National Aeronautics and Space Administration

GeneLab: “Omics” Data System for Space Biology Research

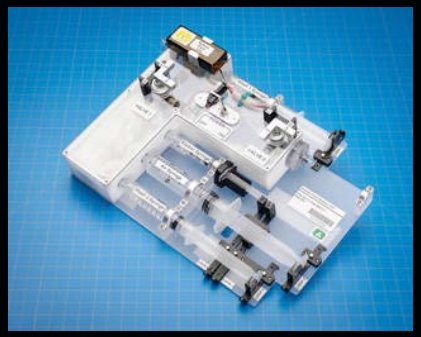
Sylvain Costes, Ph.D.
Project Manager

Marla Smithwick
Deputy Project Manager

The GeneLab Team

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.

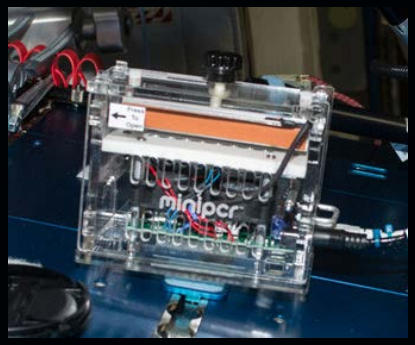


Sample Preparation Module

Oxford Nanopore MinION Gene Sequencer

Cepheid Smart Cycler qRT-PCR

Reaction tube containing lyophilized chemical assay bead (proprietary)



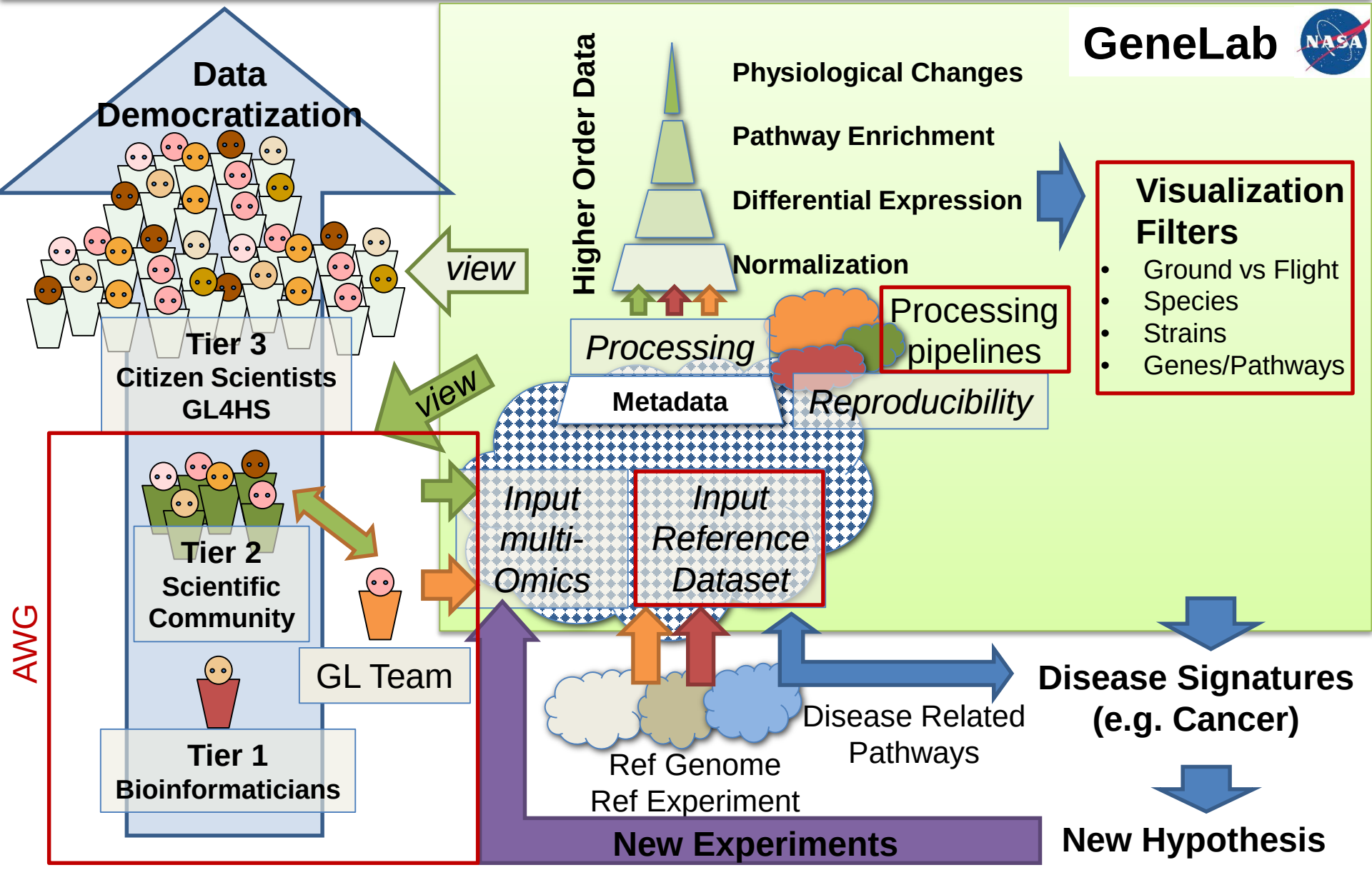
Mini-PCR

ISS Based Research

- 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 Reseach-1 hardware in the Microgravity Science Glovebox aboard the International Space Station.

Three-tier Client Strategy to Democratize Data



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:

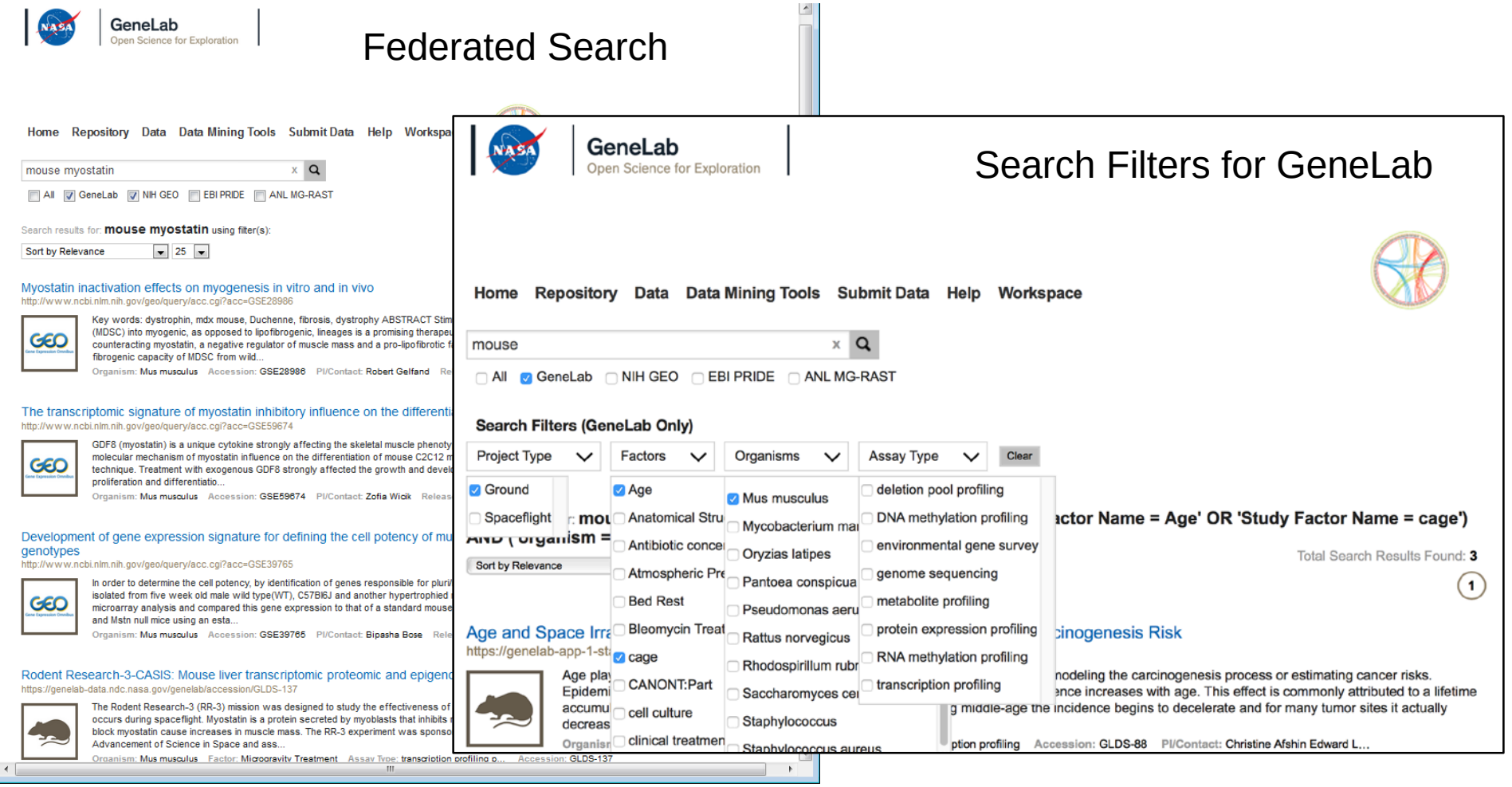
1. Peer-reviewed publications describing AWG's comprehensive analysis.
2. Consensus data analysis pipelines relevant to AWG domains to be used on the GLDS will help domains harmonize their analyses.
   a) Summer interns will process all data based on AWG recommendation
   b) Processed "higher-order" 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

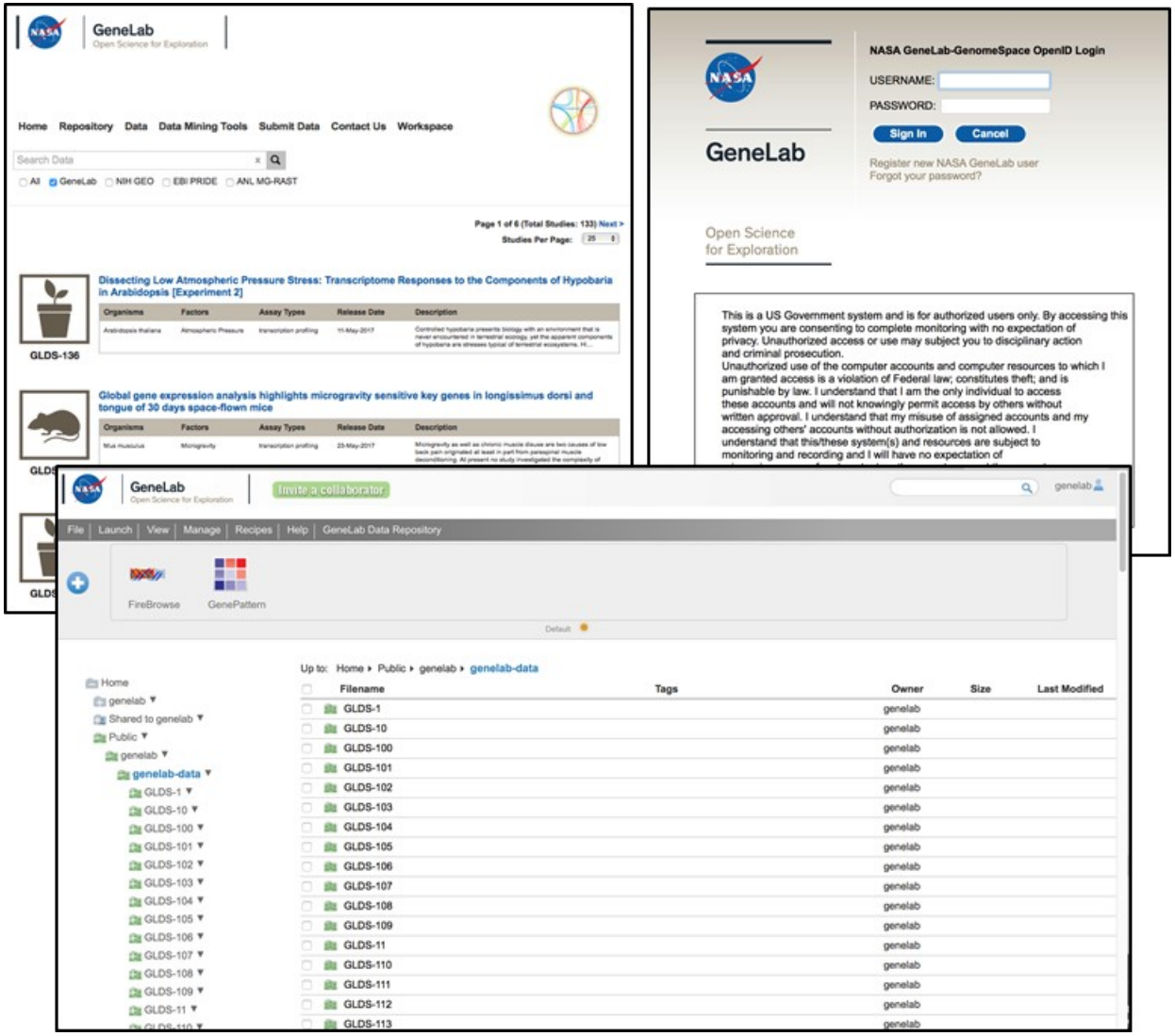
GLDS Phase 2 (Release 2.0) Google-like Search, Federated Search

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



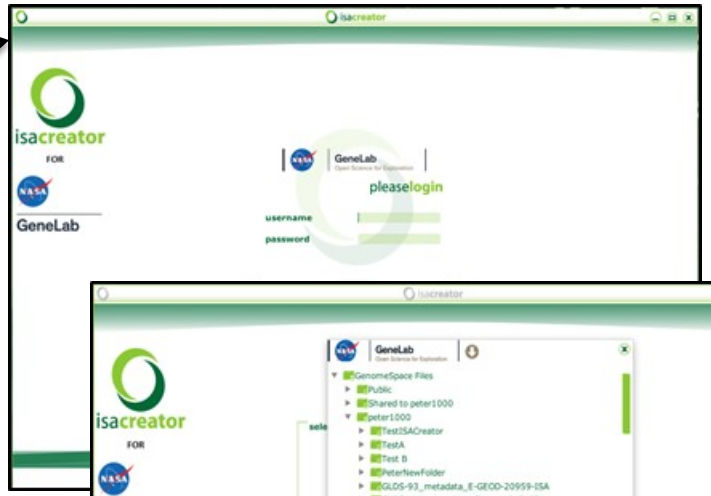
GLDS Phase 2 (Release 2.0) Customized NASA Collaborative Workspace

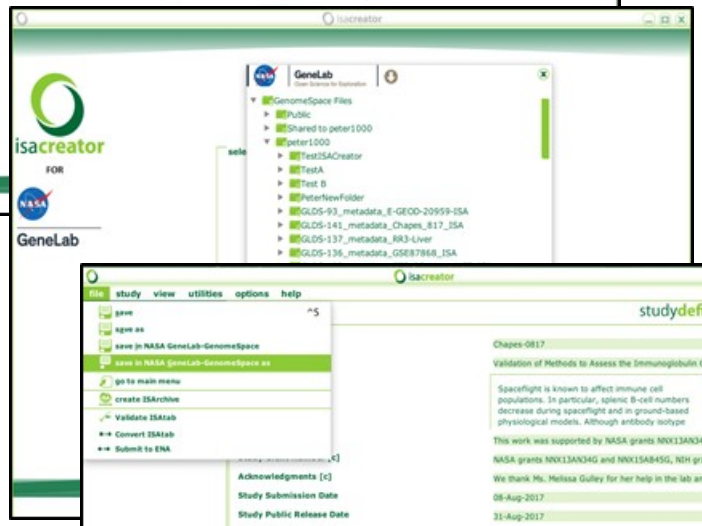
User Account Mgmt., Access Controls (e.g., Private, Shared, Public Folders)

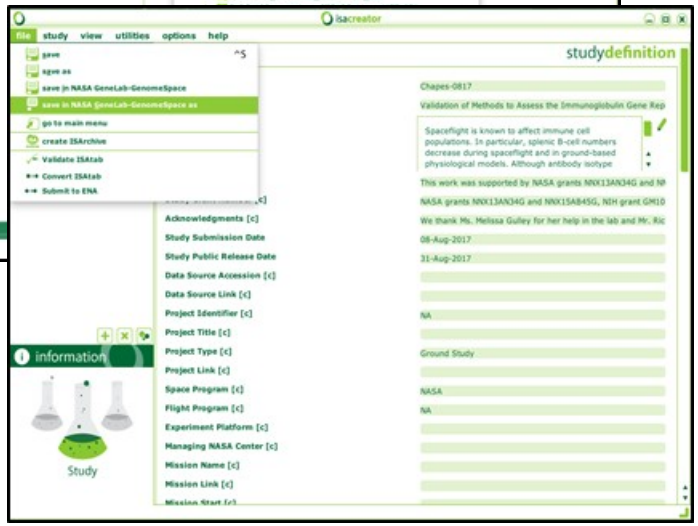


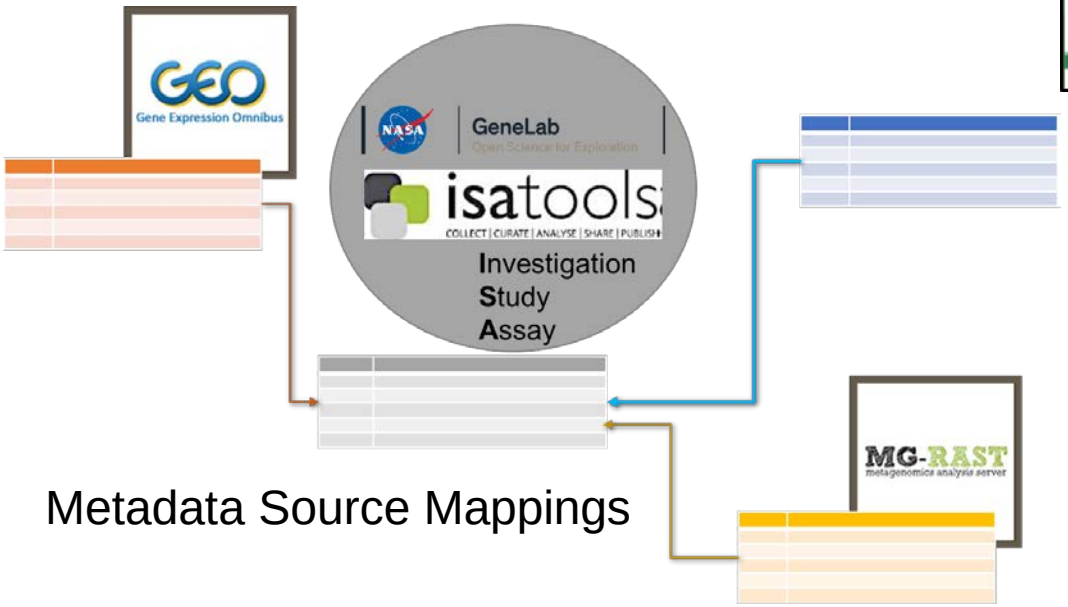
GLDS Phase 2 (Release 2.0) Metadata Curation via ISACreator Tool

