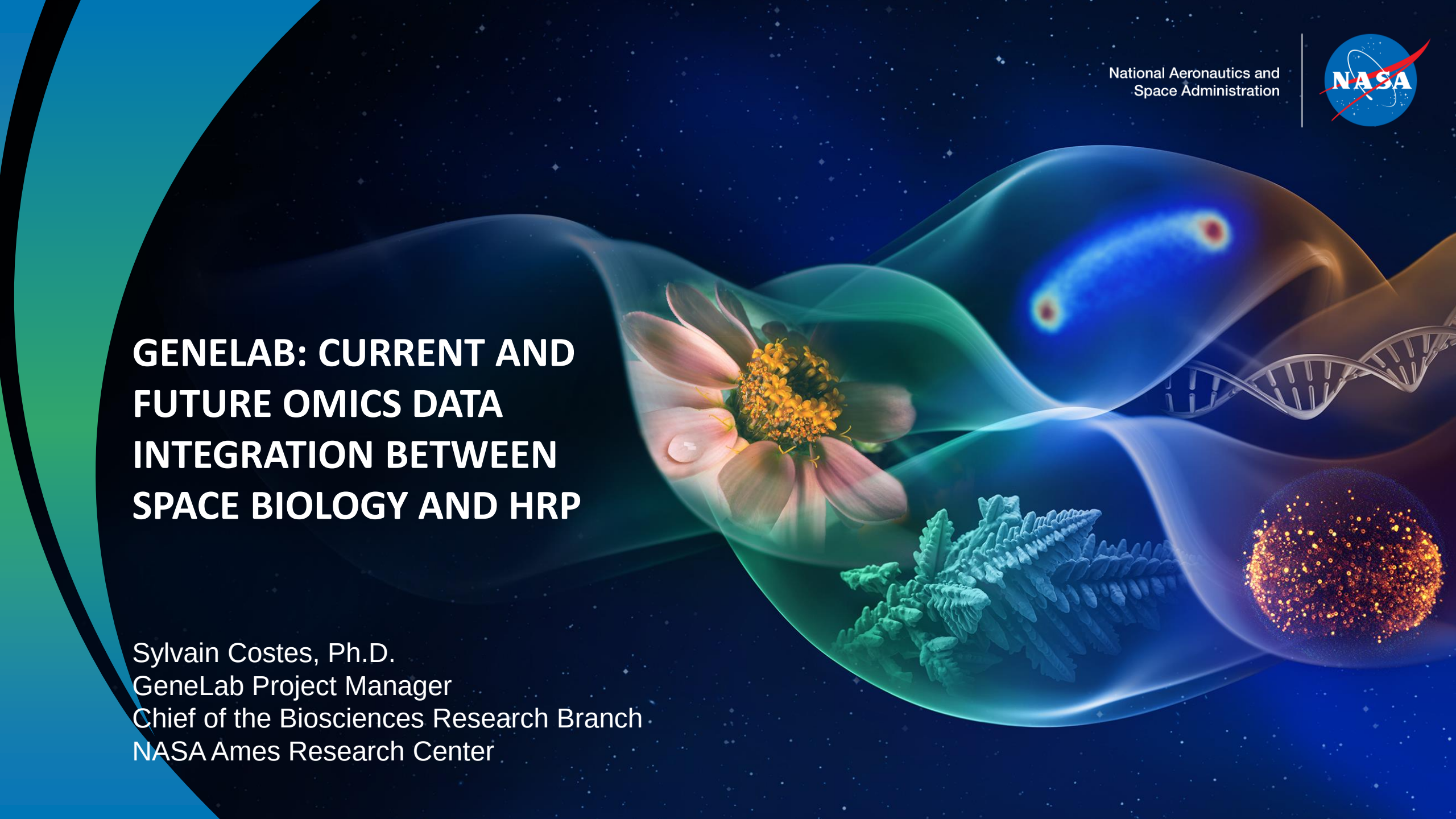
National Aeronautics and Space Administration



GENELAB: CURRENT AND FUTURE OMICS DATA INTEGRATION BETWEEN SPACE BIOLOGY AND HRP

Sylvain Costes, Ph.D.
GeneLab Project Manager
Chief of the Biosciences Research Branch
NASA Ames Research Center

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

and unlock discoveries not possible on Earth.

Reaction tube containing lyophilized chemical assay bead (proprietary)

Cepheid Smart Cycler qRT-PCR

Mini-PCR

Open Science

"We define open science as a collaborative culture enabled by technology that empowers the **open sharing of data, information, and knowledge** within the scientific community and the wider public to accelerate scientific research and understanding."

Ramachandran, R., Bugbee, K., & Murphy, K. J. Moving from Open Data to Open Science. Earth and Space Science, Wiley Publication https://doi.org/10.1029/2020EA001562

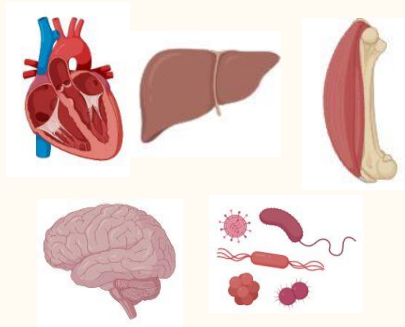
NASA Biological Open Science Resources

Biospecimen Sharing Program (BSP)

- *Dissects and preserves* rodent tissues from Flight and Ground investigations
- *Coordinates* internal tissue sharing

NASA Biological Institutional Scientific Collection (NBISC)

- *Collection* of non-human specimens and space microbial culture



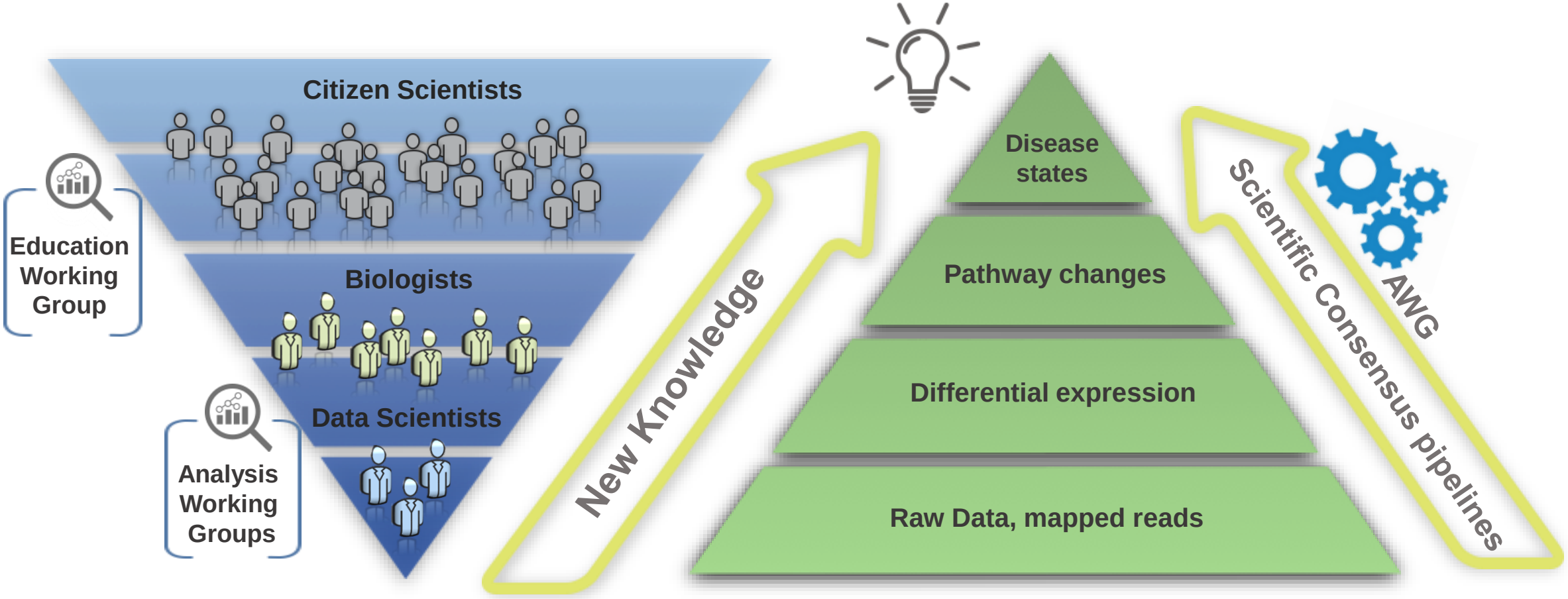
Ames Life Sciences Data Archive (ALSDA)

- *Collects and curates phenotypic research,* mission, project and imaging data

GeneLab (GL)

- *Collects and curates* omics data

GeneLab Omics Data Democratization



GeneLab Power Users

GeneLab Analysis Working Groups (AWGs) *consist of* ***200+ scientists*** *from multiple space agencies, international institutions, and industry. Scientists meet monthly with each group to analyze data in the GeneLab repository.*
We invite you to join - https://genelab.nasa.gov/awg/join!

