GeneLab: Open Science for Exploration

Jonathan Galazka, Ph.D
Project Scientist

The GeneLab Team
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.
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.

Omics Acquisition in Space is Now a Reality

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

Tier 1: Bioinformaticians

Tier 2: Scientific Community

Tier 3: Citizen Scientists

GL4HS

GL Team

Higher Order Data

Physiological Changes
Pathway Enrichment
Differential Expression
Normalization

Visualization Filters
- Ground vs Flight
- Species
- Strains
- Genes/Pathways

Disease Signatures (e.g., Cancer)

New Hypothesis

Input
multi-Omics

Input Reference Dataset

Processing pipelines

Reproducibility

Disease Related Pathways

New Experiments

GeneLab

Disease Signatures (e.g., Cancer)
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

**Phase 1**
Searchable Data
FY2014 –2015

**Phase 2**
Data Exchange
FY2016-2017

**Phase 3**
Tool Integration
FY2018– 2019

**Phase 4**
Maintenance
FY2020 – 2021

---

**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

---

Oct 1, 2017
GLDS 2.0 Release

April 23rd, 2018
AWG workshop

June 2018
Summer interns

Nov 2018
March 2019
Visualization Portal

Oct , 2018
GLDS 3.0 + ASGSR

Oct 1, 2019
GLDS 4.0 Release

---

01-25-2018 AWG KICK-OFF
GLDS Phase 2 (Release 2.0)
Google-like Search, Federated Search

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

Federated Search

Search Filters for GeneLab
User Account Mgmt., Access Controls (e.g., Private, Shared, Public Folders)
GeneLab-GenomeSpace Integration with ISACreator for Streamlining Data Processing Operations

Metadata Source Mappings

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

01-25-2018 AWG KICK-OFF
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

**Data Growth Since 2014**

- 1/31/2016
- 3/6/2017
- 4/10/2018

**Distribution by organism type**

- 32.5%
- 23.75%
- 20.63%
- 5.63%
- 1.88%

**Majority is spaceflight samples**

- 53.06%
- 46.94%

**Distribution by assay type**

- 72.53%
- 7.69%
- 0.55%
- 1.65%
- 3.85%
- 8.79%
- 2.2%
- 75.69%
- 23.31%
69 Ground Data Sets

Types of Ground Data
- Radiation only: 23 (33.33%)
- Microgravity Only: 2 (2.9%)
- Radiation + Microgravity: 18 (26.09%)
- Other: 26 (37.68%)

28 Ground Radiation Data Sets by Organism
- Homo Sapiens: 11
- Mus Musculus: 14
- Rattus norvegicus: 1
- Microbes: 1
- Plant: 1

Irradiation Ground Datasets On GeneLab
LET range: 0.1 to 170 keV/µm

Dose and Dose rate distribution
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
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
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.

![Comparing Key Genes from Meta-Analysis with Genes in Blaber et al Paper](image)
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.
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
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
#1 Risk: Data Reproducibility

\[ \text{Log}(r_m) = 0.2836 \]

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

\[ \text{Log}(R_S) = \text{Log}(r_m) + \text{Log}(E_1) - \text{Log}(E_1) \]
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

- Mice Sacrificed on ISS
- Mice flown on STS and Sacrificed after Re-entry
- Process after mice are sacrificed
- Sample Processing
- Data Sharing
- Data Collection & Curation
- Next Generation Research
- Data Submission
- Modeling and Validation

- Extensor Digitorum Longus Muscle
- Soleus Muscle
- Gastrocnemius Muscle
- Quadriceps
- Tibialis Anterior Muscle
- Adrenal Glands
- Kidney
- Liver
- Skin

- Time in Space for Mice (days)
- Space Shuttle (STS) Missions

- International Space Station (ISS) Missions
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)
Mission-specific analysis: RR-1 Transcriptomics

Samples cluster by tissue type

- Eye, Gastroc, Adrenal
- Liver
- Muscle

Many changes in muscles

Samples do not cluster by flight/ground

Changes to muscle myosin types

Muscle structure

- DE Genes (FDR 5%)
  - Up
  - Down

- Group:
  - Adrenal
  - Extensor
  - Eye
  - Gastroc
  - Kidney
  - Liver
  - Quadriceps
  - Soleus
  - Tibialis
Impact of Microgravity on Liver Tissue: STS135 & RR1 Intersect Venn Diagram Analysis