GeneLab-GenomeSpace Integration with ISACreator for Streamlining Data Processing Operations



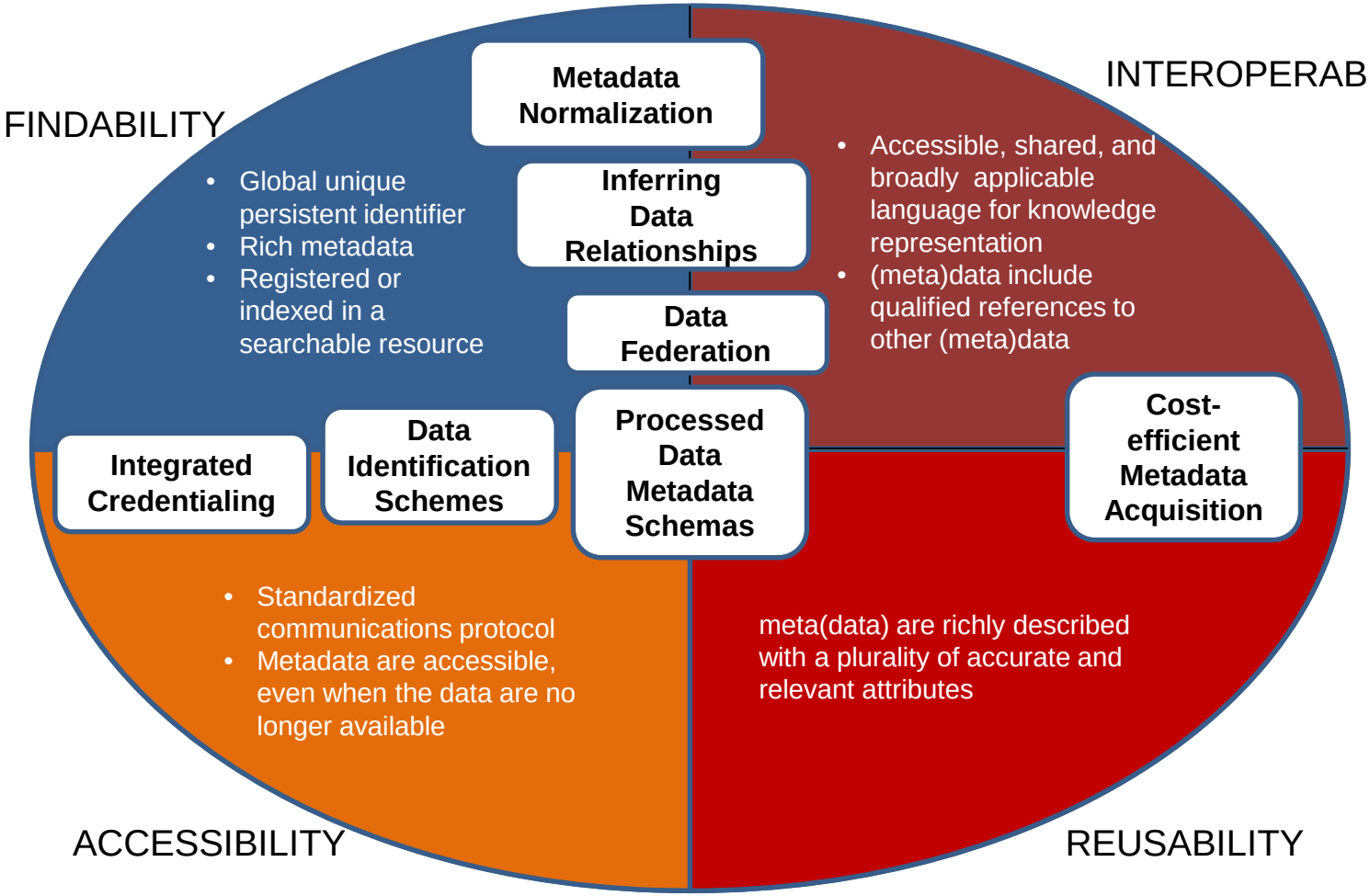




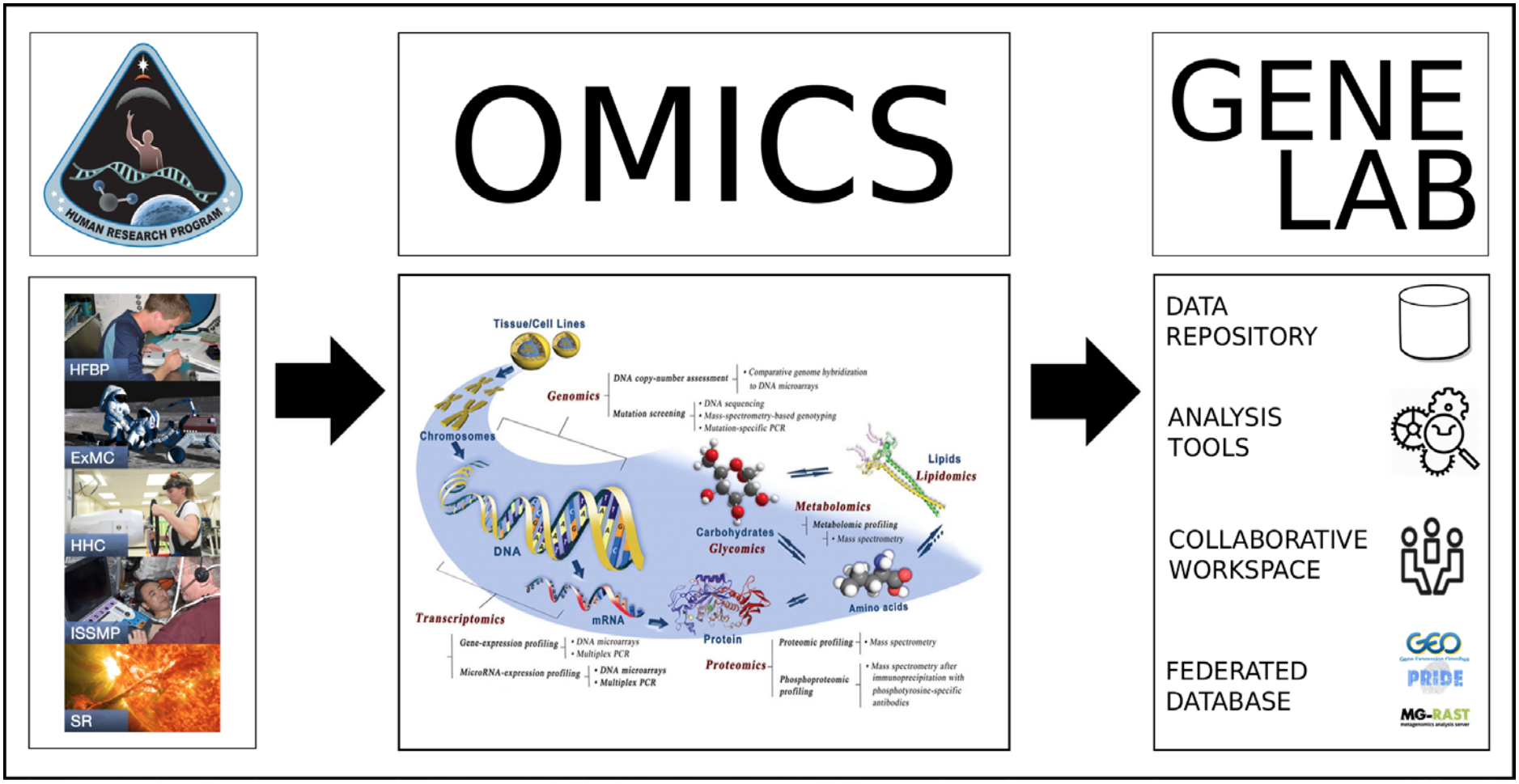


Metadata Source Mappings

GeneLab has adopted the FAIR principle

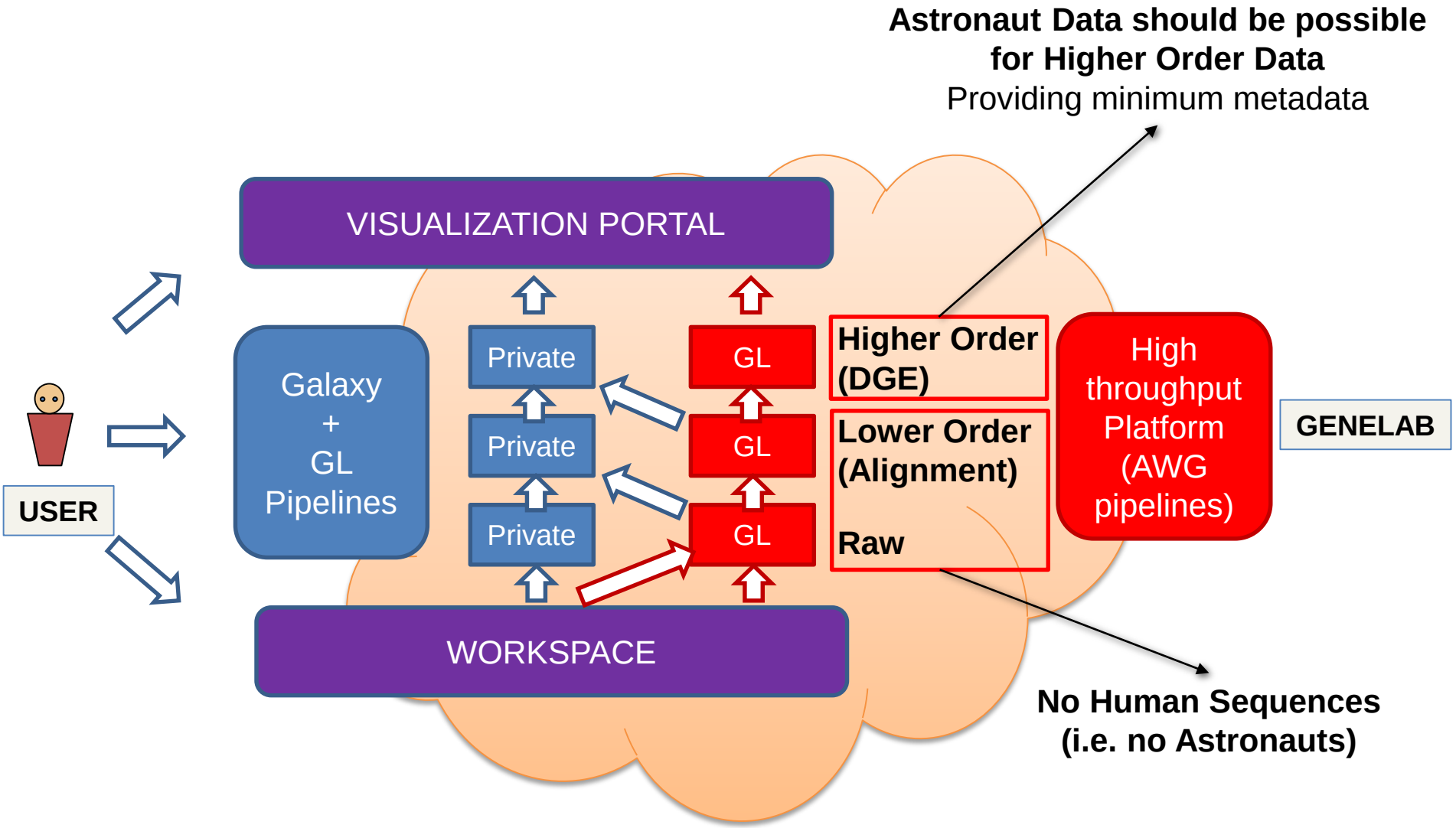


HRP, GeneLab and Omics

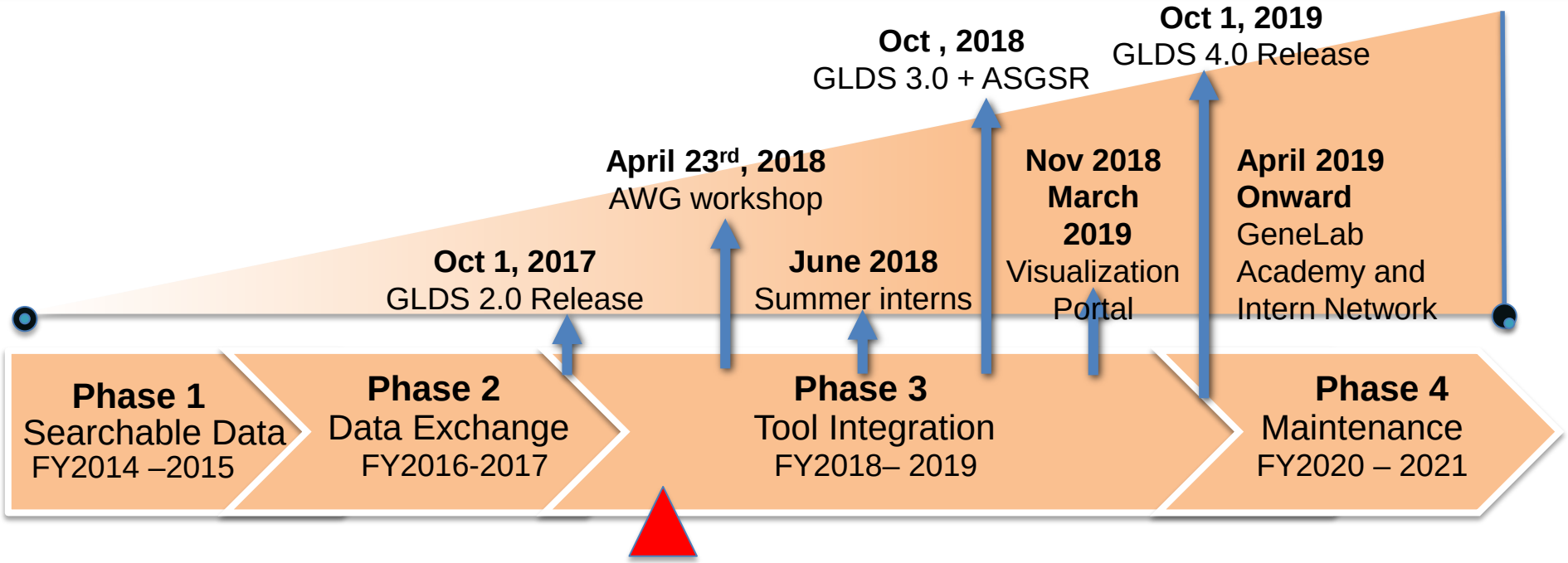


The GeneLab database infrastructure provides a platform for storage, retrieval and analysis of omics datasets – with the ultimately goal to support the missions of HRP

GLDS 4.0


Astronaut Data should be possible
for Higher Order Data
Providing minimum metadata
VISUALIZATION PORTAL
Galaxy
+
GL
Pipelines
Private
Private
Private
GL
GL
GL
Higher Order
(DGE)
Lower Order
(Alignment)
Raw
High
throughput
Platform
(AWG
pipelines)
GENELAB
USER
WORKSPACE
No Human Sequences
(i.e. no Astronauts)

Phased Implementation



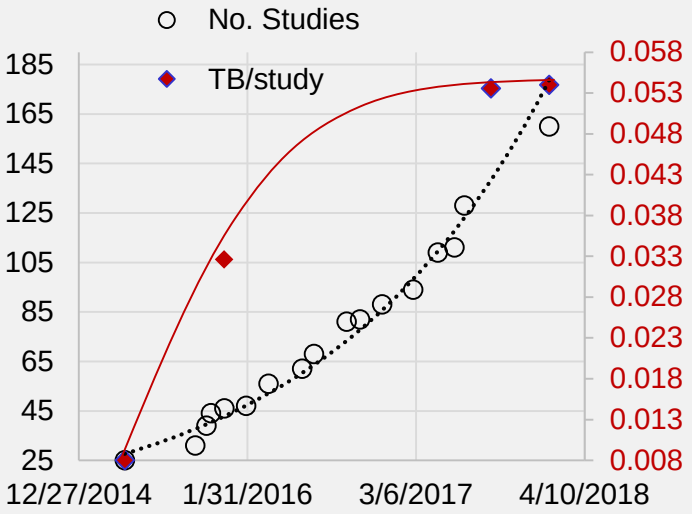
Data System
- ✓ Public Website
- ✓ Searchable Data Repository
- ✓ Top Level Requirements
- ✓ New Data and Legacy Data