ANIMAL

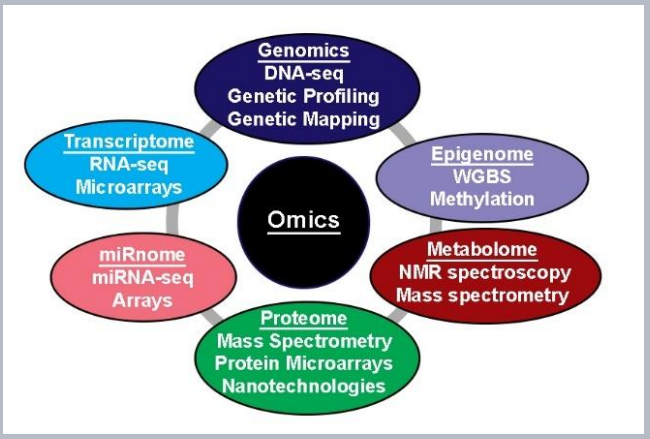
68 members

Facilitates the use of omics in understanding basic mechanisms by which animals and constituent tissues and cells adapt to the spaceflight environment.

MULTI-OMICS

128 members

Interactions between the different omics to provide complete understanding of the entire system begin studied.



PLANTS

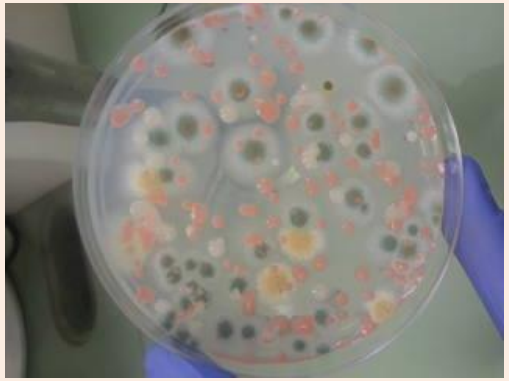
58 members

Share and discuss the latest developments in ***AstroBotany*** *– the discipline of botany concerned with interactions between plant biology and space environment.*

MICROBES

50 members

Focuses on analyzing microbial datasets within GeneLab that includes gene-expression, proteomic, metabolomic and environmental metagenomic datasets.



Opportunities for Students & Educators

GeneLab for High Schools (GL4HS): A four-week intensive training summer program for rising high school juniors and seniors to learn bioinformatics and computational biology methods and techniques to analyze space omics data.

Learn more and apply at: https://www.nasa.gov/ames/genelab-for-high-schools (Applications open!)

GeneLab for Colleges/Universities (GL4U):

For educators and students to learn how analyze omics data using GeneLab standard pipelines and space-relevant data

Access to course materials: https://github.com/nasa/GeneLab-Training/tree/main/GL4U

SLSTP: Provides undergraduate students entering their junior or senior years, and entering graduate students, with professional experience in space life science disciplines including:

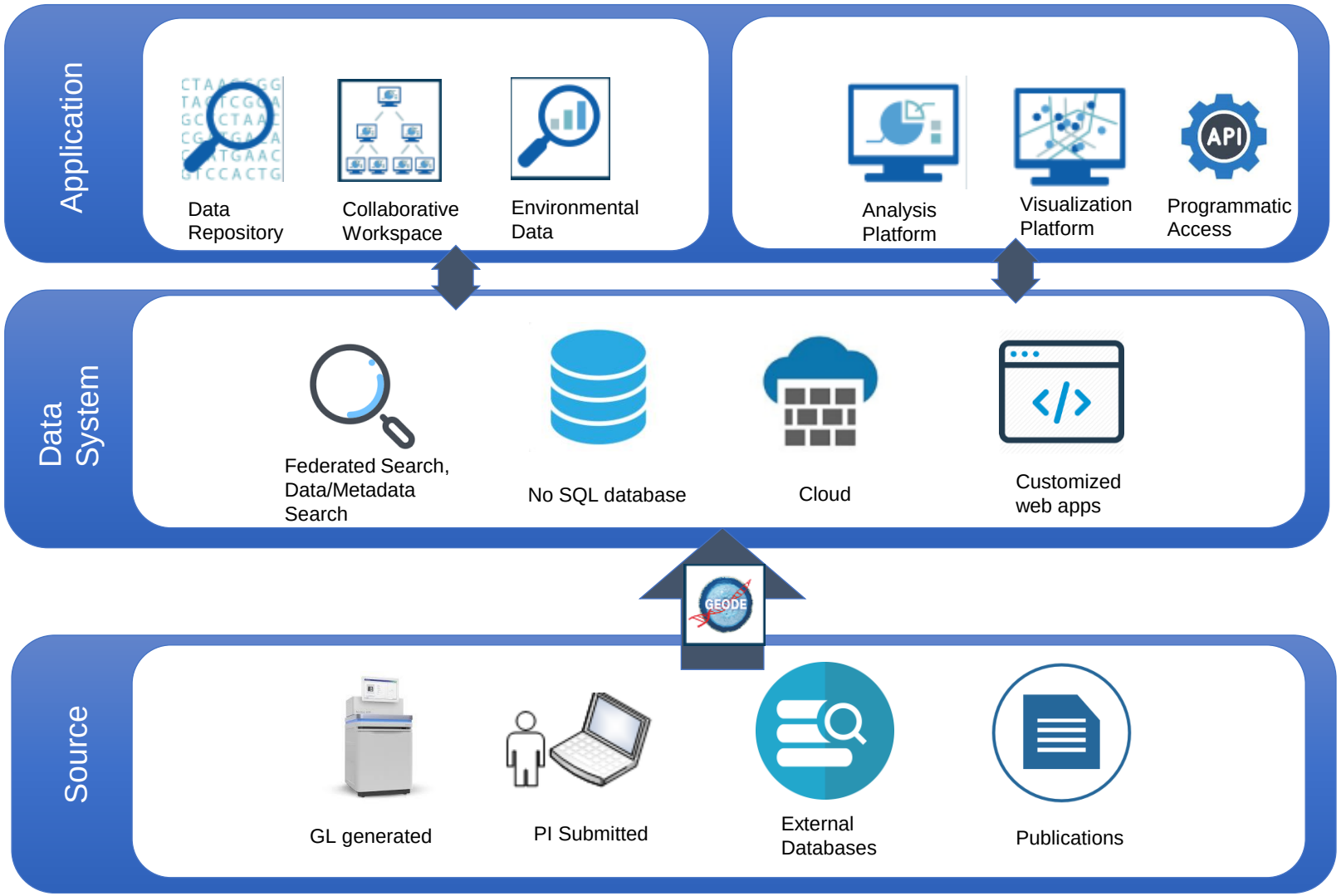
https://www.nasa.gov/ames/research/space-life-sciences-training-program

NASA BPS Open Science Enterprise Solution: genelab.nasa.gov

- Open access data
- **FAIR (Findable, Accessible, Interoperable, Reusable)**
- Controlled access tools
- API - internal and external
- User Friendly Interface
- Tutorials
- Self-service Submission Portal

- Federated search – GEO, PRIDE, MG-RAST, **ALSDA**
- Database & Cloud – Scalable, easy access, fast
- Web apps
  - Data Access & Management
  - Security
  - Operation
  - Governance and Integration
- Open Source software – *no maintenance cost for software*

