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
• 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

- Cytoscape

RStudio

October 2017

ASGSR 2017
• 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
Investigation
Study
Assay

GeneLab
Open Science for Exploration

isatools
COLLECT | CURATE | ANALYSE | SHARE | PUBLISH

MG-RAST
metagenomics analysis server

October 2017
ASGSR 2017
### Accession: GLDS-131

<table>
<thead>
<tr>
<th>Study Title</th>
<th>Rodent Research-3-CASIS: Mouse liver transcriptomic proteomic and epigenomic data</th>
</tr>
</thead>
<tbody>
<tr>
<td>Study Description</td>
<td>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.</td>
</tr>
</tbody>
</table>

### Accession: GSE466

<table>
<thead>
<tr>
<th>Title</th>
<th>mRNA expression in regenerated mdx mouse skeletal muscle</th>
</tr>
</thead>
<tbody>
<tr>
<td>Summary</td>
<td>… 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. …</td>
</tr>
</tbody>
</table>
Federated Search Example 1

"mouse" "myostatin"

Myostatin inactivation effects on myogenesis in vitro and in vivo

The transcriptomic signature of myostatin inhibitory influence on the differentiation of mouse C2C12 myoblasts

Development of gene expression signature for defining the cell potency of muscle derived stem cells (MDSC) from mice of different genotypes

Rodent Research-3-CASIS: Mouse liver transcriptomic proteomic and epigenomic data
https://genelab-data.info.nasa.gov/genelab/hicosokn/0905-137
Federated Search Example 2

“mouse”
“liver”

Search results for **mouse liver** using Internet:

Sort by Relevance

<table>
<thead>
<tr>
<th>1 2 3 Next &gt;&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total Results Found: 65</td>
</tr>
</tbody>
</table>

**Rodent Research-3-CASIS: Mouse liver transcriptomic proteomic and epigenomic data**

- **Organism**: Mus musculus
- **Factor**: Microgravity
- **Assay Type**: transcription profiling
- **Accession**: GLDS-137
- **Release/Publication Date**: 23 Aug 2017

**STS-135 Liver Metabolomics**

- **Organism**: Mus musculus
- **Factor**: Microgravity
- **Assay Type**: metabolite profiling
- **Accession**: GLDS-108
- **Release/Publication Date**: 28 Apr 2017

**Female mouse liver quantitative analysis**

- **Organism**: Mus musculus (Mouse)
- **Accession**: PXD000296
- **Release/Publication Date**: 09 Jul 2016

**The comparative proteomics research of Toxoplasma infection in mice liver**

- **Organism**: Mus musculus (Mouse)
- **Accession**: PXD003399
- **Release/Publication Date**: 31 Mar 2016
• 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

**GEO DataSets**

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

**Search results**

**Items:** 7

1. [Evaluation of in vitro macrophage differentiation during space flight](https://genelab-data.nrc.nasa.gov/genelab/accession/GLDS-50/)

   **(Submitted 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 the space shuttle Space 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-Fas as well as c-Fms.

   **Organism:** Mus musculus

   **Exp. type:** Expression profiling by array

   **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 the space shuttle Space 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-Fas as well as c-Fms. RNA was preserved during the flight and was used to perform microarray. We found that there were significant differences in the macrophages that developed in space compared to controls maintained on Earth. We found that there were significant changes in the distribution of expressed CD11b, CD31, F4/80, Mac2, Ly6C and c-Fos. However, the changes in c-Fms expression and no consistent pattern of advanced macrophage differentiation during space flight. We also found a pattern of transcript would be consistent with a relatively normal differentiation outcome but proliferation by the bone marrow macrophages that were assessed after space flight. There was also a surprising pattern of space flight influence on the expression of macrophages from mouse bone cells.

   **MESH:** Space Flight/Macroage

   **Overall design:** Transcript profiling of 2 total treatment groups and 4 total samples

   **Web link:** https://genelab-data.nrc.nasa.gov/genelab/accession/GLDS-50/
Next Steps/Directions

• 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