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

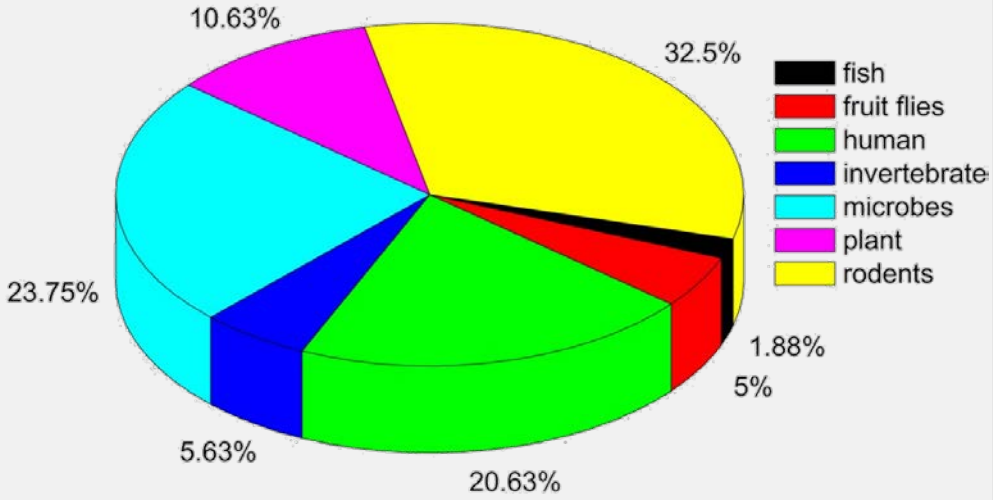
Data System
- Integrated Platform across model organisms
- Build Community via AWG
- Provide access to biocomputational tools for omics analysis
- Provide collaboration framework and tools

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

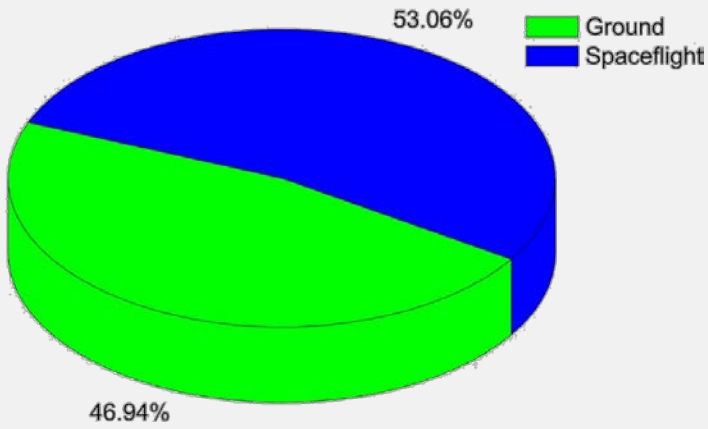
GeneLab Database: 154 data sets



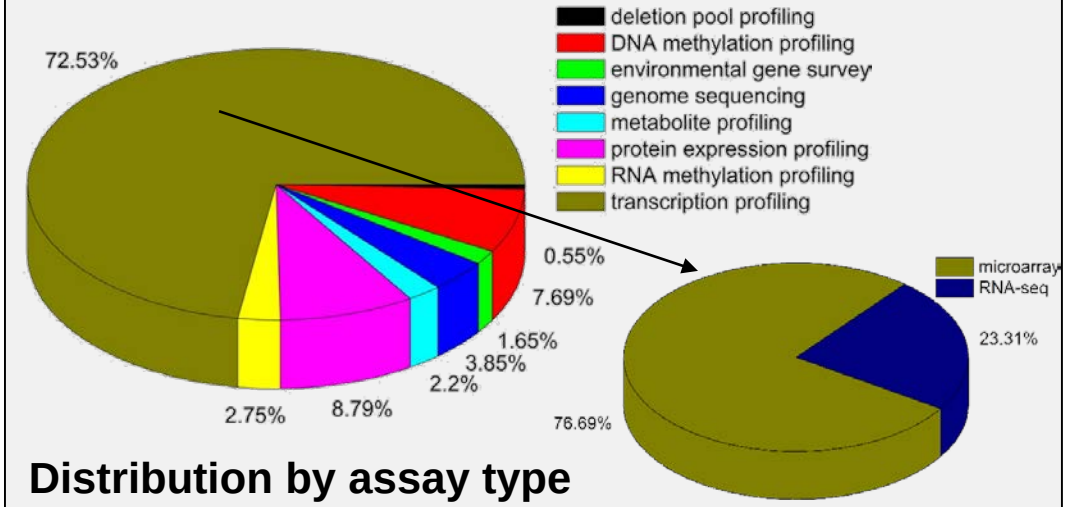
Data Growth Since 2014



Distribution by organism type

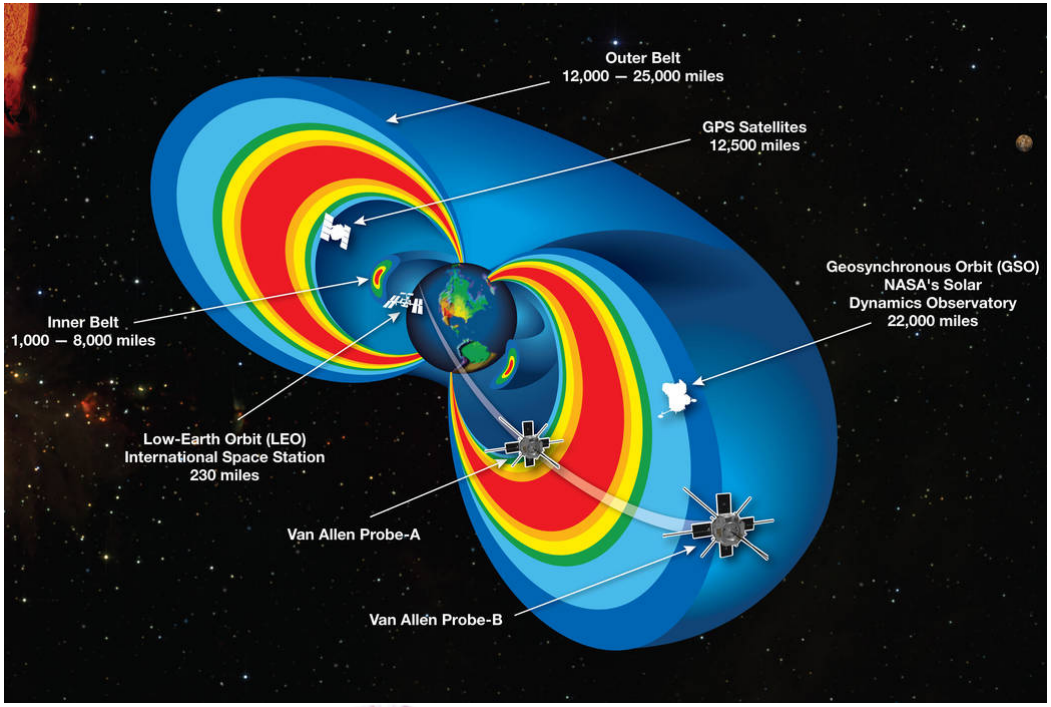


Majority is spaceflight samples

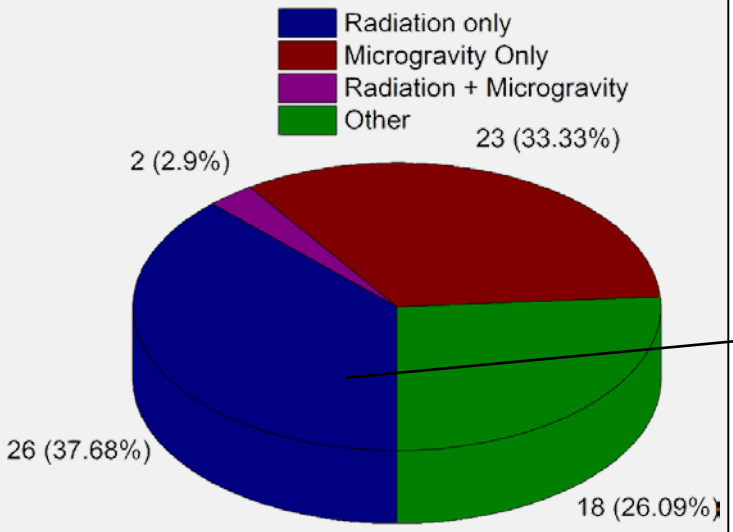


Distribution by assay type

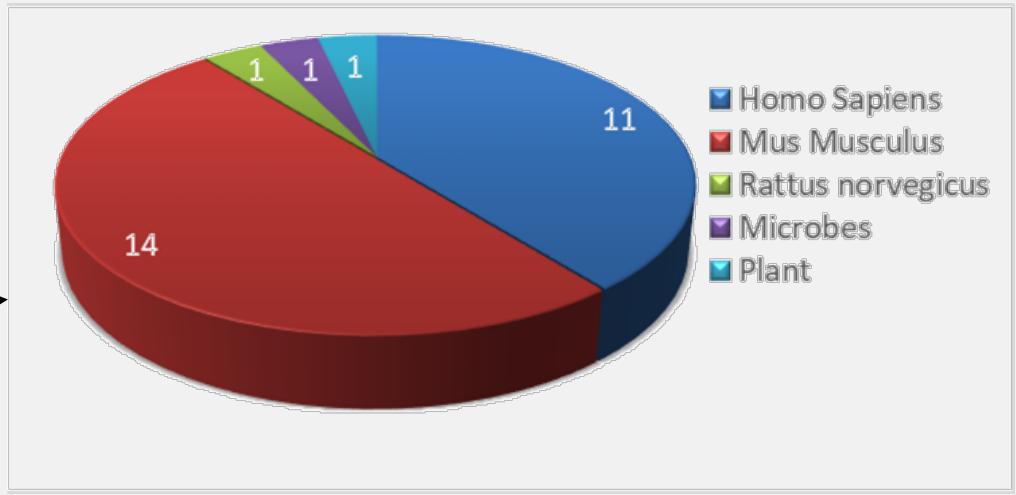
Earth's magnetic field protects us from cosmic radiation