- Multiple data sources
  - Standard metadata organization
  - Open file formats



GeneLab Data Metrics

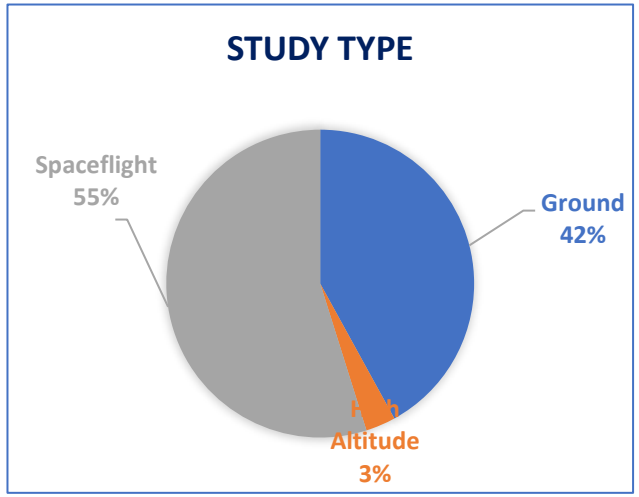
352
Studies

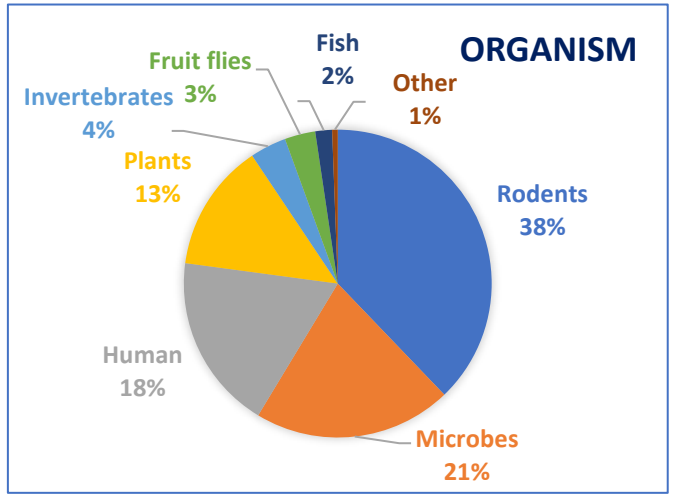
400
Datasets

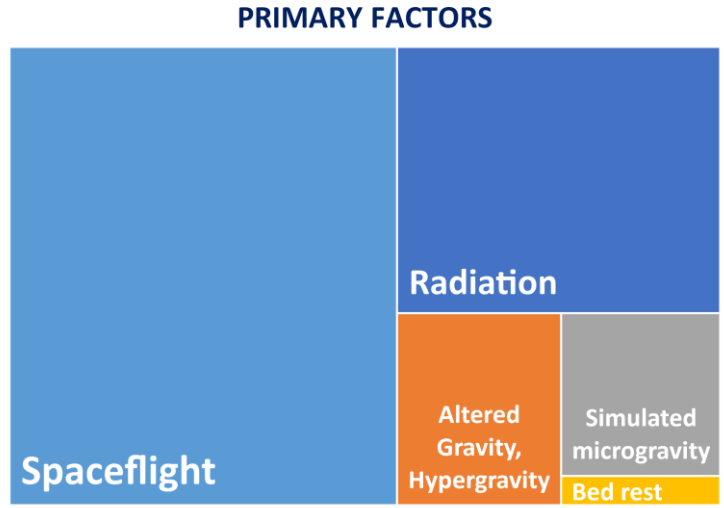
45
Species

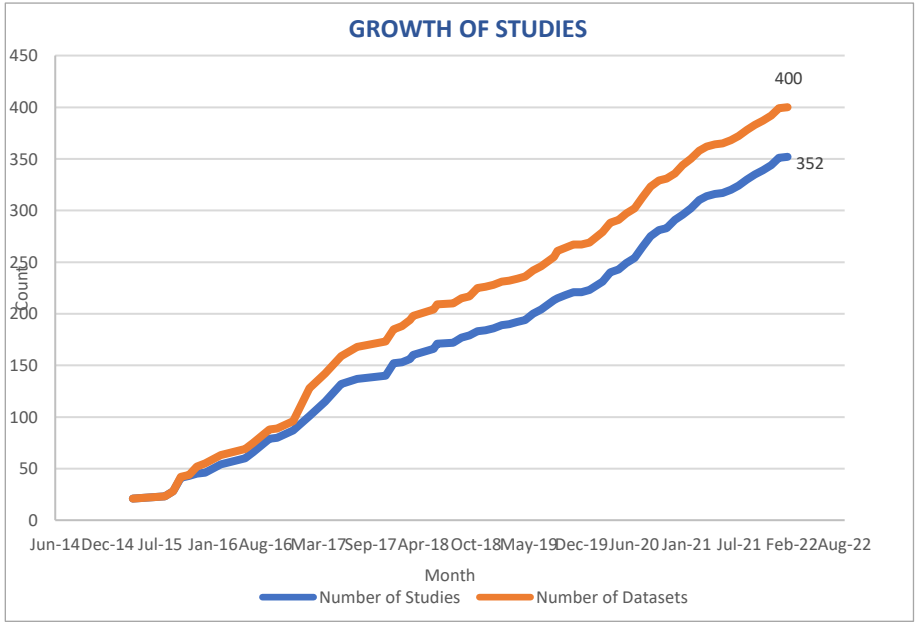
>10
Assays

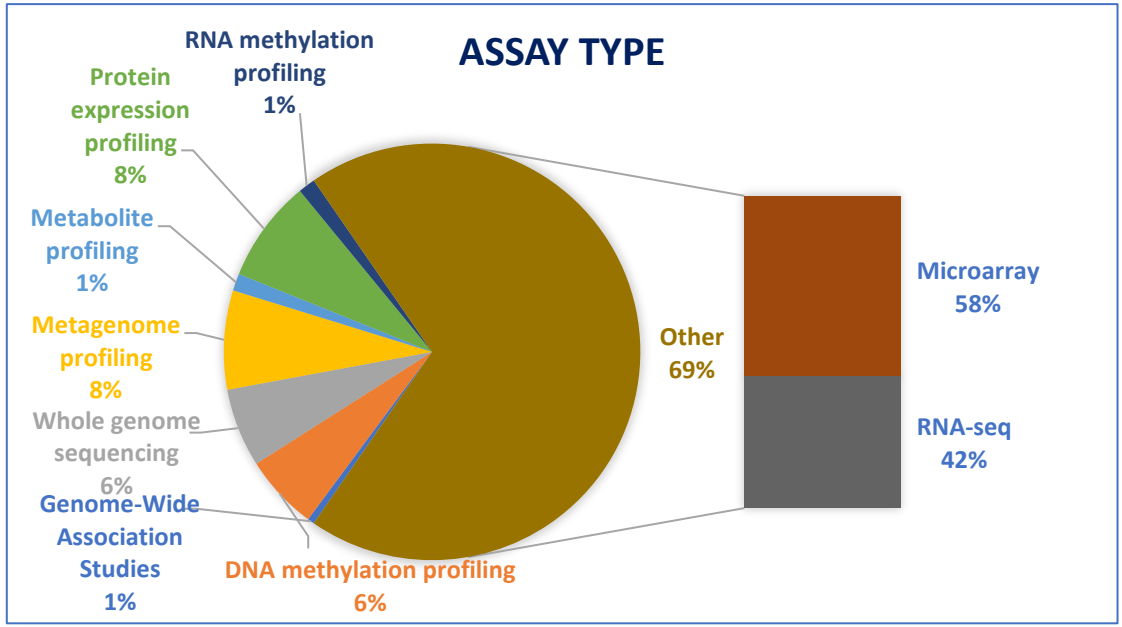
>135TB
Data











Re-use of Data and Enabling New Discoveries

38 enabled publications (**10 publications produced by the AWGs**) using data available in GeneLab.