RR1-Protein  RR1-RNA

541  12  337

RR1-Protein  STS135-RNA

386  167  303

STS135-RNA  RR1-RNA

400  70  279

Mice Sacrificed on ISS

Time in Space for Mice (days)

0  10  20  30  40  50  60  70  80  90  100

Space Shuttle (STS) Missions

Mice flown on STS

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

RR1 Transcriptomics

RR1 Proteomics

STS135 Transcriptomics
<table>
<thead>
<tr>
<th>Pathway</th>
<th>STS-135 RNA</th>
<th>RR1 Protein</th>
<th>RR1 RNA</th>
</tr>
</thead>
<tbody>
<tr>
<td>mmu01100:Metabolic pathways</td>
<td>****</td>
<td>****</td>
<td>****</td>
</tr>
<tr>
<td>mmu01200:Carbon metabolism</td>
<td>****</td>
<td>****</td>
<td>**</td>
</tr>
<tr>
<td>mmu01130:Biosynthesis of antibiotics</td>
<td>****</td>
<td>****</td>
<td>**</td>
</tr>
<tr>
<td>mmu01212:Fatty acid metabolism</td>
<td>****</td>
<td>****</td>
<td>**</td>
</tr>
<tr>
<td>mmu00640:Propanoate metabolism</td>
<td>****</td>
<td>****</td>
<td>*</td>
</tr>
<tr>
<td>mmu00620:Pyrurate metabolism</td>
<td>**</td>
<td>*</td>
<td>*</td>
</tr>
<tr>
<td>mmu00380:Tryptophan metabolism</td>
<td>**</td>
<td>***</td>
<td>**</td>
</tr>
<tr>
<td>mmu00520:Amino sugar and nucleotide sugar metabolism</td>
<td>*</td>
<td>*</td>
<td>*</td>
</tr>
<tr>
<td>mmu00190:Oxidative phosphorylation</td>
<td>****</td>
<td>**</td>
<td>*</td>
</tr>
<tr>
<td>mmu00280:Valine, leucine and isoleucine degradation</td>
<td>****</td>
<td>****</td>
<td>NS</td>
</tr>
<tr>
<td>mmu04146:Peroxisome</td>
<td>****</td>
<td>**</td>
<td>NS</td>
</tr>
<tr>
<td>mmu04141:Protein processing in endoplasmic reticulum</td>
<td>****</td>
<td>*</td>
<td>NS</td>
</tr>
<tr>
<td>mmu00020:Citrate cycle (TCA cycle)</td>
<td>****</td>
<td>**</td>
<td>NS</td>
</tr>
<tr>
<td>mmu03013:RNA transport</td>
<td>****</td>
<td>**</td>
<td>NS</td>
</tr>
<tr>
<td>mmu03010:Ribosome</td>
<td>****</td>
<td>****</td>
<td>NS</td>
</tr>
<tr>
<td>mmu00071:Fatty acid degradation</td>
<td>****</td>
<td>****</td>
<td>NS</td>
</tr>
<tr>
<td>mmu00650:Butanoate metabolism</td>
<td>****</td>
<td>****</td>
<td>NS</td>
</tr>
<tr>
<td>mmu01210:2-Oxocarboxylic acid metabolism</td>
<td>***</td>
<td>***</td>
<td>NS</td>
</tr>
<tr>
<td>mmu00630:Glyoxylate and dicarboxylate metabolism</td>
<td>***</td>
<td>****</td>
<td>NS</td>
</tr>
<tr>
<td>mmu01230:Biosynthesis of amino acids</td>
<td>**</td>
<td>***</td>
<td>NS</td>
</tr>
<tr>
<td>mmu00970:Aminoacyl-tRNA biosynthesis</td>
<td>**</td>
<td>**</td>
<td>NS</td>
</tr>
<tr>
<td>mmu05010:Alzheimer's disease</td>
<td>**</td>
<td>**</td>
<td>NS</td>
</tr>
<tr>
<td>mmu00310:Lysine degradation</td>
<td>**</td>
<td>**</td>
<td>NS</td>
</tr>
<tr>
<td>mmu05012:Parkinson's disease</td>
<td>**</td>
<td>**</td>
<td>NS</td>
</tr>
<tr>
<td>mmu03050:Proteasome</td>
<td>**</td>
<td>**</td>
<td>NS</td>
</tr>
<tr>
<td>mmu00410:beta-Alanine metabolism</td>
<td>**</td>
<td>**</td>
<td>NS</td>
</tr>
<tr>
<td>mmu00920:Sulfur metabolism</td>
<td>**</td>
<td>**</td>
<td>NS</td>
</tr>
<tr>
<td>mmu00270:Cysteine and methionine metabolism</td>
<td>**</td>
<td>*</td>
<td>NS</td>
</tr>
<tr>
<td>mmu00010:Glycolysis / Gluconeogenesis</td>
<td>**</td>
<td>*</td>
<td>NS</td>
</tr>
<tr>
<td>mmu05016:Huntington's disease</td>
<td>*</td>
<td>**</td>
<td>NS</td>
</tr>
<tr>
<td>mmu00072:Synthesis and degradation of ketone bodies</td>
<td>*</td>
<td>**</td>
<td>NS</td>
</tr>
<tr>
<td>mmu00250:Alanine, aspartate and glutamate metabolism</td>
<td>*</td>
<td>**</td>
<td>NS</td>
</tr>
<tr>
<td>mmu00860:Porphyrin and chlorophyll metabolism</td>
<td>*</td>
<td>*</td>
<td>NS</td>
</tr>
<tr>
<td>mmu04932:Non-alcoholic fatty liver disease (NAFLD)</td>
<td>*</td>
<td>*</td>
<td>NS</td>
</tr>
<tr>
<td>mmu01040:Biosynthesis of unsaturated fatty acids</td>
<td>**</td>
<td>NS</td>
<td>*</td>
</tr>
<tr>
<td>mmu04922:Glucagon signaling pathway</td>
<td>**</td>
<td>NS</td>
<td>*</td>
</tr>
<tr>
<td>mmu00061:Fatty acid biosynthesis</td>
<td>**</td>
<td>NS</td>
<td>*</td>
</tr>
<tr>
<td>mmu04710:Circadian rhythm</td>
<td>*</td>
<td>NS</td>
<td>*</td>
</tr>
</tbody>
</table>
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