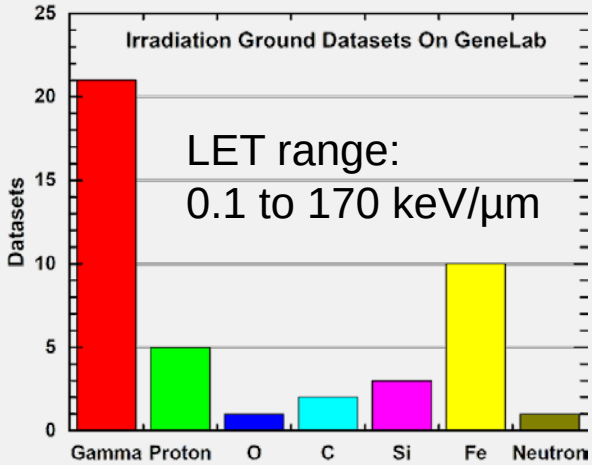
69 Ground Data Sets



Types of Ground Data

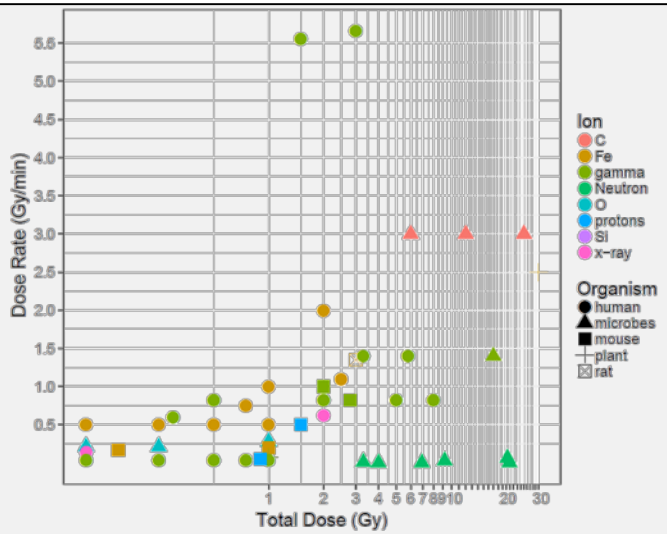


28 Ground Radiation Data Sets by Organism

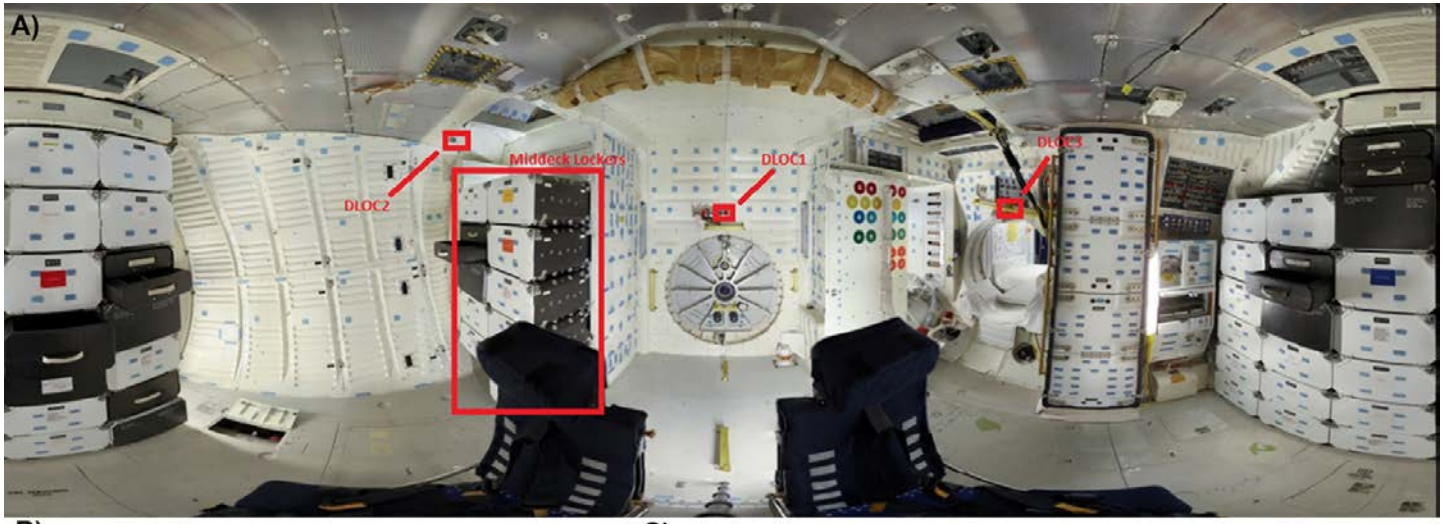


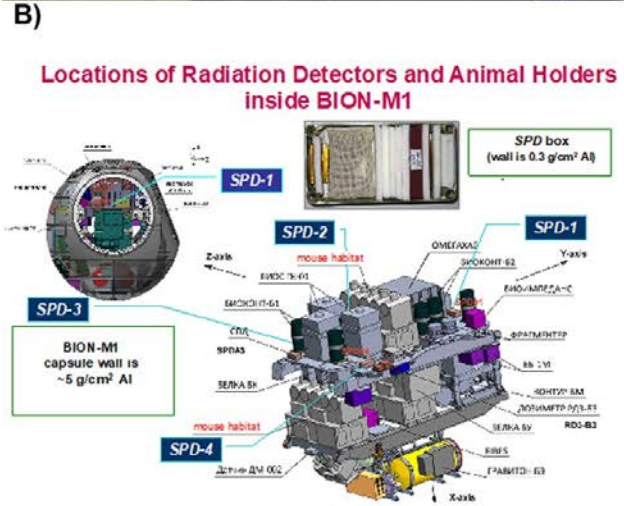
Radiation Quality Distribution

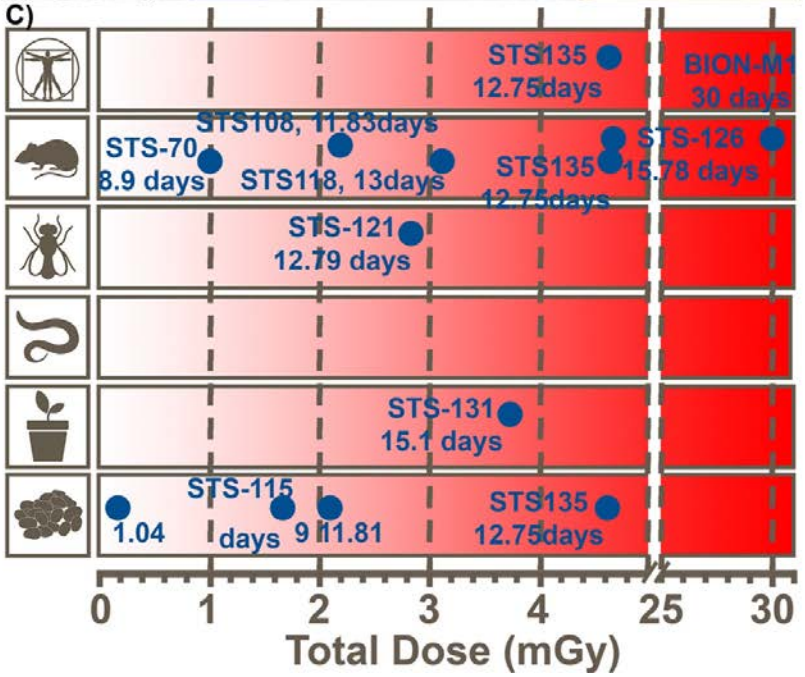
Dose and Dose rate distribution



STS Samples: Radiation Dosimetry





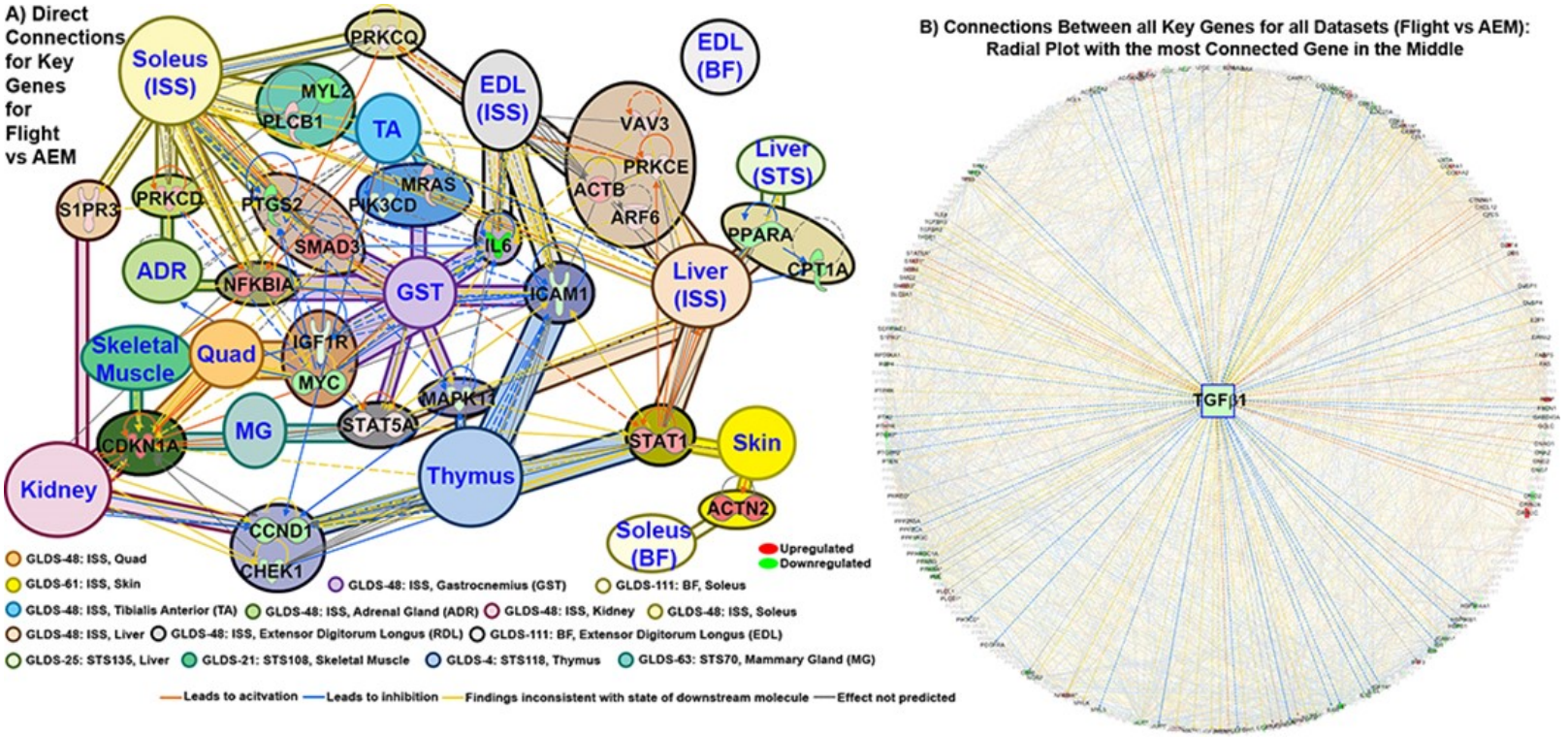


Building a database to support studies on human health and countermeasures

Future analysis capabilities

1. Cohort comparison

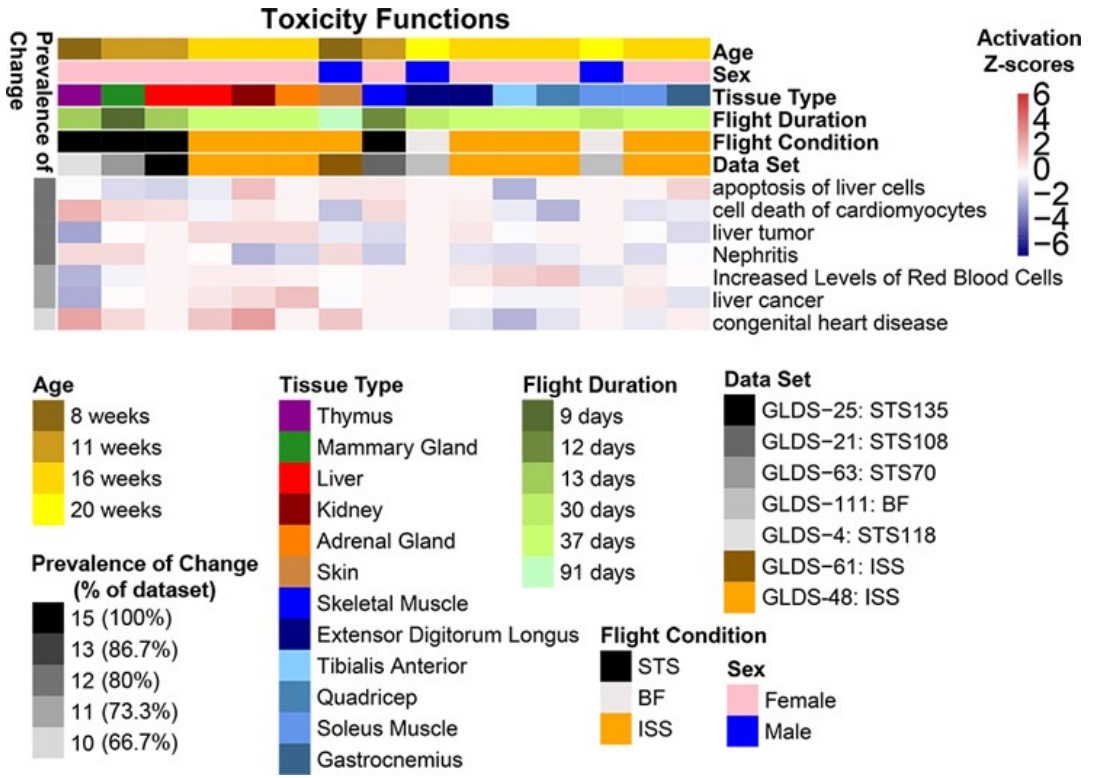
- Display the expression of a gene query or its frequency of differential regulation based on sex, species, tissue, or age
  - •Example: From a systems biology analysis, TGFβ1 was found to be a master regulator impacting spaceflight



Building a database to support studies on human health and countermeasures

Future analysis capabilities

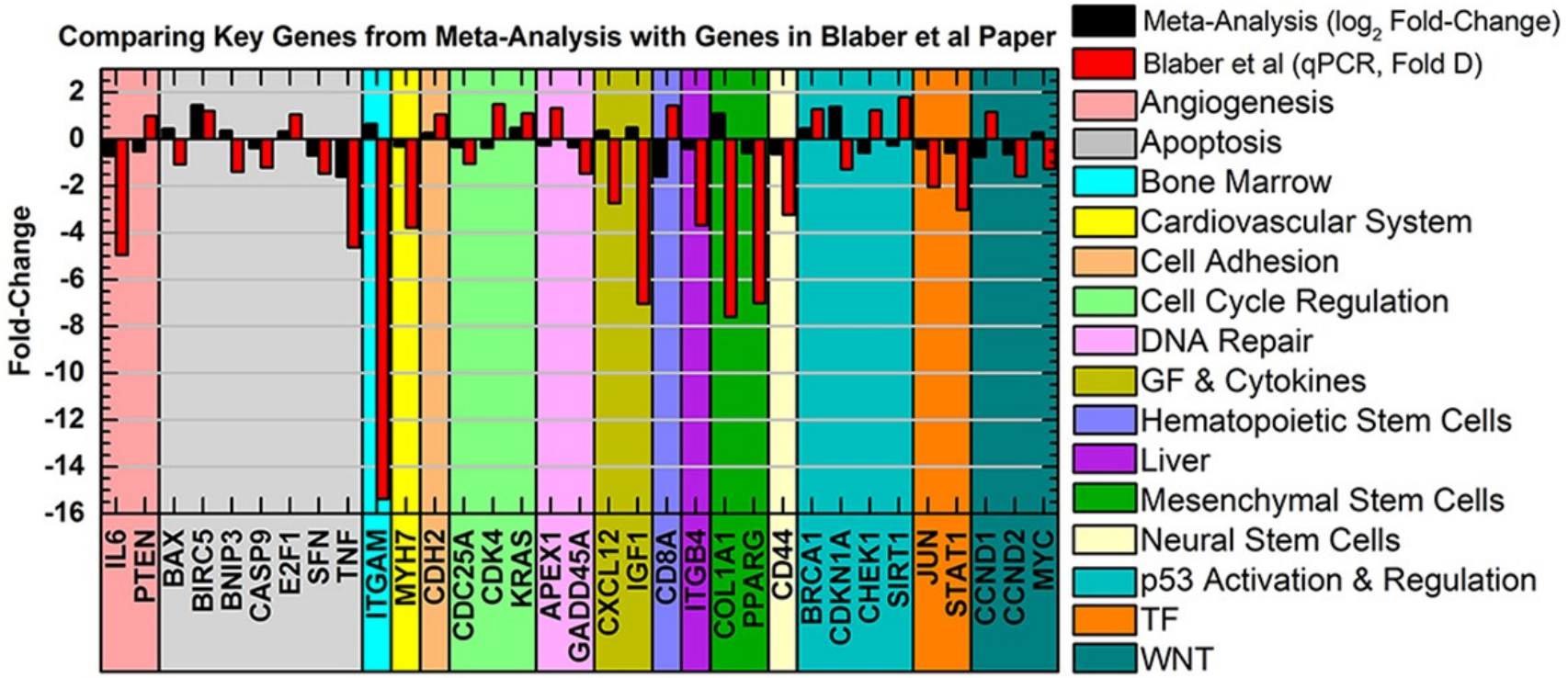
2. Relevance to human disease
   - Display the expression of a query gene or its frequency of differential regulation in disease types
     - Example: Using the GeneLab data we are able to make predictions on impact on health and risk of diseases due to space flight



Building a database to support studies on human health and countermeasures

Future analysis capabilities

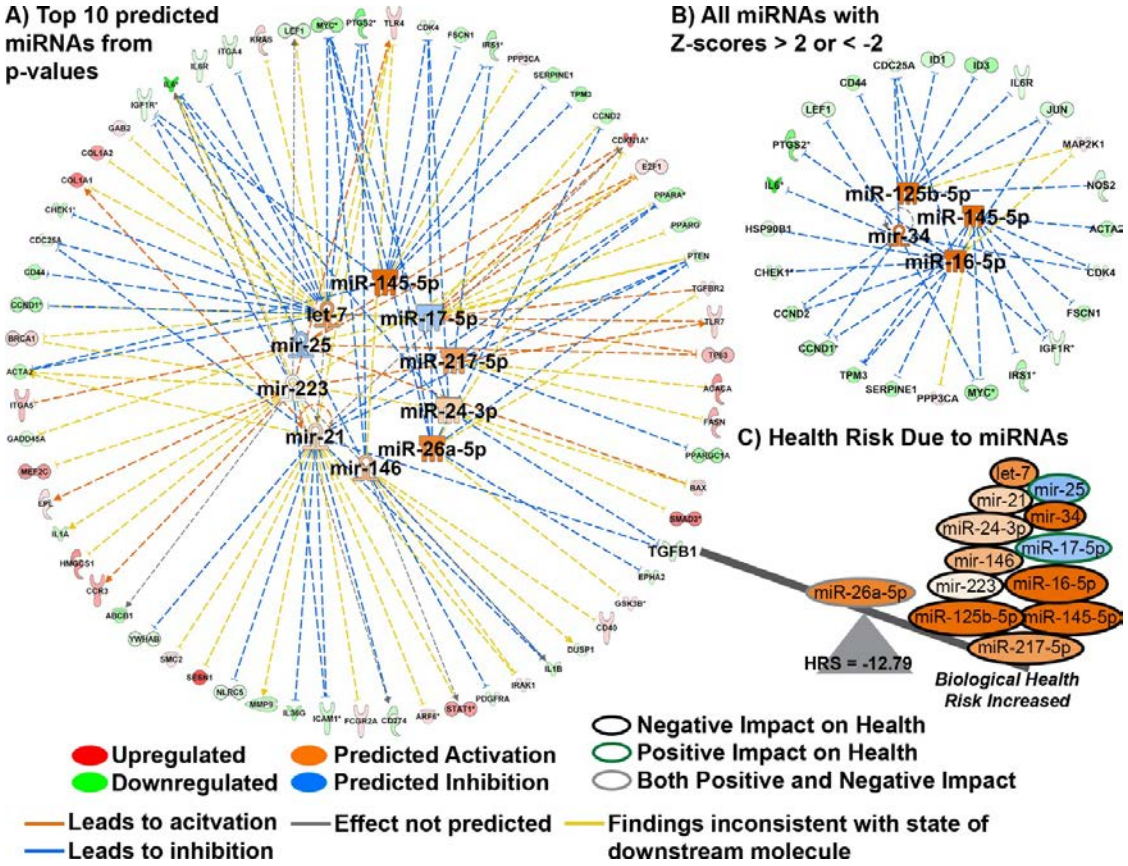
3. Tissue expression
   - Display the expression of a query gene based on cell or tissue type
     - Example: Can make direct comparisons from of key genes to data from the literature.



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.



Current studies

Sylvain Costes, Ph.D.
Afshin Beheshti, Ph.D.

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

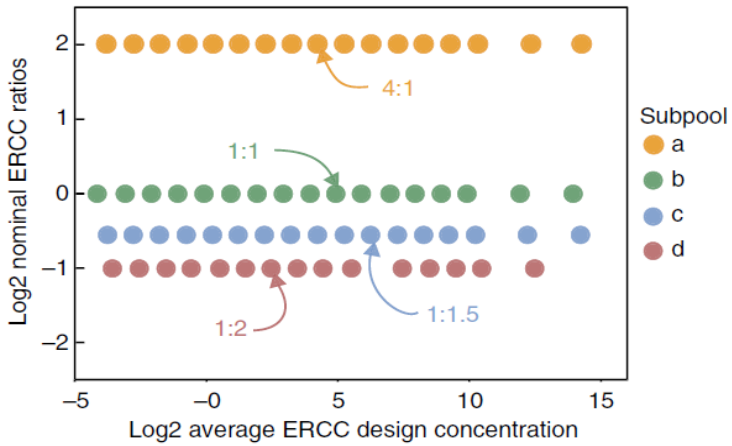
ARTICLE

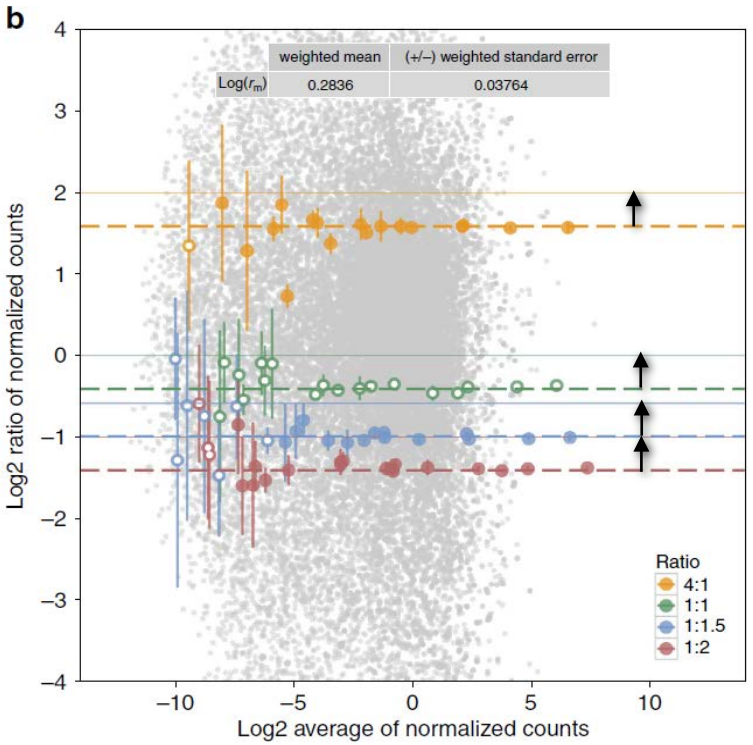
Received 11 Aug 2014 | Accepted 1 Sep 2014 | Published 25 Sep 2014

DOI: 10.1038/ncomms6125

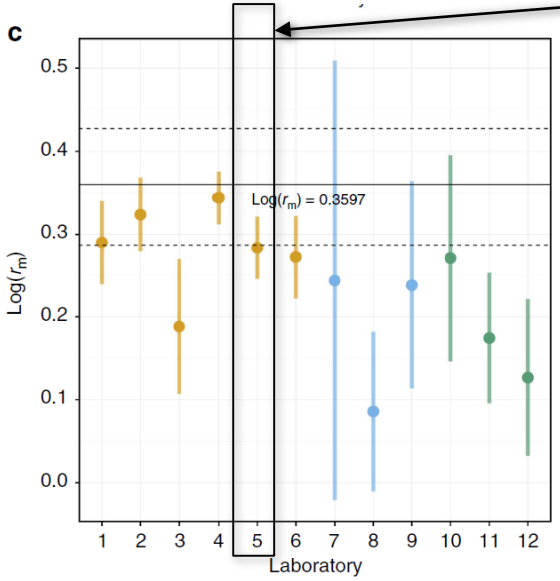
Assessing technical performance in differential gene expression experiments with external spike-in RNA control ratio mixtures