Mammalian and Invertebrate Models as Complementary Tools for Gaining Mechanistic Insight on Muscle Responses to Spaceflight

by Thomas Cahill [1], Henry Cope [2], Joseph J. Bass [3], Eliah G. Overbey [4], Rachel Gilbert [5,6], Willian Abraham da Silveira [1,7], Amber M. Paul [5,8,9], Tejaswini Mishra [10], Raúl Herranz [11], Sigrid S. Reinsch [5], Sylvain V. Costes [5], Gary Hardiman [1,12], Nathaniel J. Szewczyk [3,13] and Candice G. T. Tahimic [5,14,*]

ELSEVIER

Computational and Structural Biotechnology Journal

Volume 19, 2021, Pages 2223-2235

Reanalysis of the Mars500 experiment reveals common gut microbiome alterations in astronauts induced by long-duration confinement

N.J.B. Brereton [a,1], F.E. Pitre [a,1], E. Gonzalez [b,c,1]

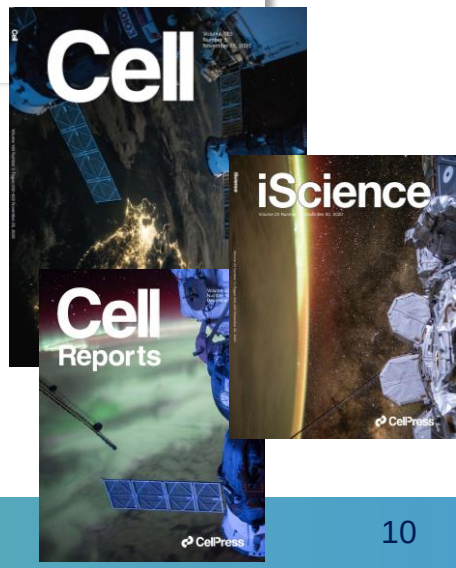
An Integrative Network Science and Artificial Intelligence Drug Repurposing Approach for Muscle Atrophy in Spaceflight Microgravity

Vidya Manian*, Jairo Orozco-Sandoval and Victor Diaz-Martinez

Rad-Bio-App: a discovery environment for biologists to explore spaceflight-related radiation exposures

Richard Barker, Sylvain V. Costes, Jack Miller, Samrawit G. Gebre, Jonathan Lombardino & Simon Gilroy

npj Microgravity **7**, Article number: 15 (2021) | Cite this article

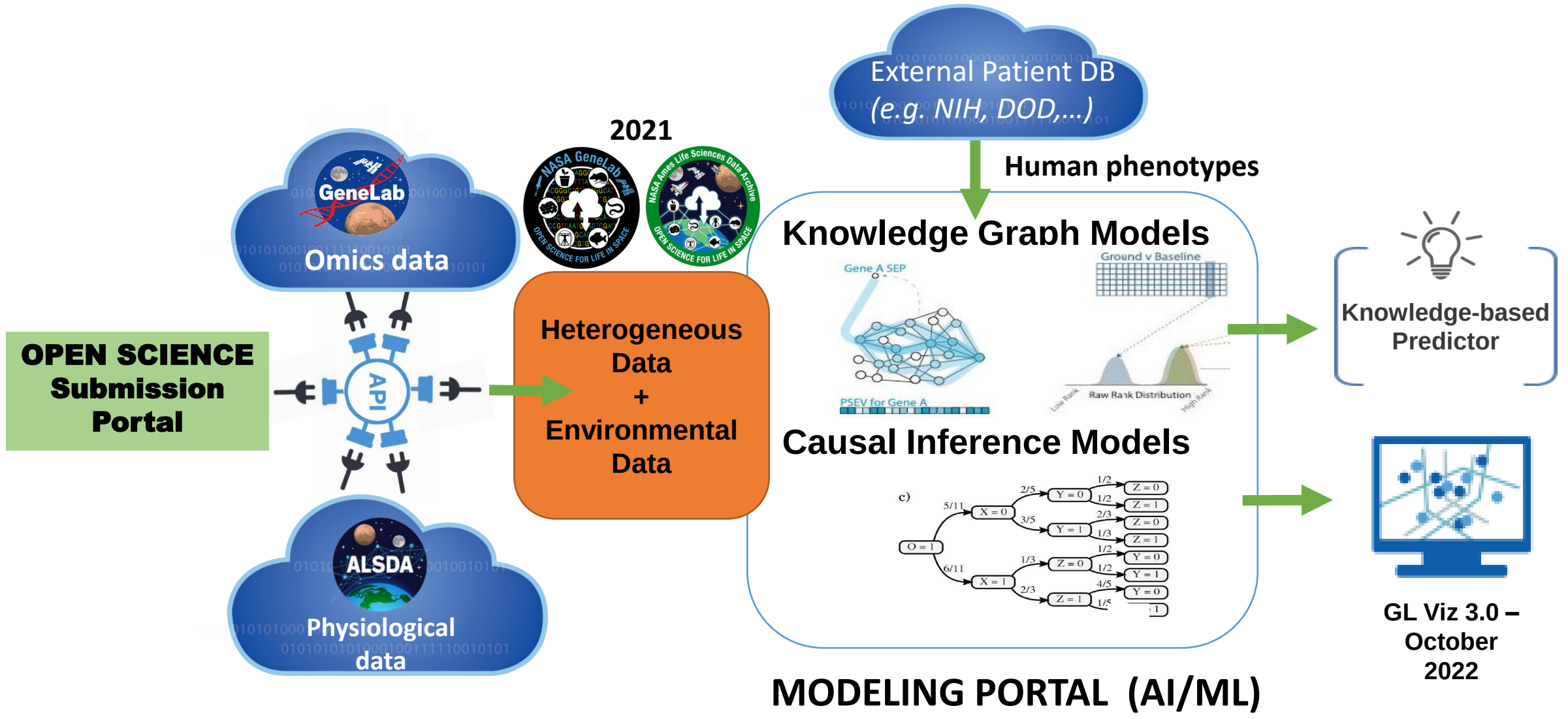


Cell: The biology of spaceflight package

A coordinated package of 29 scientific papers published in five *Cell Press* journals featuring **9** papers utilizing data or resources in GeneLab. *High impact research highlights:*

- Comparative Transcriptomics Identifies Neuronal and Metabolic Adaptations to Hypergravity and Microgravity in Caenorhabditis elegans, *iScience*
- Comprehensive Multi-omics Analysis Reveals Mitochondrial Stress as a Central Biological Hub for Spaceflight Impact, *Cell*
- NASA GeneLab RNA-seq consensus pipeline: standardized processing of short-read RNA-seq data, *iScience*

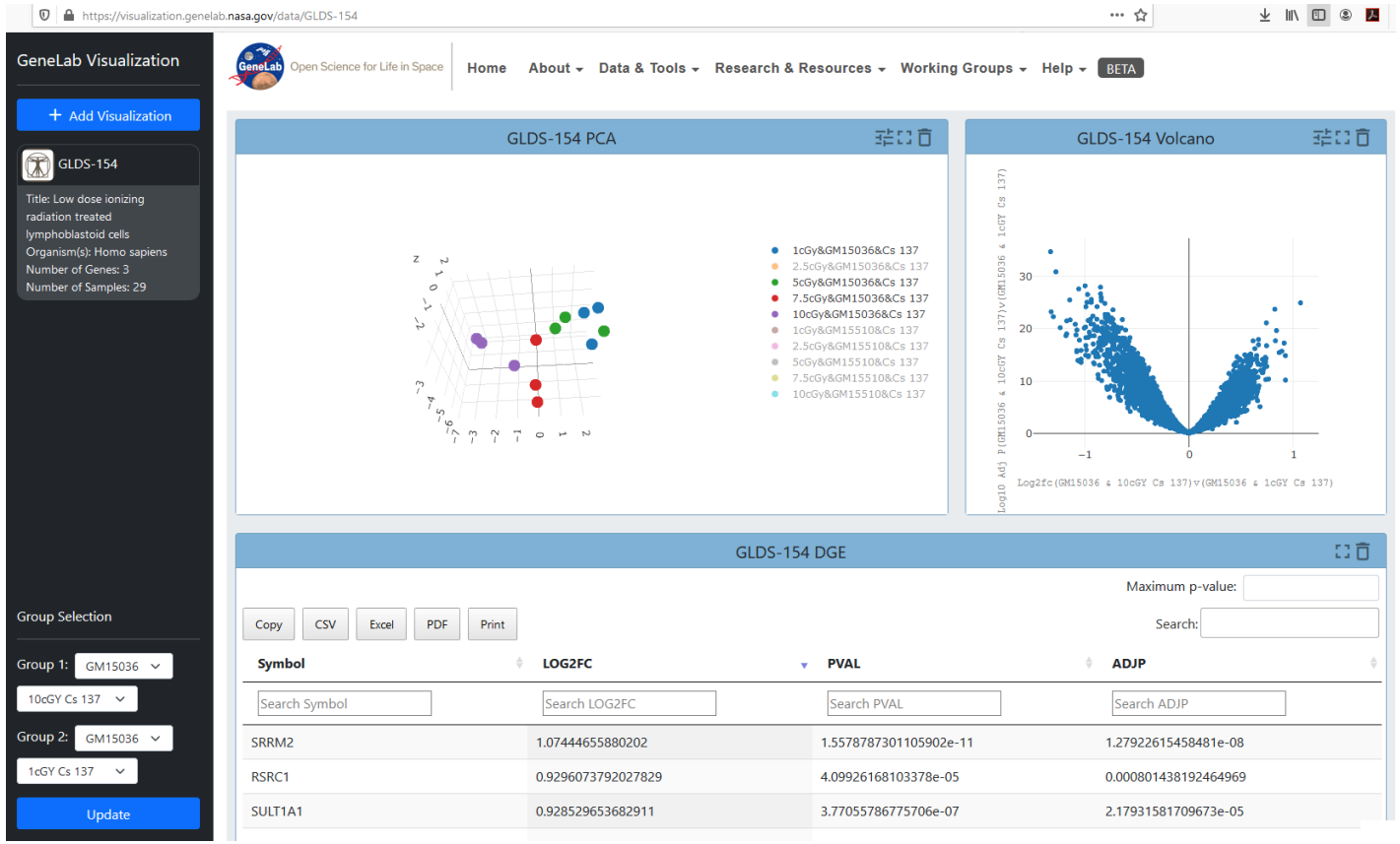
From multiple databases to a knowledge-based system

