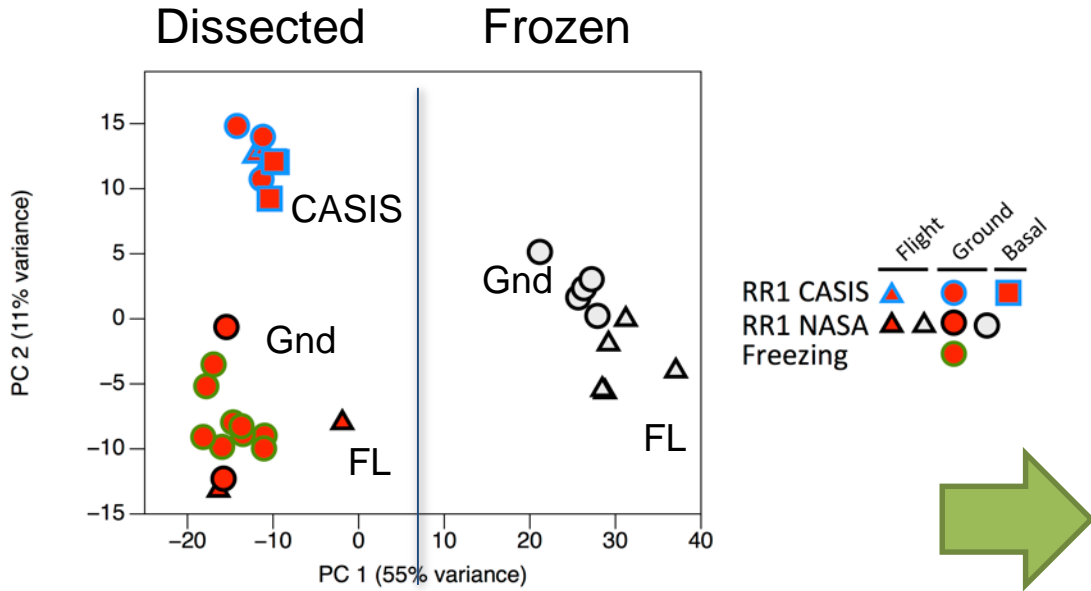
N =  9 C57BL/6J female mice, 12 weeks



Dissection

Storage Time

Freezing Before Dissection Changes RNA



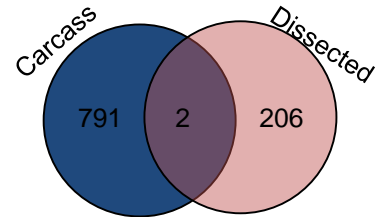
New experimental design to understand:

1. Is this effect specific to liver?
2. Are drugs used for euthanasia creating a system effect?
3. Can conclusions be reached by having proper controls?

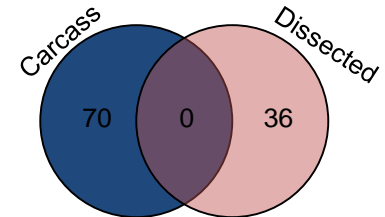
Principal component analysis of liver samples:

- Triangles - flight samples
- Circles - ground samples
- Squares - basal controls
- Red fill - dissected
- Gray fill - frozen carcass
- Blue outline: RR1 CASIS
- Black outline: RR1 NASA,
- Green outline: Freezing study

DGE



GSEA



The screenshot shows the GeneLab website layout. On the left is a navigation menu with links for Home, About, Latest Data Releases, GeneLab Data System, Analysis Working Groups, Upcoming Events, Newsletters, FAQ, and Contact Us. The main content area is divided into several sections: 'LATEST NEWS' featuring a 'GeneLab Analysis Working Group' sign-up button and a news item about Sylvain Costes; 'DATA REPOSITORY' with a search bar and a table of data releases; 'ANALYSIS WORKING GROUPS' with a 'Join the GeneLab Analysis Working Groups' link; 'SOCIAL MEDIA' with icons for ResearchGate, Facebook, and Twitter; and 'SOLICITATION' with a link to a new NASA solicitation. At the bottom right, there is a large graphic titled 'OMICS' with the tagline 'A Journey to See More Than Ever Before'.

Engage broadest community of researchers, industry, and citizen scientists to advance innovations

<https://genelab.nasa.gov>

- Weekly social media posts:
 - @NASA Ames **Facebook**
 - **Twitter** #GeneLab
 - **ResearchGate**: <https://www.researchgate.net/project/Omics-for-Space-Biology-The-GeneLab-project>
- GeneLab database listed in science journals:
 - *Scientific Data*, Oxford e-Research
- GeneLab issues Digital Object Identifiers (DOI) via DataCite
- Customer Support: Respond and resolve all inquiries from science community, academia, public

GeneLab Acknowledgements



Chris Barreras
Afshin Beheshti
Dan Berrios
Valery Boyko
Sonja Caldwell
Jairon Camarillo
Kaushik Chakravarty
Egle Cekanaviciute
John Costa
Sylvain Costes (PM)
Marie Dinh
Sandy Dueck
Homer Fogel
Jon Galazaka (PS)
Samrawit Gebre
Dennis Heher
Lynn Hutchison
Yared Kidane
San-Huei Lai Polo
Tristan Le
Qiang Li
Shu-Chun Lin
Sneha Raghunandan
Shayoni Ray
Sigrid Reinsch
David Smith
Marla Smithwick
Hao Thai
Khai Peter Tran
Andrew Williamson

TO JOIN GENELAB AWG, CONTACT:
sylvain.v.costes@nasa.gov



Genelab.nasa.gov