#1 Risk: Data Reproducibility





Lab #5
Log(r_m) = 0.2836

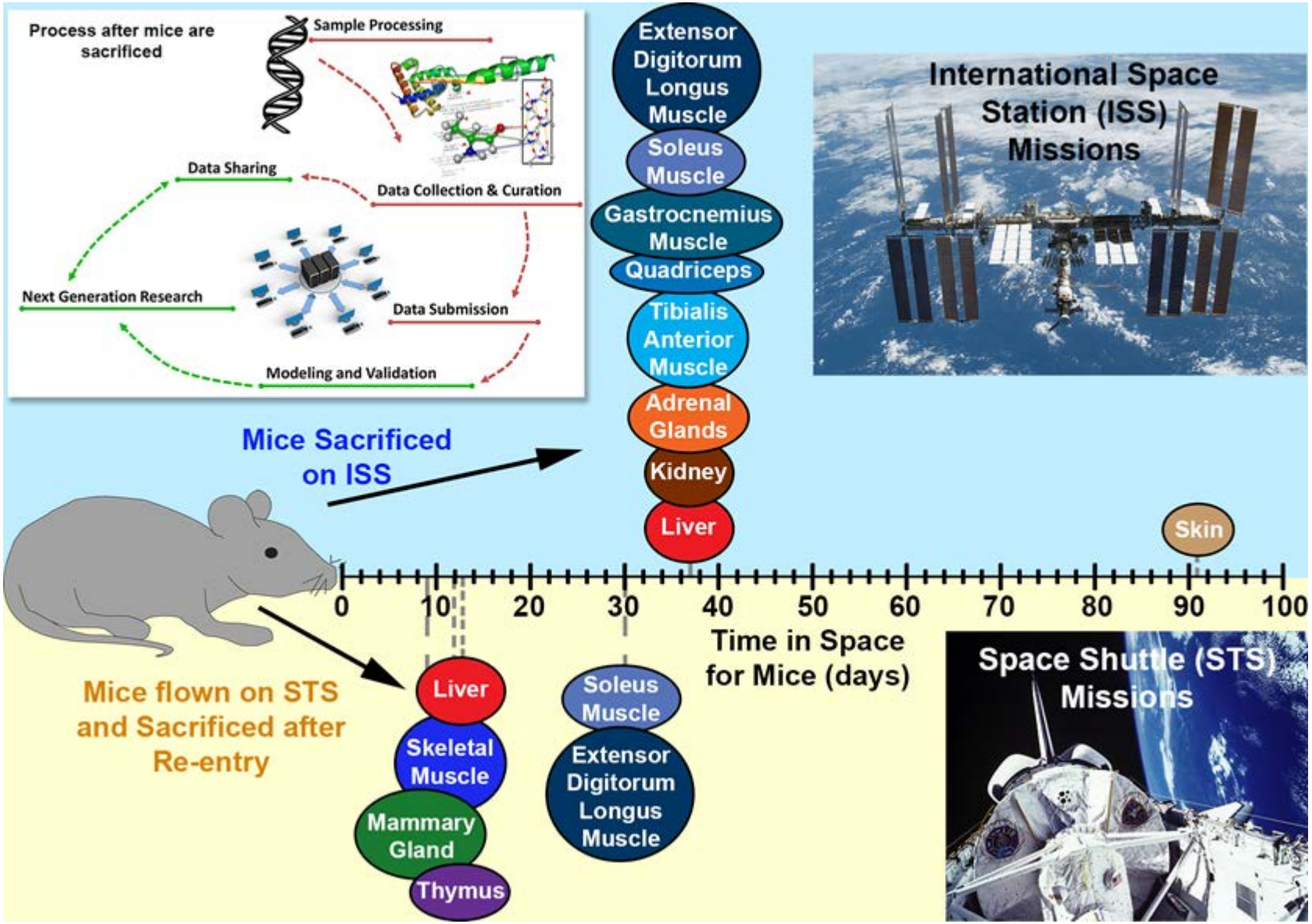


r_m = Weighted mean estimates of mRNA fraction differences for the sample set with error bars representing weighted standard errors

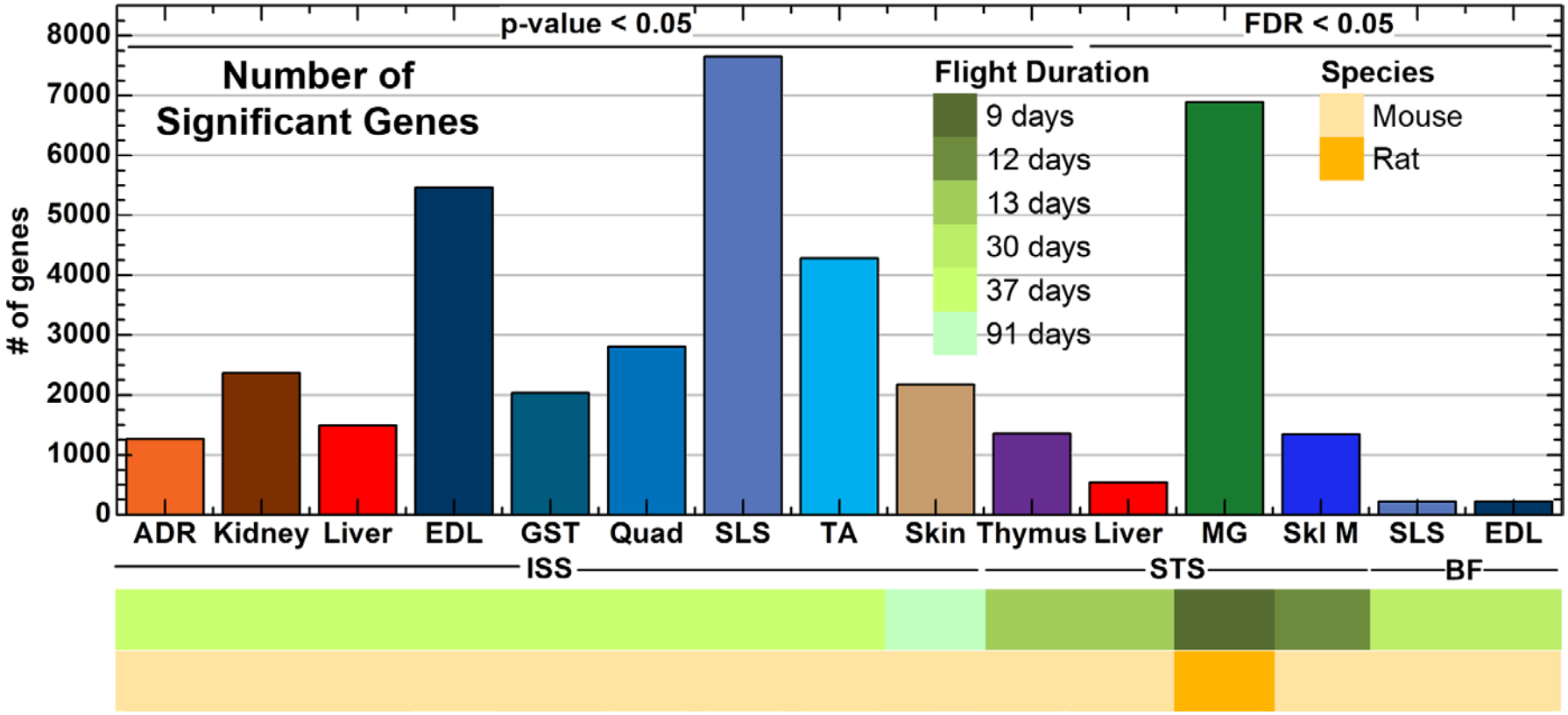
$$R_S = r_m \left(\frac{E_1}{E_2}\right)_S$$

$$Log(R_S) = Log(r_m) + Log(E_1) - Log(E_1)$$

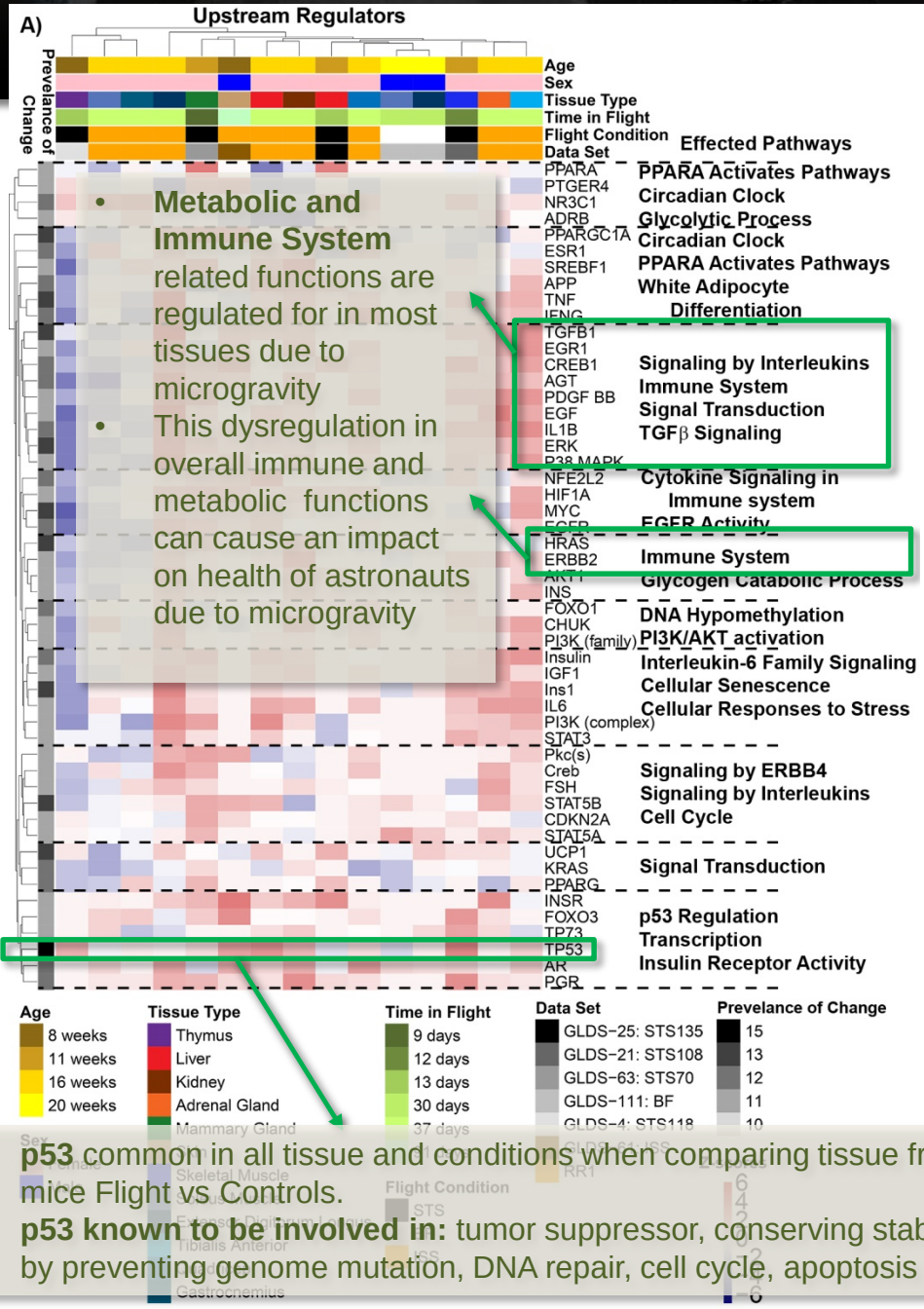
General Overview of GeneLab Mice Data



Number of Significant Genes from Multiple Datasets



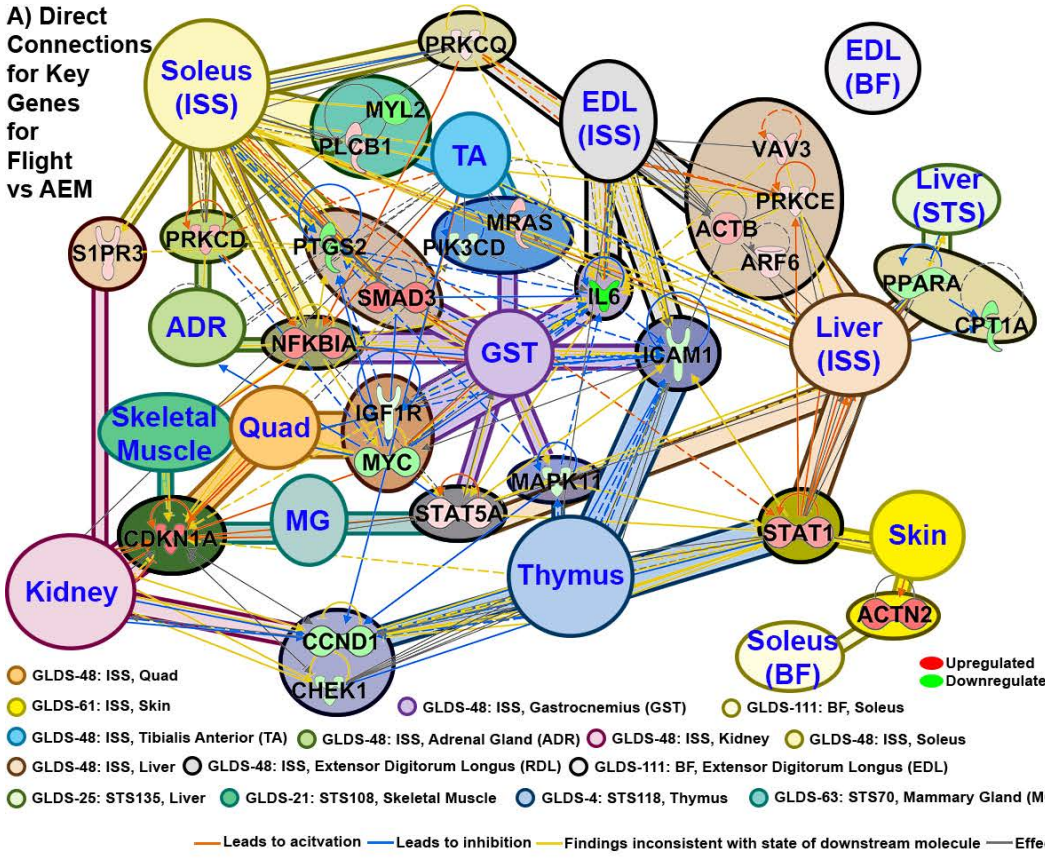
Predicted Master Regulators



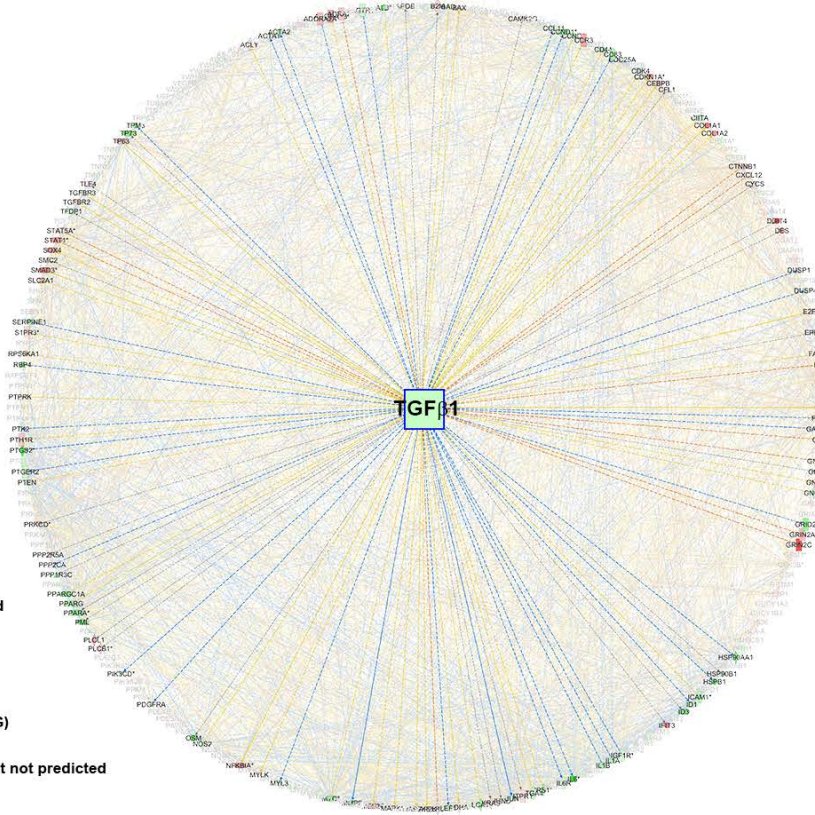
- **Metabolic and Immune System** related functions are regulated for in most tissues due to microgravity
- This dysregulation in overall immune and metabolic functions can cause an impact on health of astronauts due to microgravity

- **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)

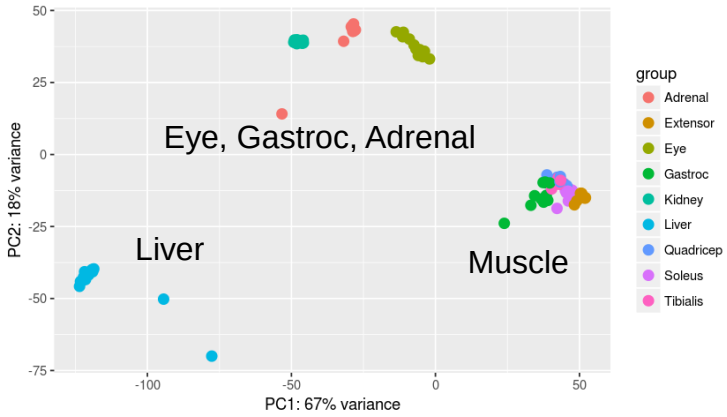


B) Connections Between all Key Genes for all Datasets (Flight vs AEM): Radial Plot with the most Connected Gene in the Middle