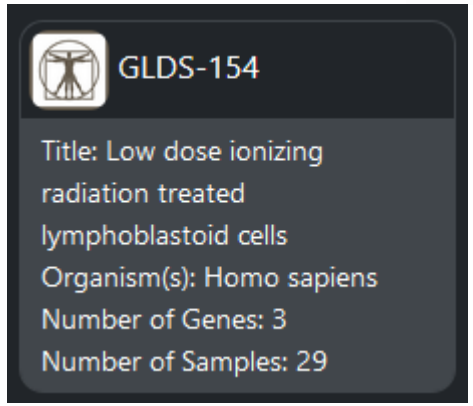


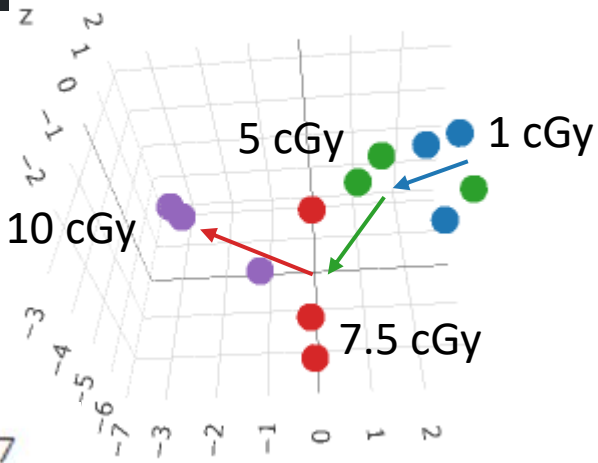
Increasing accessibility of complex data

Application

Visualization Platform







Wu P, Coleman M, Wyrobek AJ. "Low dose ionizing radiation treated lymphoblastoid cells", GeneLab, Version 3, http://doi.org/10.26030/hs0p-6w85

GeneLab radiation data: enabling AI/ML meta-analysis

AI4LS
Artificial Intelligence for Life in Space

6 individual radiation exposure gene expression datasets:

GLDS-71: Immediate Transcriptional Changes in Response to High Dose Radiation Exposure Version 4

0.3 Gy – Cs-137

GLDS-152: Transcription profiling of human peripheral blood to development gene expression signatures for practical radiation biodosimetry Version 2

0.5 Gy – gamma ray

GLDS-156: Identifying radiation exposure biomarkers from mouse blood transcriptome Version 1

2 Gy – Cs-137

GLDS-157: Gene expression in human peripheral blood 48 hours after exposure to ionizing radiation Version 2

0.5 Gy – gamma ray

GSE124612: Transcriptomic responses in mouse blood during the first week after in vivo gamma irradiation

1.5 Gy – gamma ray

GSE62623: Gene expression in mouse blood following low dose-rate or acute x-ray exposure

1.1 Gy – X ray

Data Pipeline:

Merge datasets

↓

Gene-homologue mapping

↓

Data normalization

↓

Combined cross-organism radiation exposure dataset

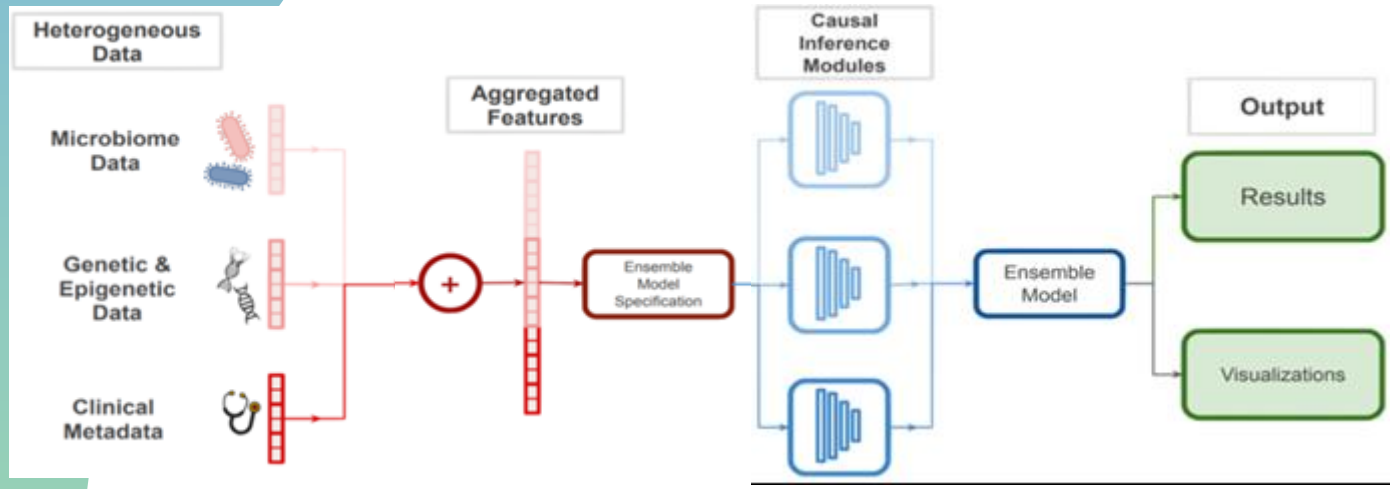
Combined Dataset:

- ~25,000 human-mouse gene homologues
- 455 total samples

- Gamma irradiation or non-irradiated controls:
  - *ex vivo* irradiation of human blood samples
  - *in vivo* irradiated mouse whole blood

Causal inference in complex biological data

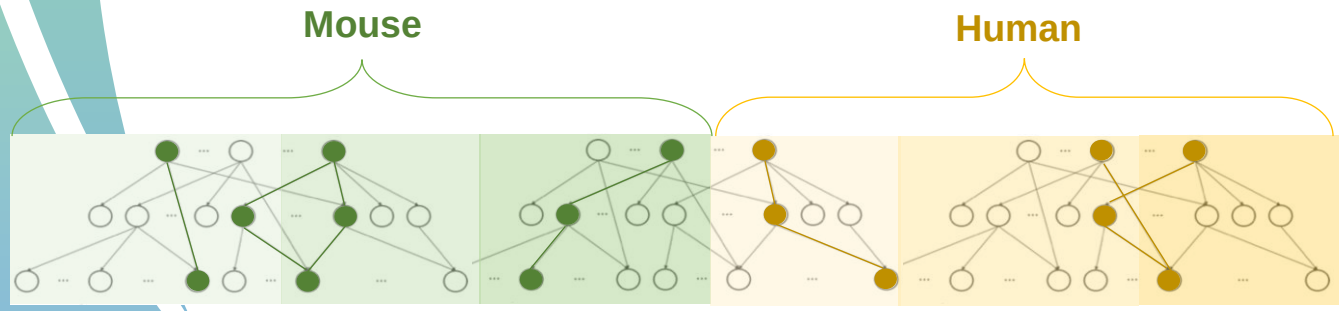
CRISP Platform Overview



Generalizable to different data types with the ability to learn causal drivers of any feature within a dataset