Ground Control  Flight

Frozen Carcass  Dissected

% change

Ground Control (N=7)  Flight (N=7)

* (p ≤ 0.012)
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

<table>
<thead>
<tr>
<th>GeneLab study</th>
<th>Mission</th>
<th>Species</th>
<th>CO₂ (ppm)</th>
<th>Tissue type</th>
</tr>
</thead>
<tbody>
<tr>
<td>GLDS-21</td>
<td>STS-108</td>
<td>mouse</td>
<td>~3000</td>
<td>skeletal muscle (gastrocnemius)</td>
</tr>
<tr>
<td>GLDS-111</td>
<td>BF</td>
<td>mouse</td>
<td>~600</td>
<td>soleus muscle</td>
</tr>
<tr>
<td>GLDS-111</td>
<td>BF</td>
<td>mouse</td>
<td>~600</td>
<td>extensor digitorum longus muscle</td>
</tr>
<tr>
<td>GLDS-25</td>
<td>STS-135</td>
<td>mouse</td>
<td>~3000</td>
<td>liver</td>
</tr>
<tr>
<td>GLDS-63</td>
<td>STS-70</td>
<td>rat</td>
<td>~3000 (est.)</td>
<td>mammary gland</td>
</tr>
</tbody>
</table>
PCA Plots Suggest Strong Cage Effect

A) GLDS-25: STS-135
Murine Liver

Legend
- AEM Control
- Vivarium Control

B) GLDS-21: STS-106
Murine skeletal muscle

Legend
- Mouse Group

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-83: STS-70
Rat Mammary Gland

Legend
- Rodent Group
- AEM Control
- Vivarium Control

Tissue Type
- Soleus
- EDL
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 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.
Liver collection for RR1

N=2  
Dissected then Frozen

N=5  
Full Carcass frozen

N=2  
Dissected then Frozen

N=5  
Full Carcass frozen
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).
Transcriptomic Data: Pre-validation Experiment

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

- 3 Fresh LN
- 3 25 min
- 3 1 YR LN

Dissection:
- 3 Days
- 1.2 YR
- 1.2 YR

Storage Time:
Freezing Before Dissection Changes RNA

**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

**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?
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](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