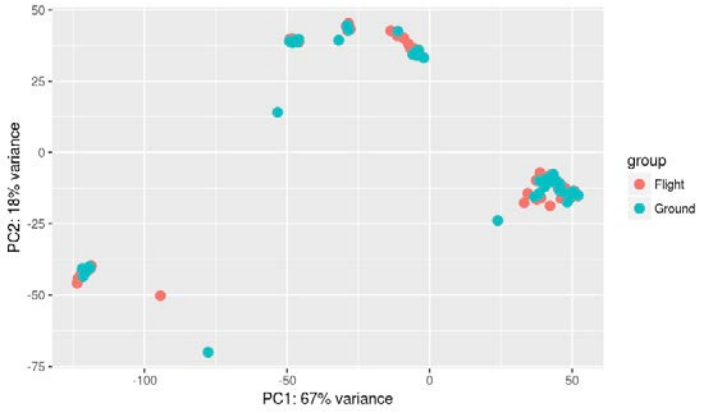


Mission-specific analysis: RR-1 Transcriptomics

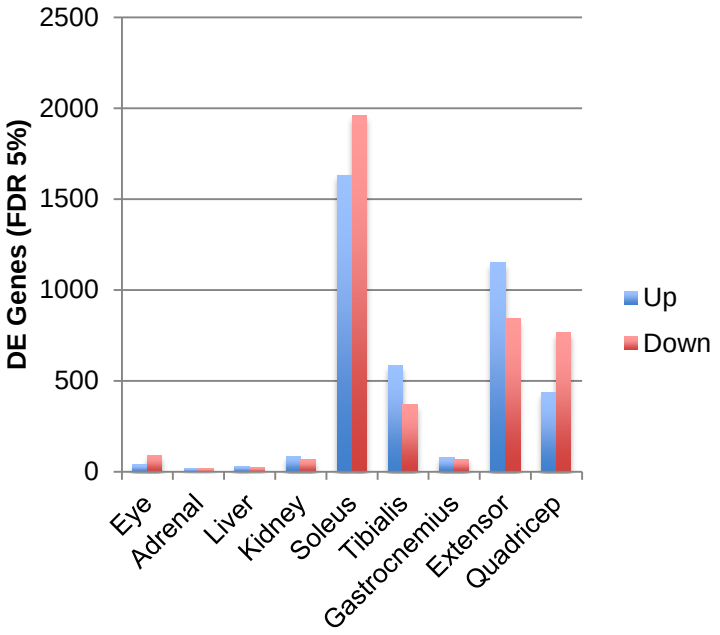
Samples cluster by tissue type



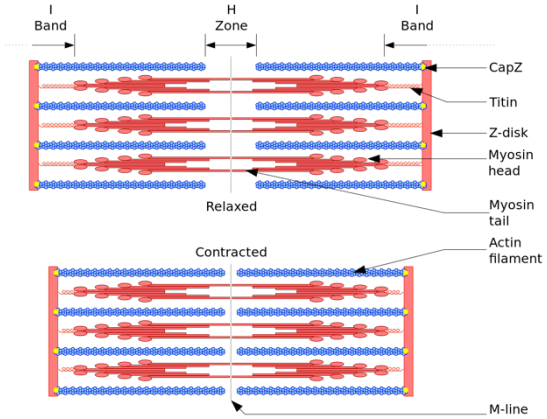
Samples do not cluster by flight/ground



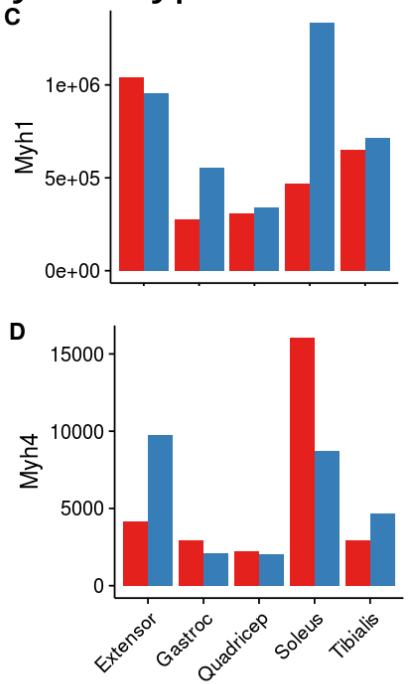
Many changes in muscles



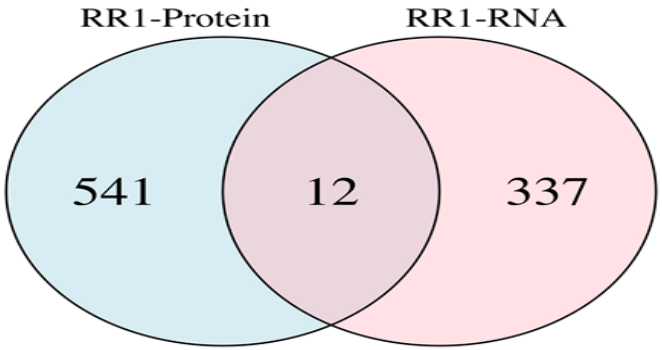
Changes to muscle myosin types

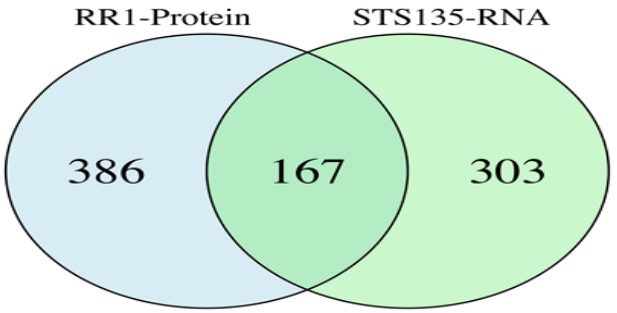


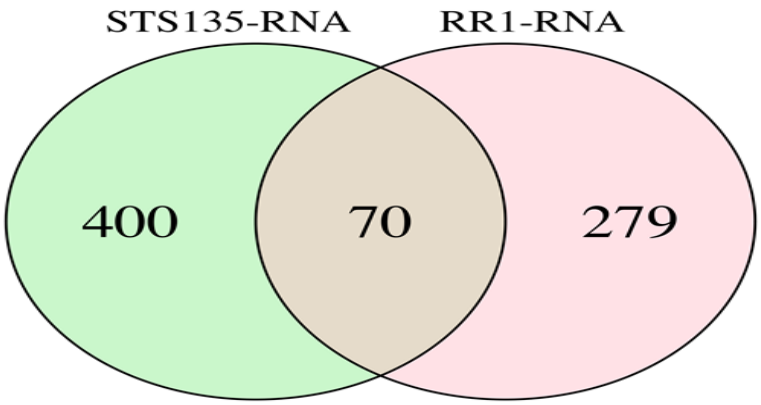
Muscle structure

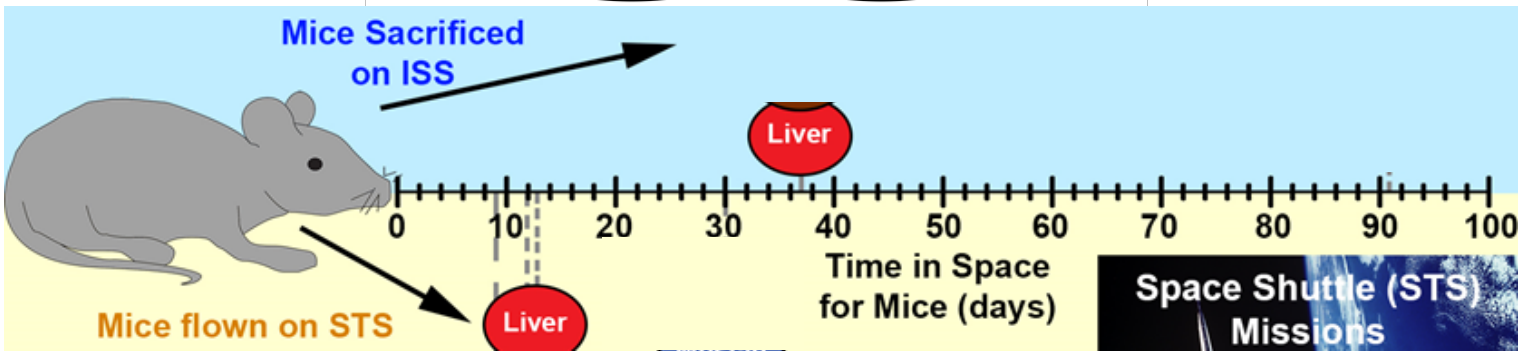


Impact of Microgravity on Liver Tissue: STS135 & RR1 Intersect Venn Diagram Analysis



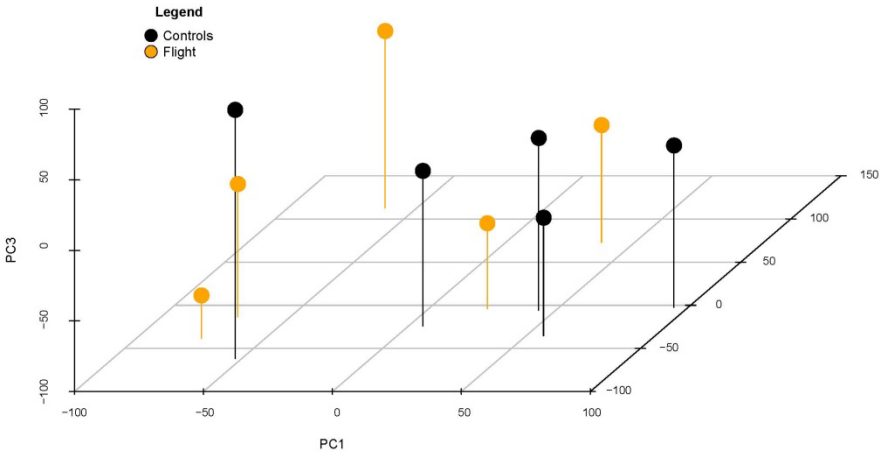




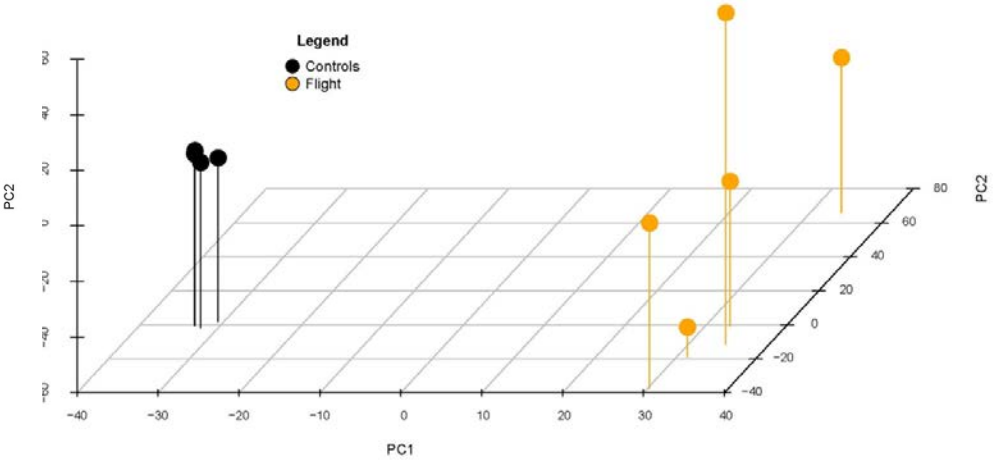


Impact of Microgravity on Liver Tissue: STS135 & RR1 Principle Component Analysis

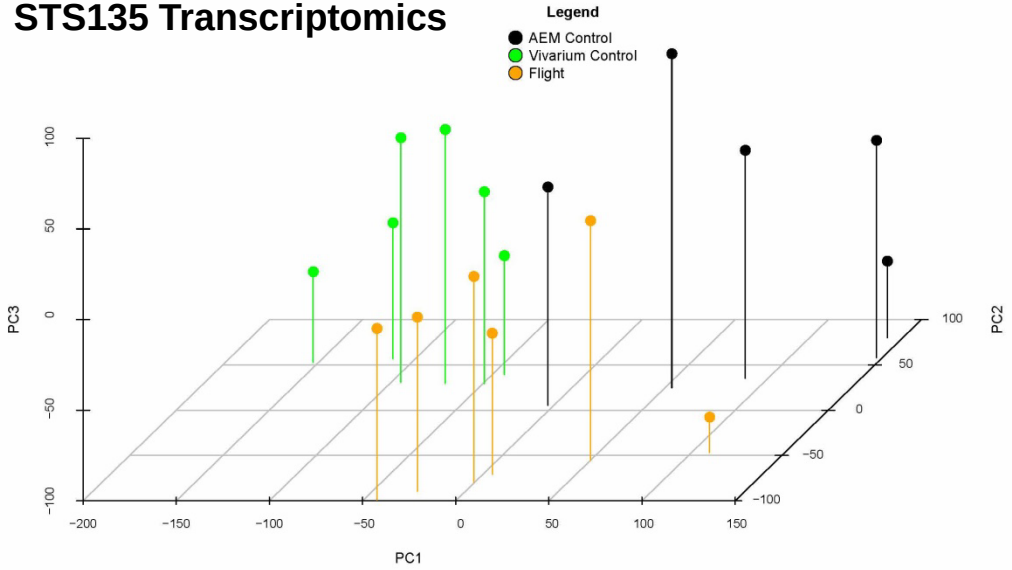
RR1 Transcriptomics



RR1 Proteomics



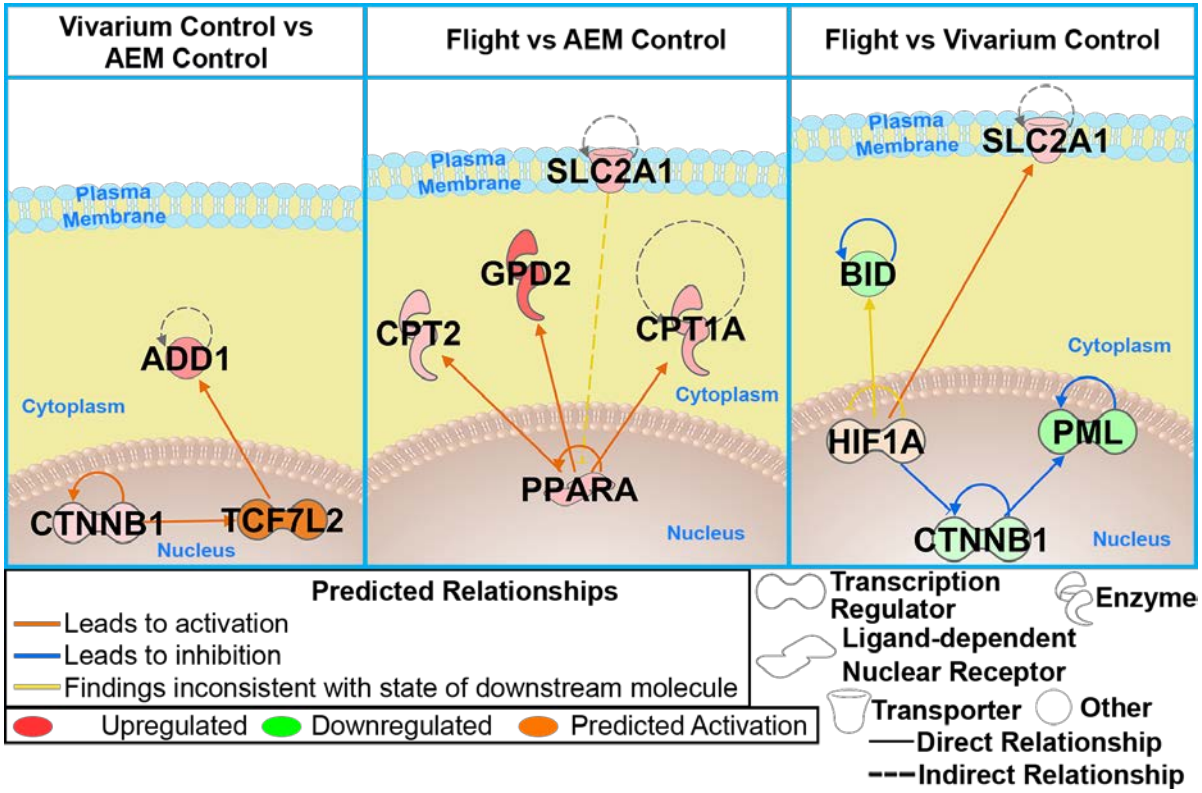
STS135 Transcriptomics



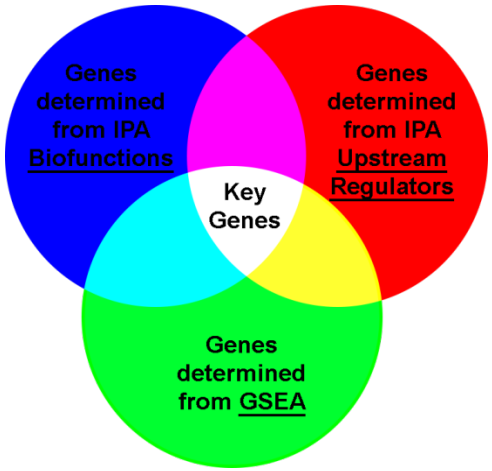
RR1 & STS135 Mice Liver KEGG Enrichment Pathways

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 – Astronauts may develop NASH disease