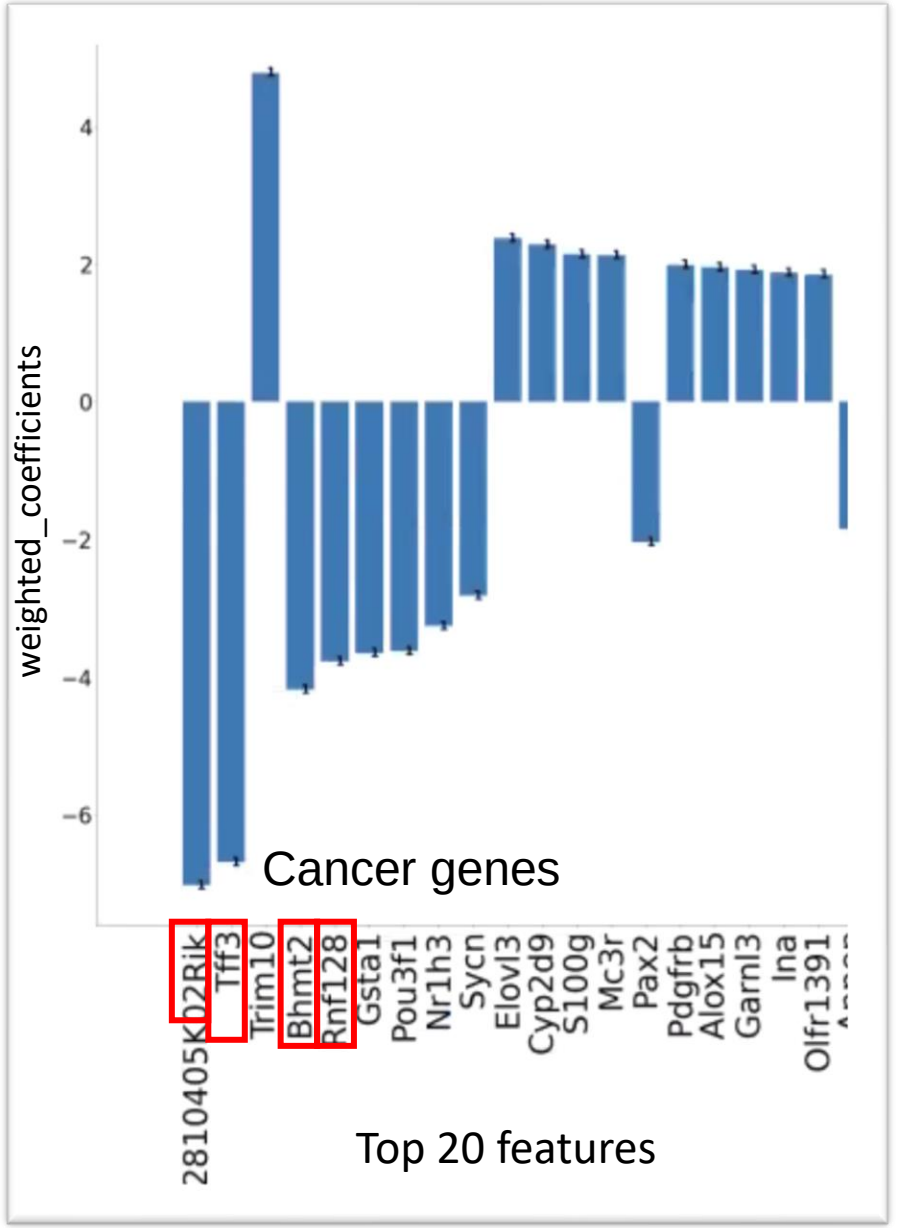
Automated dimensionality reduction using machine learning selects the most important input variables for causal analysis

An ensemble of causal inference models “votes” on the selection of causal features



TARGET: radiation state (irradiated or control)

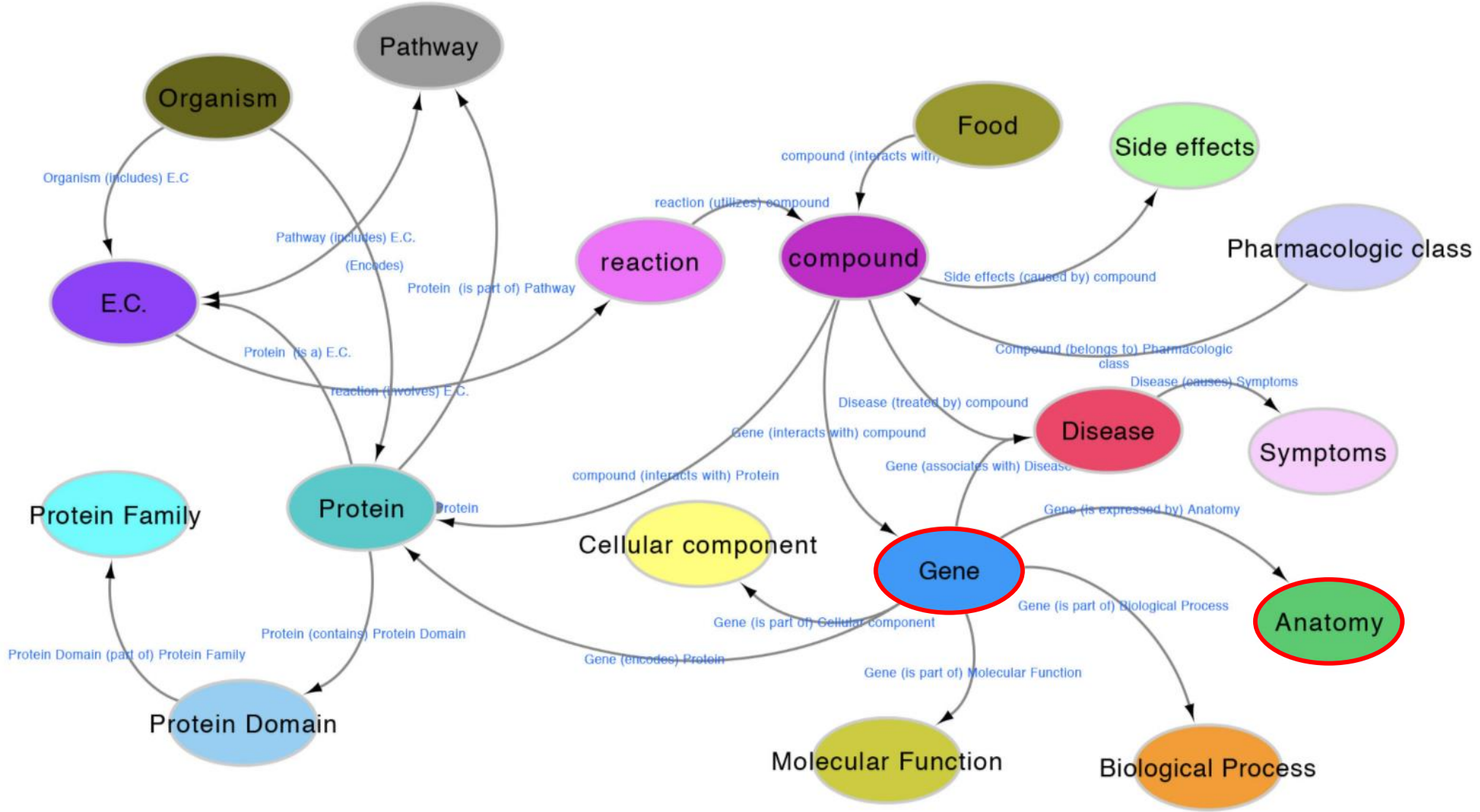
CRISP identifies features that are *causal* of a target variable



