NASA’s GeneLab Phase II: Federated Search and Data Discovery

Daniel C. Berrios MD PhD MPH
Sylvain Costes PhD
Peter Tran MS

NASA Ames Research Center
Agenda

- GeneLab
- Federated Search
  - Common Metadata Model
  - Metadata Export
- Next Steps
• Goals

  – An integrated repository and bioinformatics data system for analysis and modeling

  – Enable the discovery and validation of molecular networks that are influenced by space conditions through ground-based and flight research using next-generation omics technologies

  – Engage the broadest possible community of researchers, industry, and the general public to foster innovation

  – Strengthen international partnerships by leveraging existing capabilities and data sharing
Phased Implementation

**Phase 1**
Searchable Data
FY2014 – 2015

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

**Phase 2**
Data Exchange
FY2016-2017

- Link to Public Databases via Metadata Federation
- Integrated Search

**Phase 3**
System Integration
FY2018– 2019

- Provide collaboration framework and tools
- Build Community via collaborative science analysis & modeling
- Provide access to biocomputational tools for omics analysis

Data System

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

Data System

- Link to Public Databases via Metadata Federation
- Integrated Search

Data System

- Provide collaboration framework and tools
- Build Community via collaborative science analysis & modeling
- Provide access to biocomputational tools for omics analysis

Cytoscape
• What?
  – Search using 1 system over multiple data sources
  – For example, Google web search

• Why?
  – Facilitates discovery of data similar to known data
  – Improves search efficiency: no need to switch and search multiple source systems

• How?
  – Metadata Mapping of Data Sources
  – If systems have search interfaces:
    • Dynamic query translation
  – If systems do not have search interfaces, or for greater reliability:
    • Metadata warehousing
**Accession** | **GLDS-131**
--- | ---
**Study Title** | Rodent Research-3-CASIS: **Mouse** liver transcriptomic proteomic and epigenomic data

**Study Description** | The Rodent Research-3 (RR-3) mission was designed to study the effectiveness of a potential countermeasure for the loss of muscle and bone mass that occurs during spaceflight. **Myostatin** is a protein secreted by myoblasts that inhibits muscle cell growth and differentiation. Mutations in myostatin or drugs that block **myostatin** cause increases in muscle mass. The RR-3 experiment was sponsored by pharmaceutical company Eli Lilly and Co. and the Center for the Advancement of Science in Space and assessed the efficacy of **myostatin** inhibition to prevent skeletal muscle atrophy and weakness.…

**Accession** | **GSE466**
--- | ---
**Title** | mRNA expression in regenerated mdx **mouse** skeletal muscle

**Summary** | … A fourfold decrease in **myostatin** mRNA in the mdx muscle was noted. Differential upregulation of actin-related protein 2/3 (subunit 4), beta-thymosin, calponin, mast cell chymase, and guanidinoacetate methyltransferase mRNA in the more benign mdx was also observed. …
Federated Search Example 1

“mouse”
“myostatin”
"mouse" "liver"

Federated Search Example 2

ASGSR 2017

October 2017
• Support development of federated searches initiated using extramural systems
  – Export metadata (and data, if necessary) from GeneLab to these systems
  – Provide link to “authoritative” source data (GeneLab)
• Semi-automated (scripted) process
  – GeneLab metadata used as input
  – Data products for submission to extramural data system are the output
Extramural Federated Search

GeneLab
Open Science for Exploration

NCBI Resources How To

GEO DataSets GEO DataSets → M-CSF space flight
Create alert Advanced

Search results
Items: 7

Evaluation of in vitro macrophage differentiation during space flight
(Submitter supplied) We differentiated mouse bone marrow cells in the presence of recombinant macrophage colony stimulating (M-CSF) factor for 14 days during the flight of space shuttle Transportation System (STS)-126. We tested the hypothesis that the receptor expression for M-CSF, c-Fms was reduced. We used flow cytometry to assess molecules on cells that were preserved during flight to define the differentiation state of the developing bone marrow macrophages, including CD11b, CD31, CD44, Ly6C, Ly6G, F4/80, Mac2, c-Fos as well as c-Fms. more...

Organism: Mus musculus
Type: Expression profiling by array
Platform: GPL1261 4 Samples
Download data: CEL, CHP
Published: Full text in PMC Similar studies: Analyze with GEO2R

Microarray Profile of Gene Expression during Osteoclast Differentiation in Modeled Microgravity
(Submitter supplied) Microgravity leads to a 10-15% loss of bone mass in astronauts during space flight. Osteoclast is the multinucleated bone resorbing cell. In this study, we used NASA developed ground based Rotary Wall Vessel Bioreactor (RWVB), Rotary Cell Culturo System (RCCS) to simulate microgravity (μg) conditions and demonstrated a significant increase (2-fold) in osteoclastogenesis compared to ground based control (Xg) mouse bone marrow cultures. more...

Organism: Mus musculus
Type: Expression profiling by array
Platform: GPL7202 2 Samples

Search details
("macrophage colony-stimulating factor"[Mesh Terms] OR M-CSF[All Fields]) AND ("space flight"[Mesh] OR space)

Find related data
Database: Select

Find items

Summary
We differentiated mouse bone marrow cells in the presence of recombinant macrophage colony stimulating (M-CSF) factor for 14 days during the flight of space shuttle Transportation System (STS)-126. We tested the hypothesis that the receptor expression for M-CSF, c-Fms was reduced. We used flow cytometry to assess molecules on cells that were preserved during flight to define the differentiation state of the developing bone marrow macrophages, including CD11b, CD31, CD44, Ly6C, Ly6G, F4/80, Mac2, c-Fos as well as c-Fms. more...

Organism: Mus musculus
Type: Expression profiling by array
Platform: GPL1261 4 Samples
Download data: CEL, CHP
Published: Full text in PMC Similar studies: Analyze with GEO2R

Overall design
Transcription profiling of 2 total treatment groups and 4 total samples

Web link
https://genelab-data.ncc.nasa.gov/genelab/accession/GLDS-50/

Contributor(s)
Chapin SK

Citation(s)
• Support federated queries initiated using PRIDE, MG-RAST
  – Export metadata to these data systems

• Implement federated searches to other sources
  – MODs
  – NGOs, OGOs

• Expand search capabilities using ontologies (beyond UMLS translations) to increase discovery further
Acknowledgements

Chris Barreras
Afshin Beheshti
Valery Boyko
Sonja Caldwell
Jairon Camarillo
Kaushik Chakravarty
Egle Cekanaviciute
John Costa
Sylvain Costes
Marie Dinh
Sandy Dueck
Homer Fogel
Jon Galazaka
Samrawit Gebre
Dennis Heher
Lynn Hutchison
Yared Kidane
San-Huei Lai Polo
Tristan Le
Qiang Li
Shu-Chun Lin
Debora Pletcher
Sneha Raghunandan
Shayoni Ray
Sigrid Reinsch
Mike Skidmore
David Smith
Marla Smithwick
Olga Stotzky
Hao Thai
Serena Trieu
Andrew Williamson