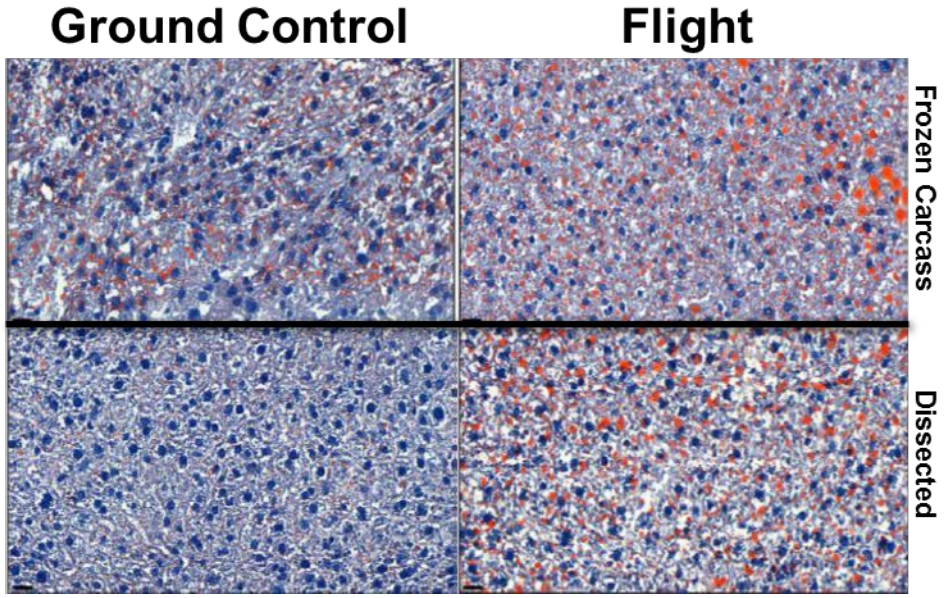


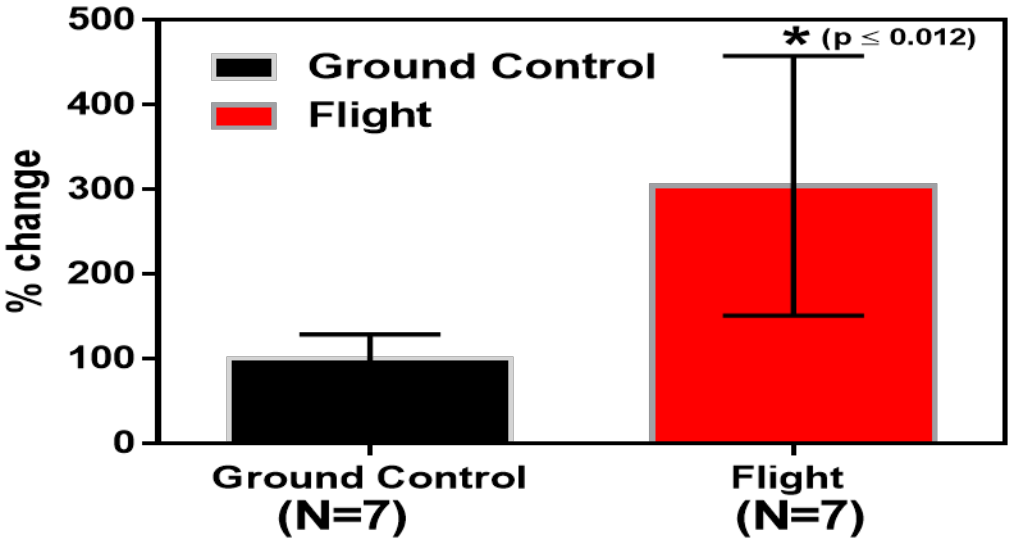
Key genes determined by the following:



- 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

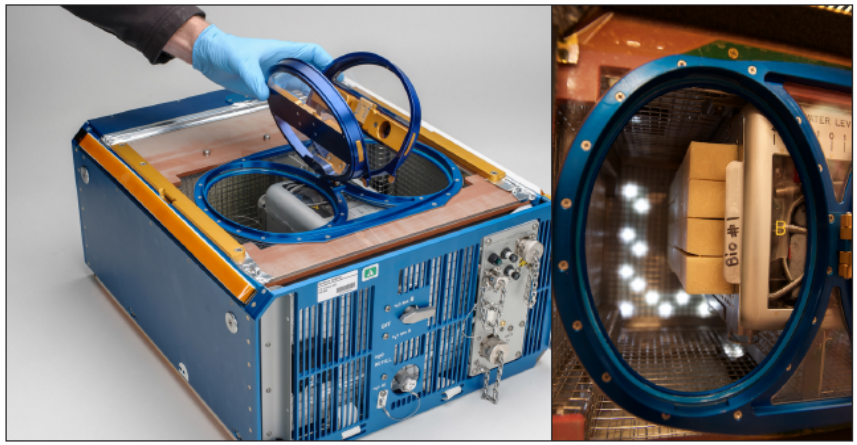




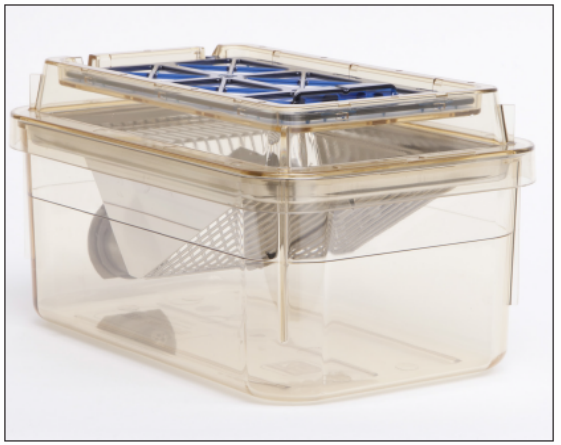
Confounding Factor 1: Cage Effects

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

Cage Types

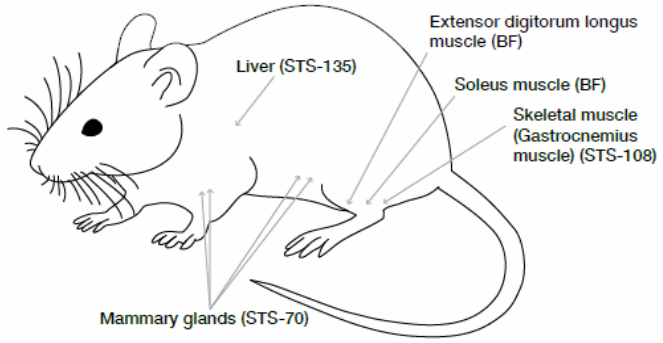


Animal Enclosure Module (AEM)



Sample vivarium cage

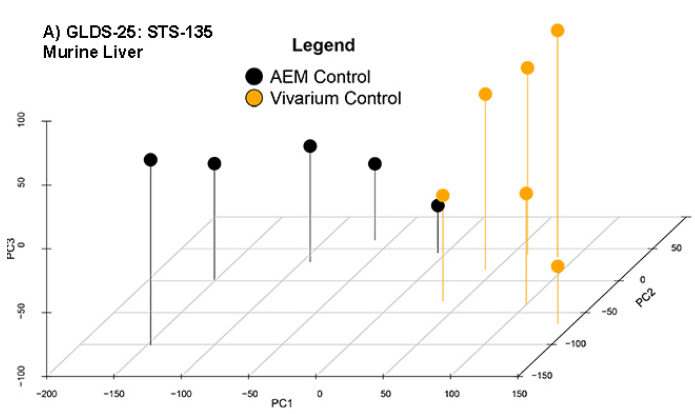
GeneLab study	Mission	Species	CO_2 (ppm)	Tissue type
GLDS-21	STS-108	mouse	~3000	skeletal muscle (gastrocnemius)
GLDS-111	BF	mouse	~600	soleus muscle
GLDS-111	BF	mouse	~600	extensor digitorum longus muscle
GLDS-25	STS-135	mouse	~3000	liver
GLDS-63	STS-70	rat	~3000 (est.)	mammary gland

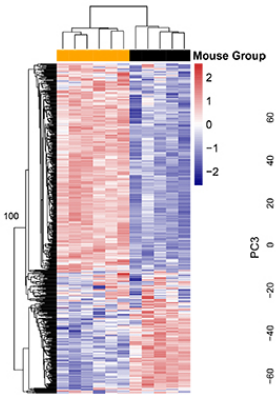


Beheshti et al, **Scientific Reports,** March 2018, DOI: 10.1038/s41598-018-22613-1

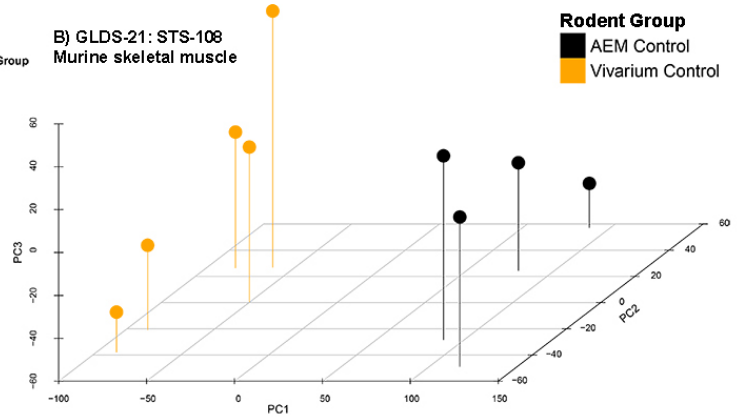
PCA Plots Suggest Strong Cage Effect

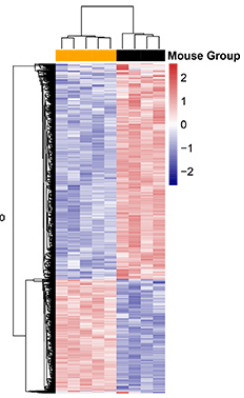
3000 ppm

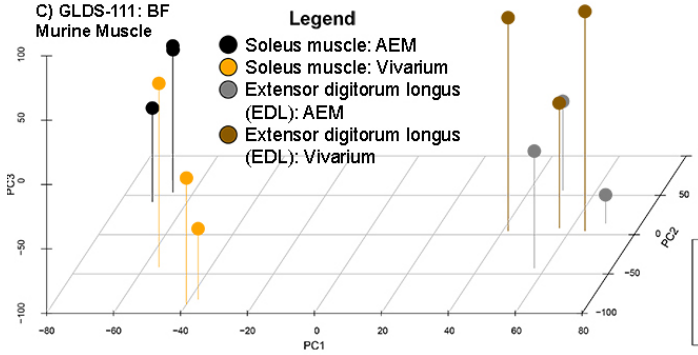


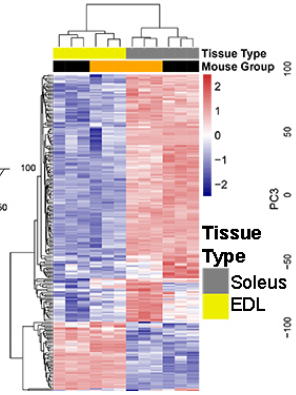


3000 ppm

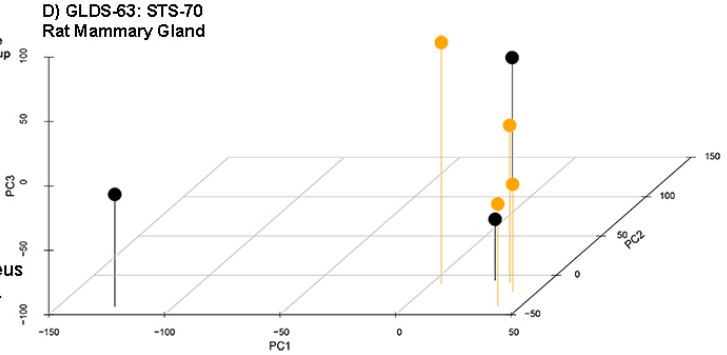


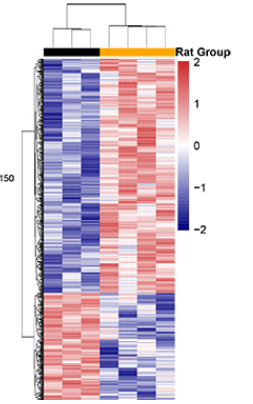






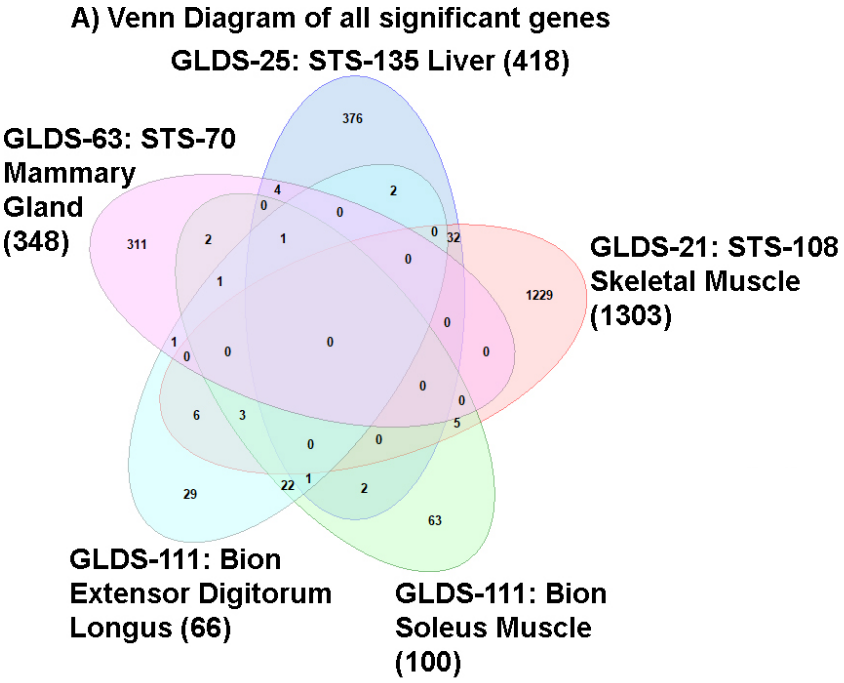
600 ppm





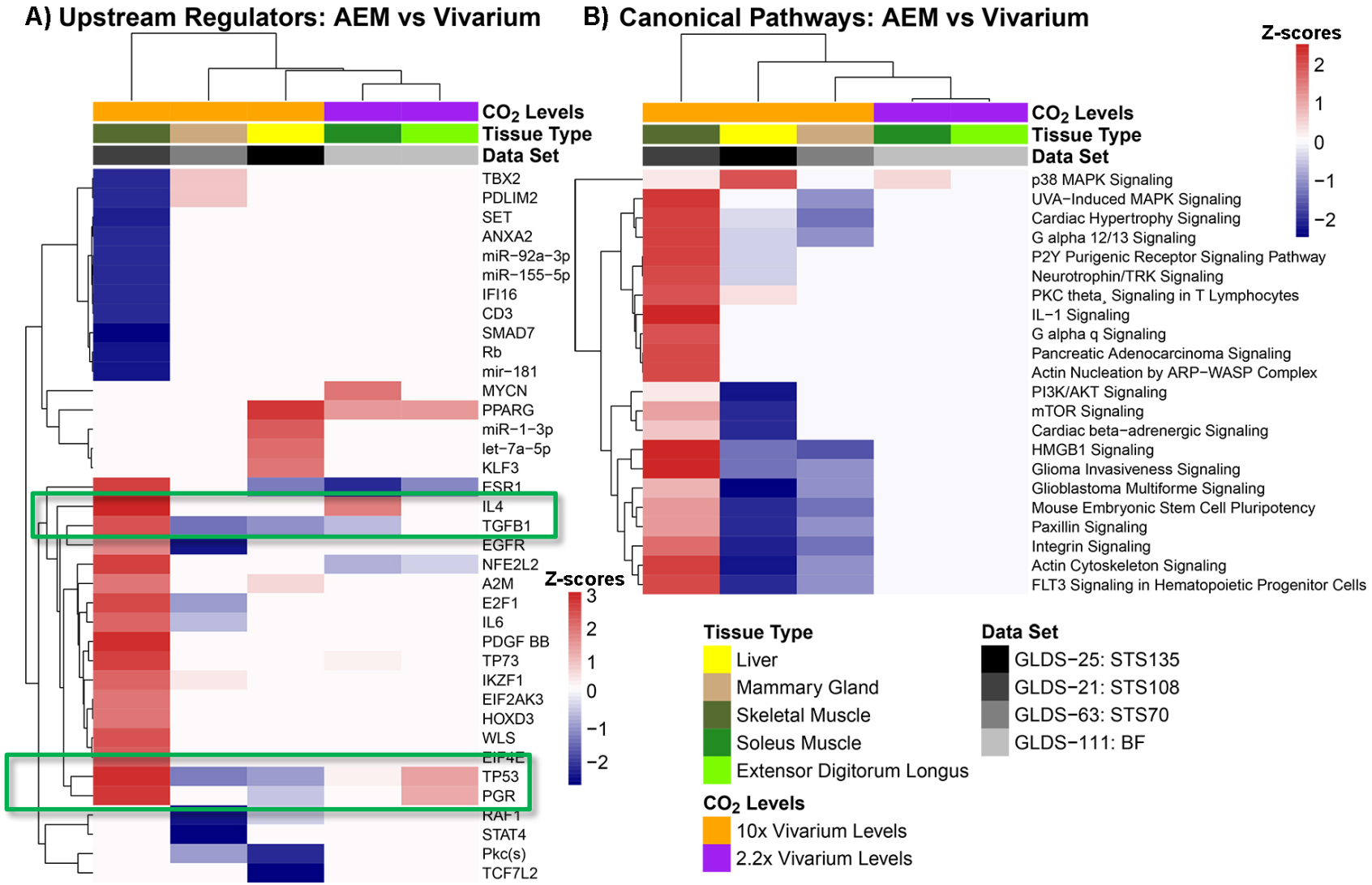
3000 ppm

Differential Gene Expression: Cage or CO2 Effect?