Developed by FDL Astronaut Health 2020 Team

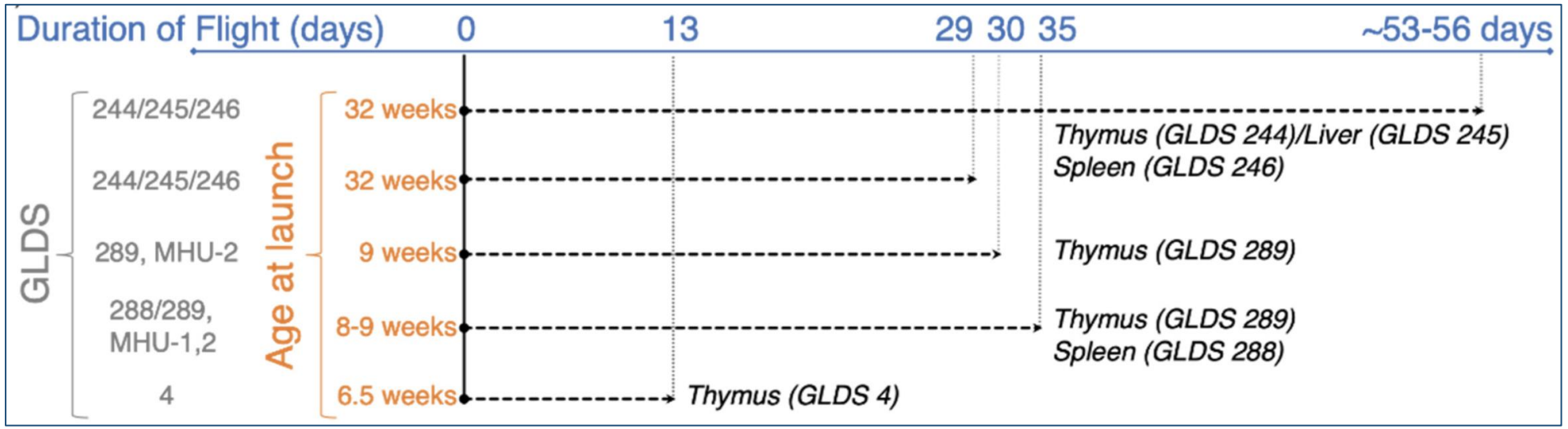
SPOKE

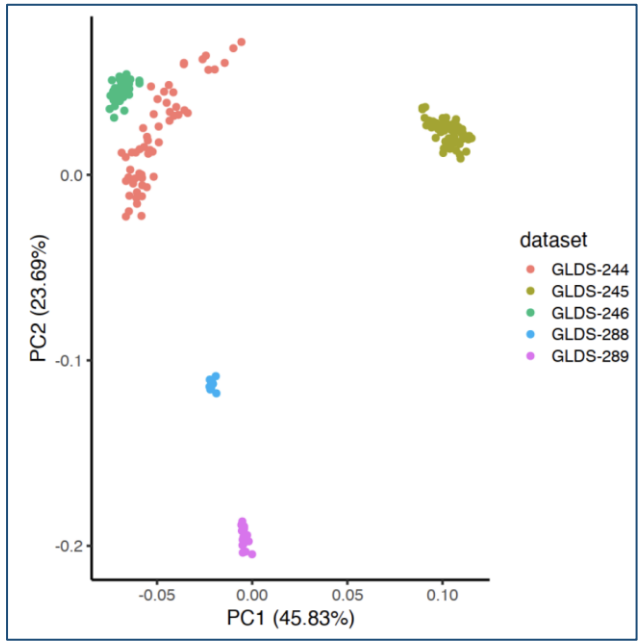
Scalable Precision Medicine Open Knowledge Engine