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

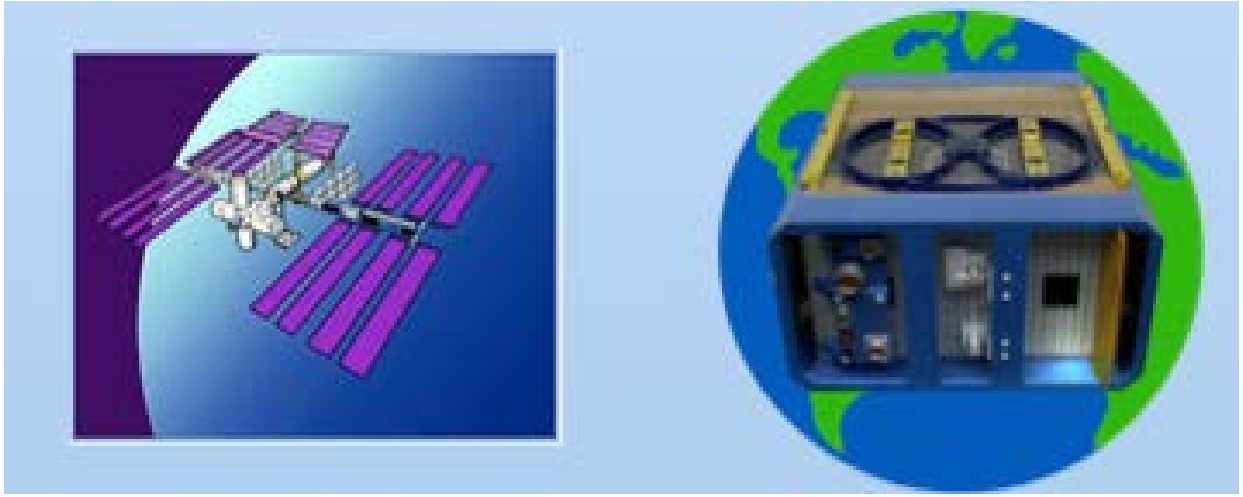
Upstream regulators and canonical pathways show response is tissue specific and highest for high CO2



Mild chronic hypoxia due to increased CO_2 levels could explain both the increase in immune responses and a reduction in metabolism – **Need to confirm with AEM experiments at ambient CO_2 levels.**

Confounding Factor 2: Preservation Methods in Space

Liver collection for RR1



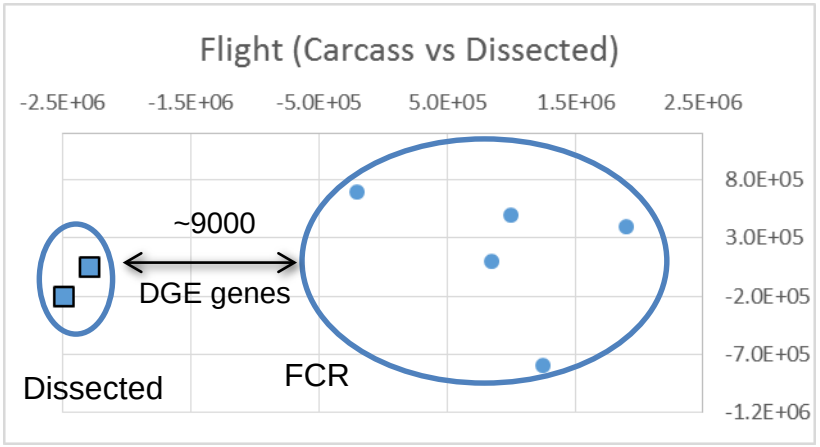
N=2 → Dissected then Frozen

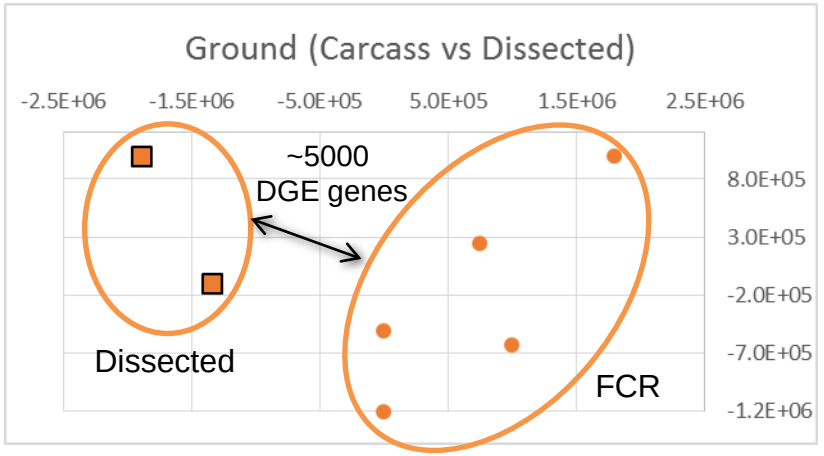
N=5 → Full Carcass frozen

N=2 → Dissected then Frozen

N=5 → Full Carcass frozen

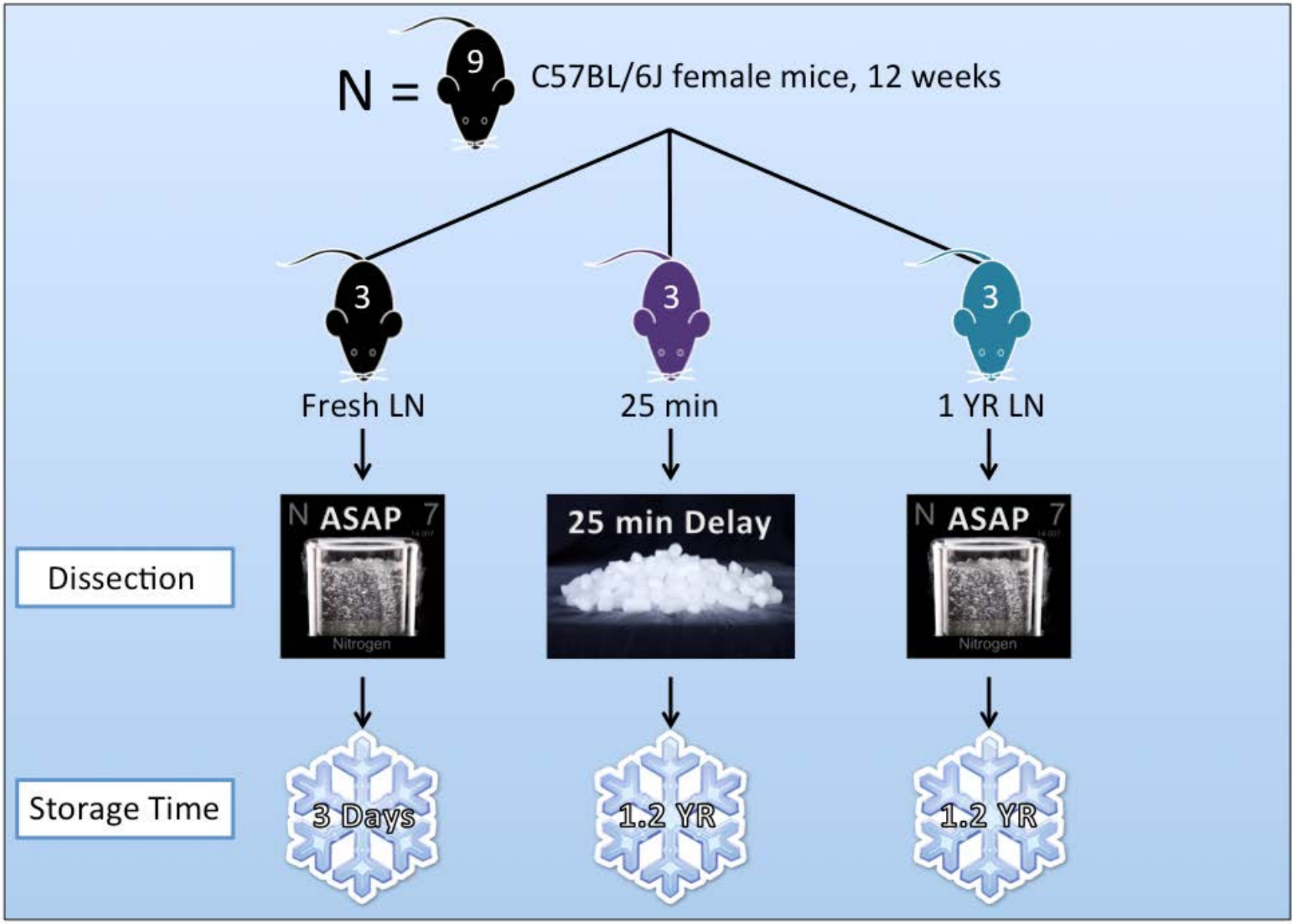
Principal Component Analysis of On-Orbit Dissected vs Frozen Carcass Livers



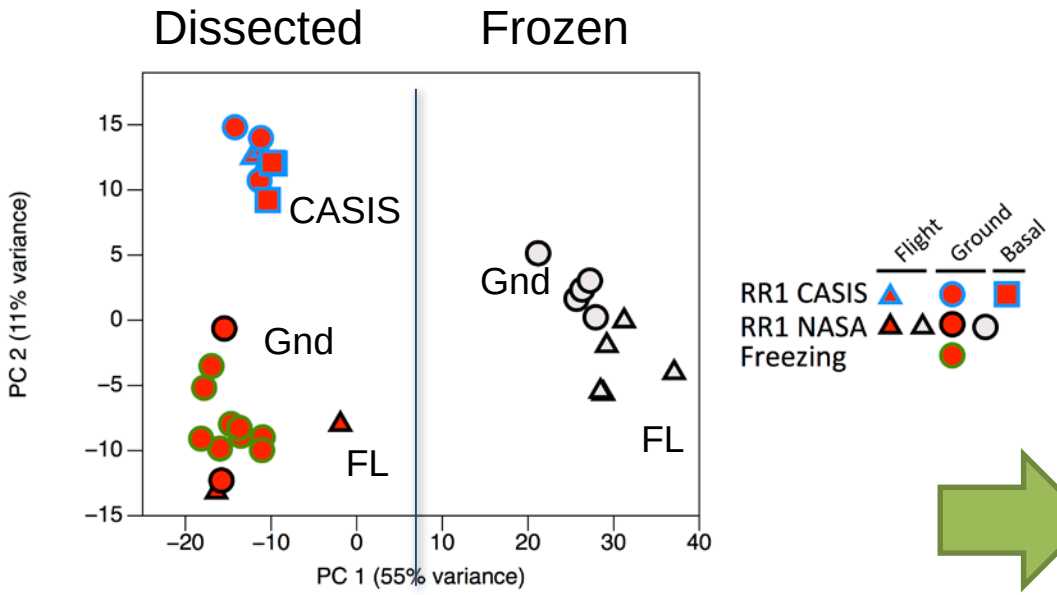


Strong separation of differentially expressed genes between FCR and frozen tissue, either in space or on the ground (worst in space) – 4000 genes in common, principally linked to catabolic pathways (i.e. tissue degradation).

Transcriptomics Data: Pre-validation Experiment



Freezing Before Dissection Changes RNA



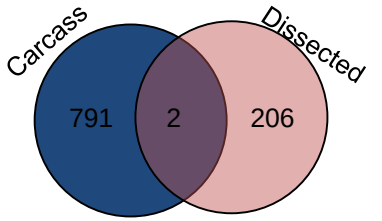
New experimental design to understand:

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?

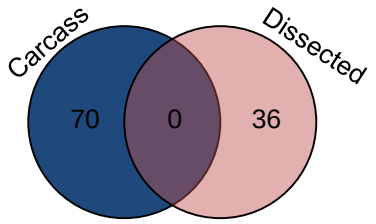
Principal component analysis of liver samples:

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

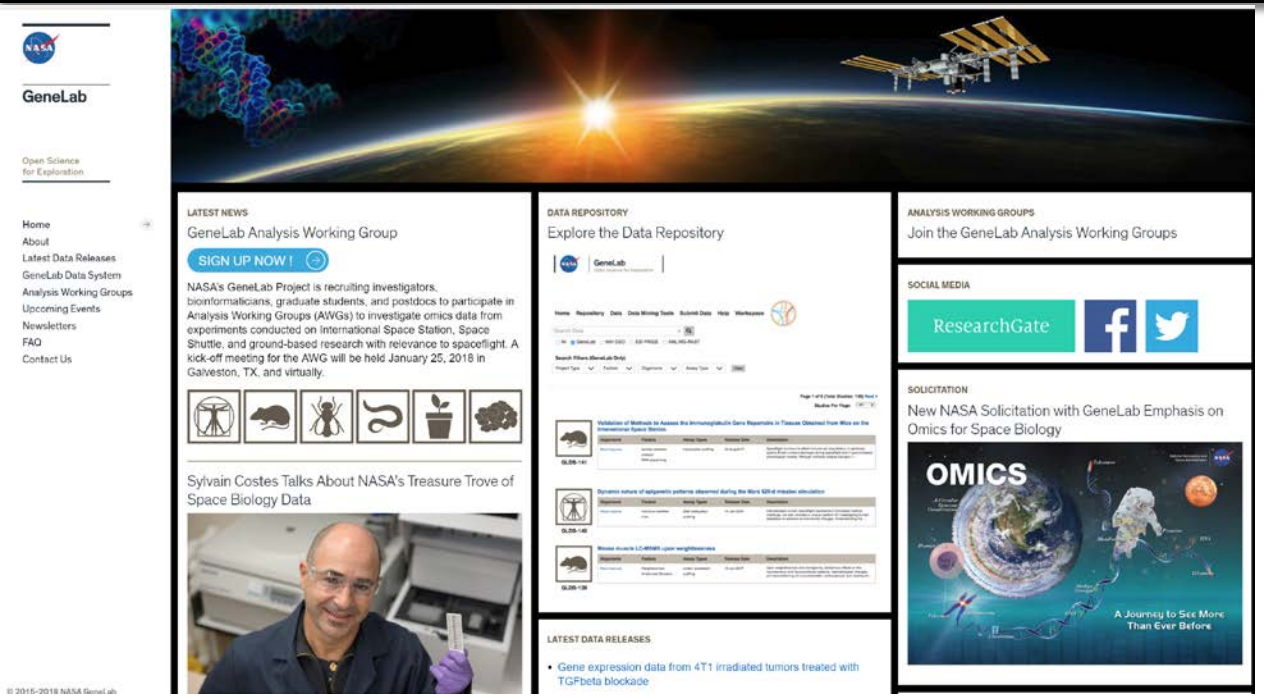
DGE



GSEA



Science Communications

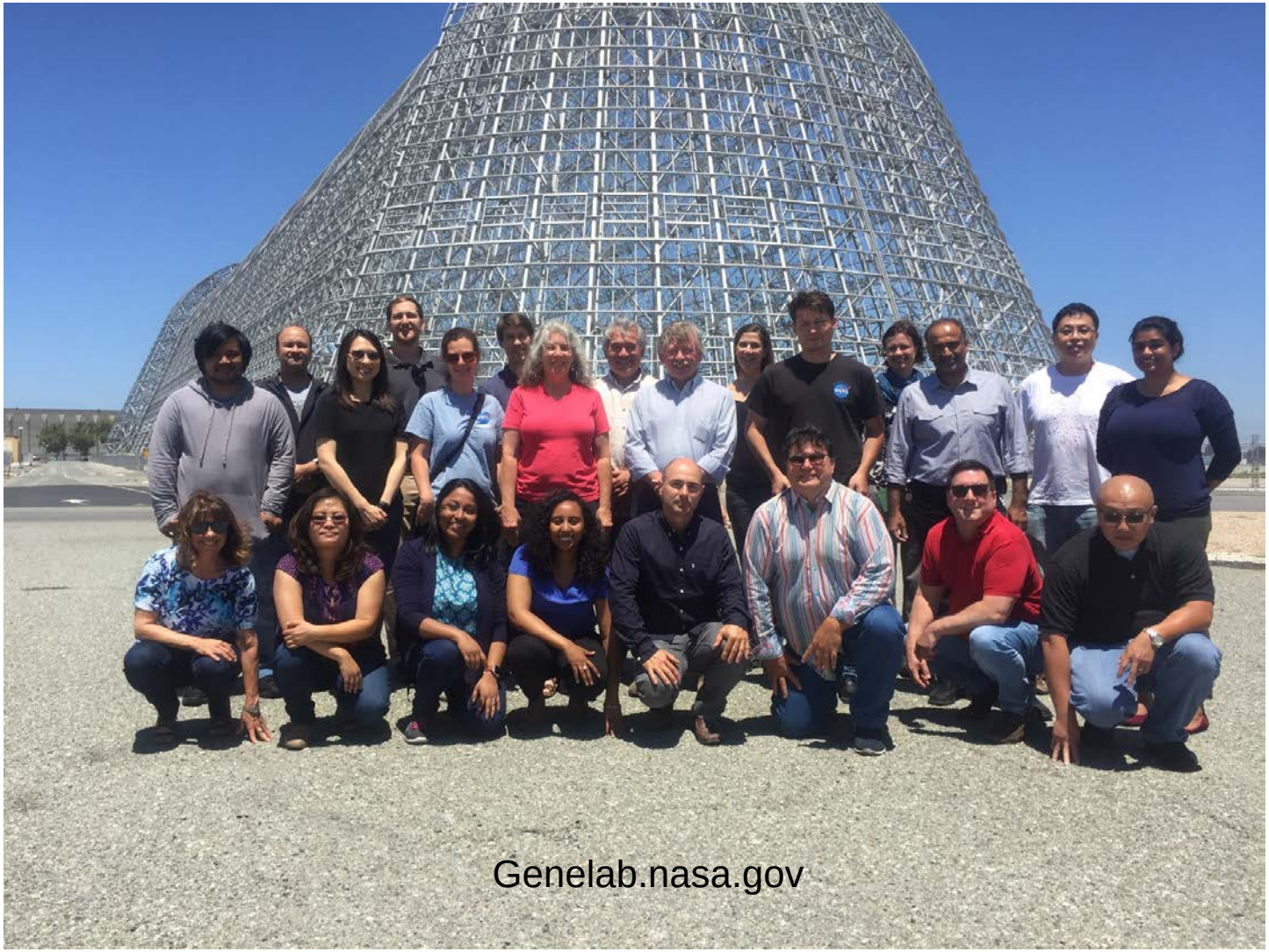


Engage broadest community of researchers, industry, and citizen scientists to advance innovations

https://genelab.nasa.gov

- Weekly social media posts:
  - @NASAAmes **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