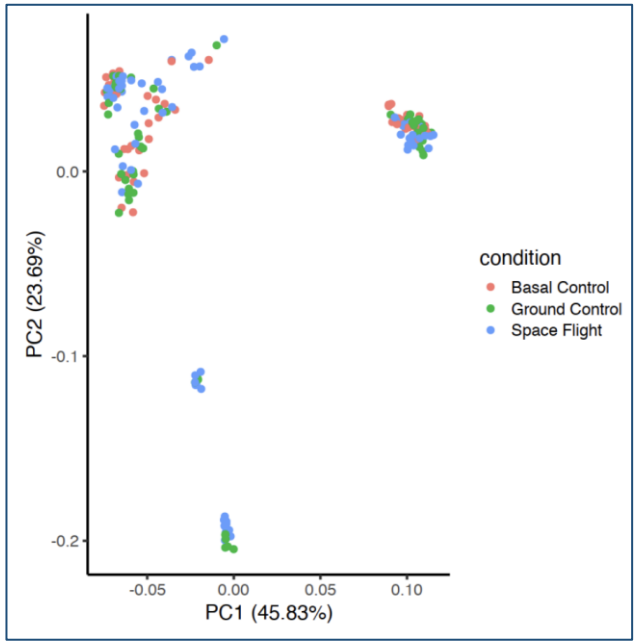


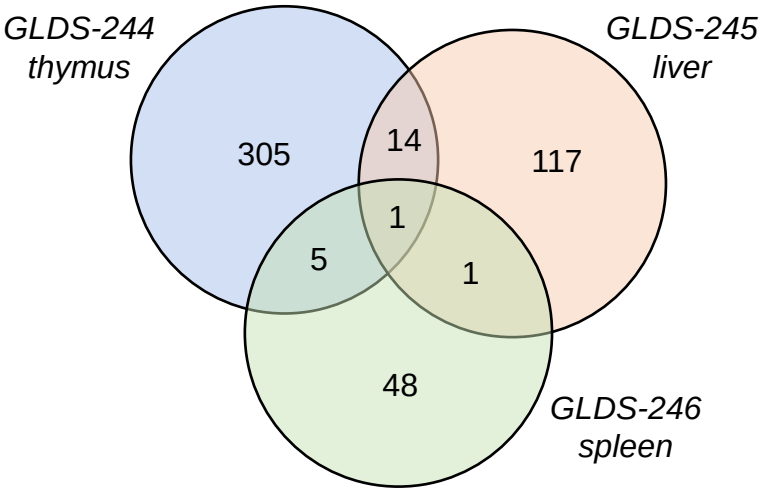
spoke.ucsf.edu/

Transcriptomic analysis of GeneLab spaceflown mouse thymus, spleen, liver

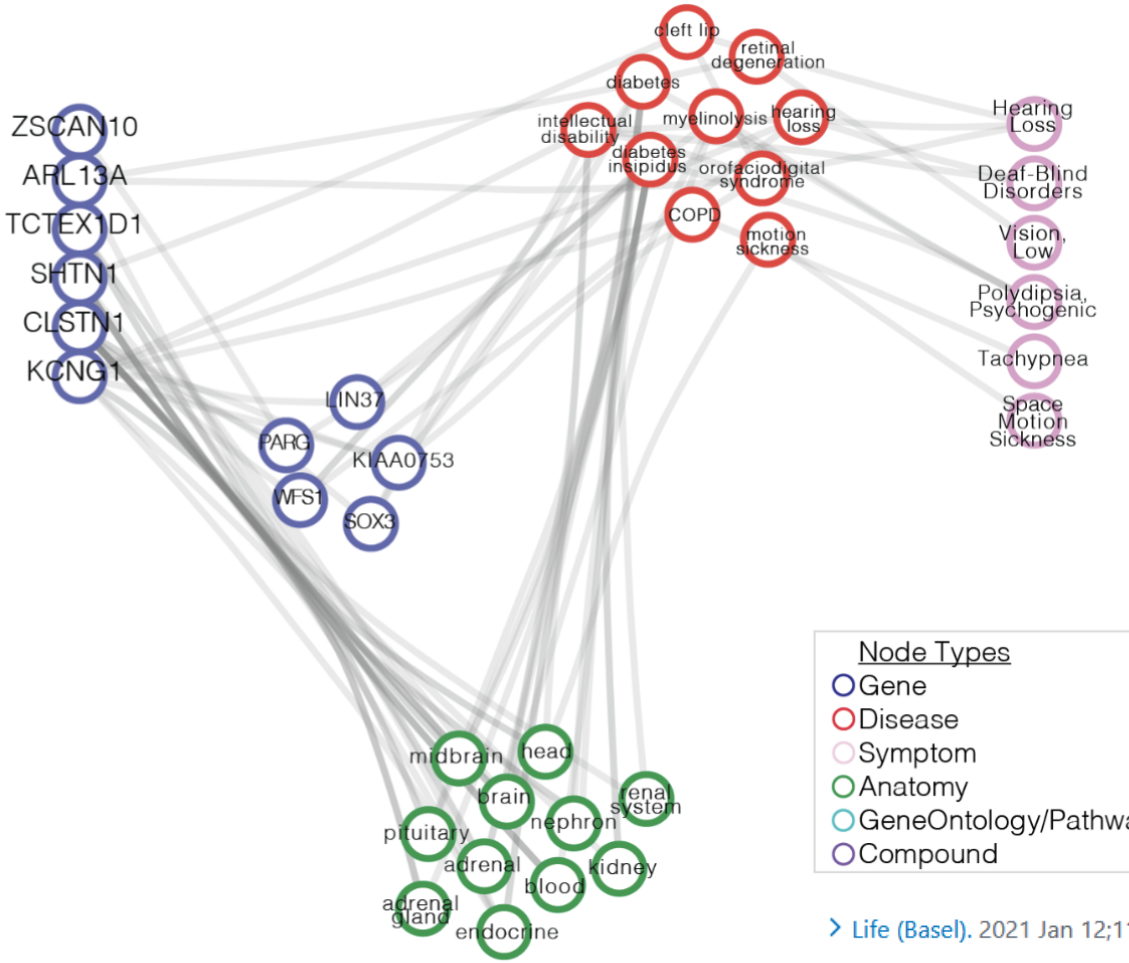


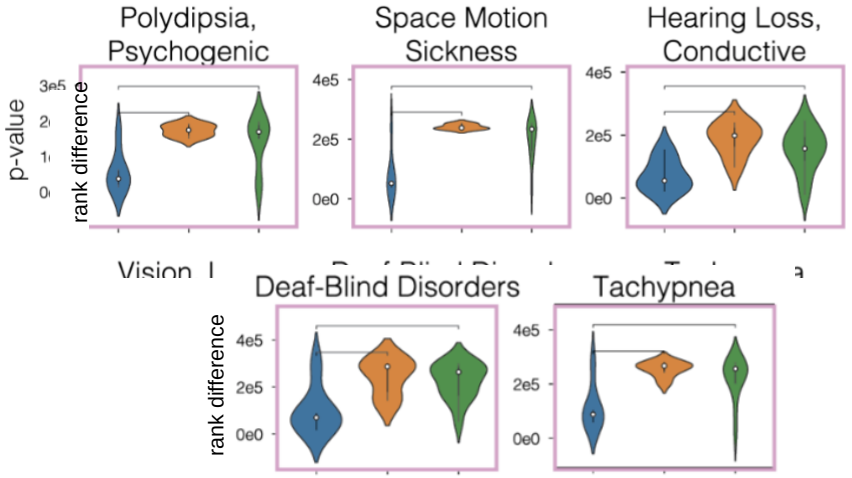






Significant *Symptom* nodes





Node Types
- Gene
- Disease
- Symptom
- Anatomy
- GeneOntology/Pathway
- Compound

Violin Plots
- Ground v Baseline
- Space v Baseline
- Space v Ground

Life (Basel). 2021 Jan 12;11(1):42. doi: 10.3390/life11010042.

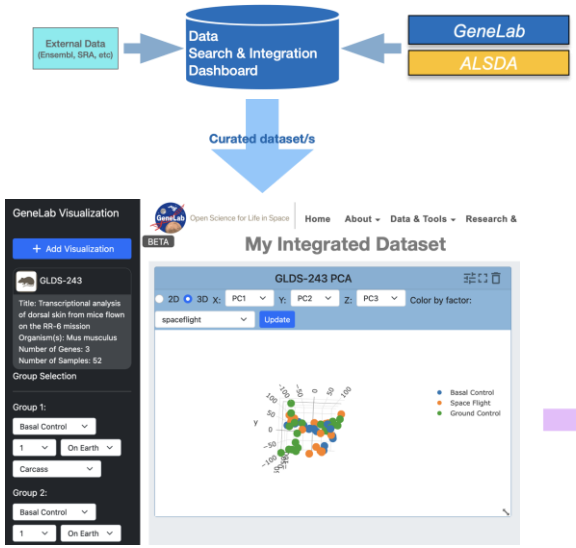
Knowledge Network Embedding of Transcriptomic Data from Spaceflown Mice Uncovers Signs and Symptoms Associated with Terrestrial Diseases

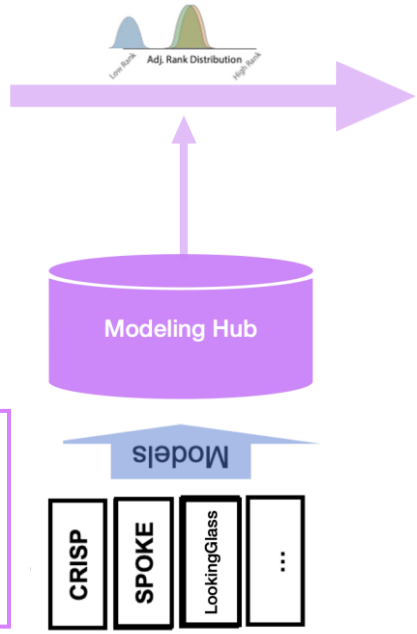
Charlotte A Nelson [1], Ana Uriarte Acuna [2 3], Amber M Paul [2 4], Ryan T Scott [2 3], Atul J Butte [5 6], Egle Cekanaviciute [2], Sergio E Baranzini [1 5 7], Sylvain V Costes [2]

spoke.ucsf.edu/

Coming soon! *GeneLab Model Hub and Analysis Portal*

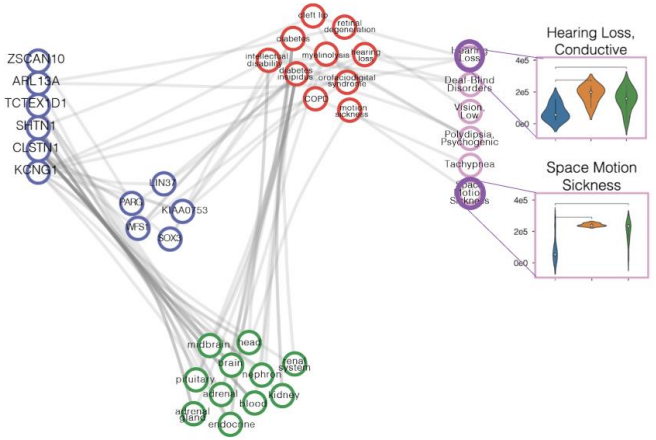
Streamlined Data Integration
of multiple datasets

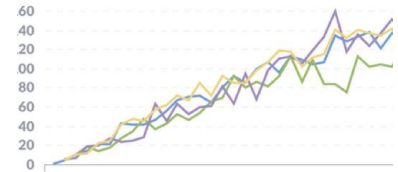


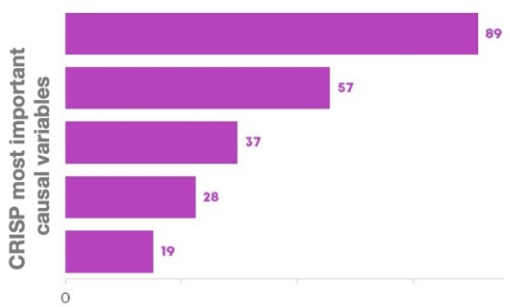


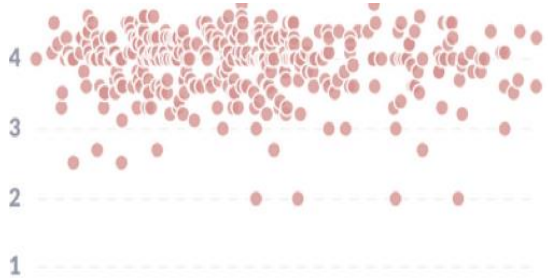
Modeling Hub
Automated pipelines for a suite of machine learning and modeling tools

Analysis Portal
Interactive analysis dashboard of modeling and analysis results









Thank you!

Twitter: @NASAGeneLab
https://genelab.nasa.gov

2017-2018

2018-2019

GeneLab is funded by the NASA Space Biology program within the NASA Science Mission Directorate's (SMD) Biological and Physical Sciences (BPS) Division