

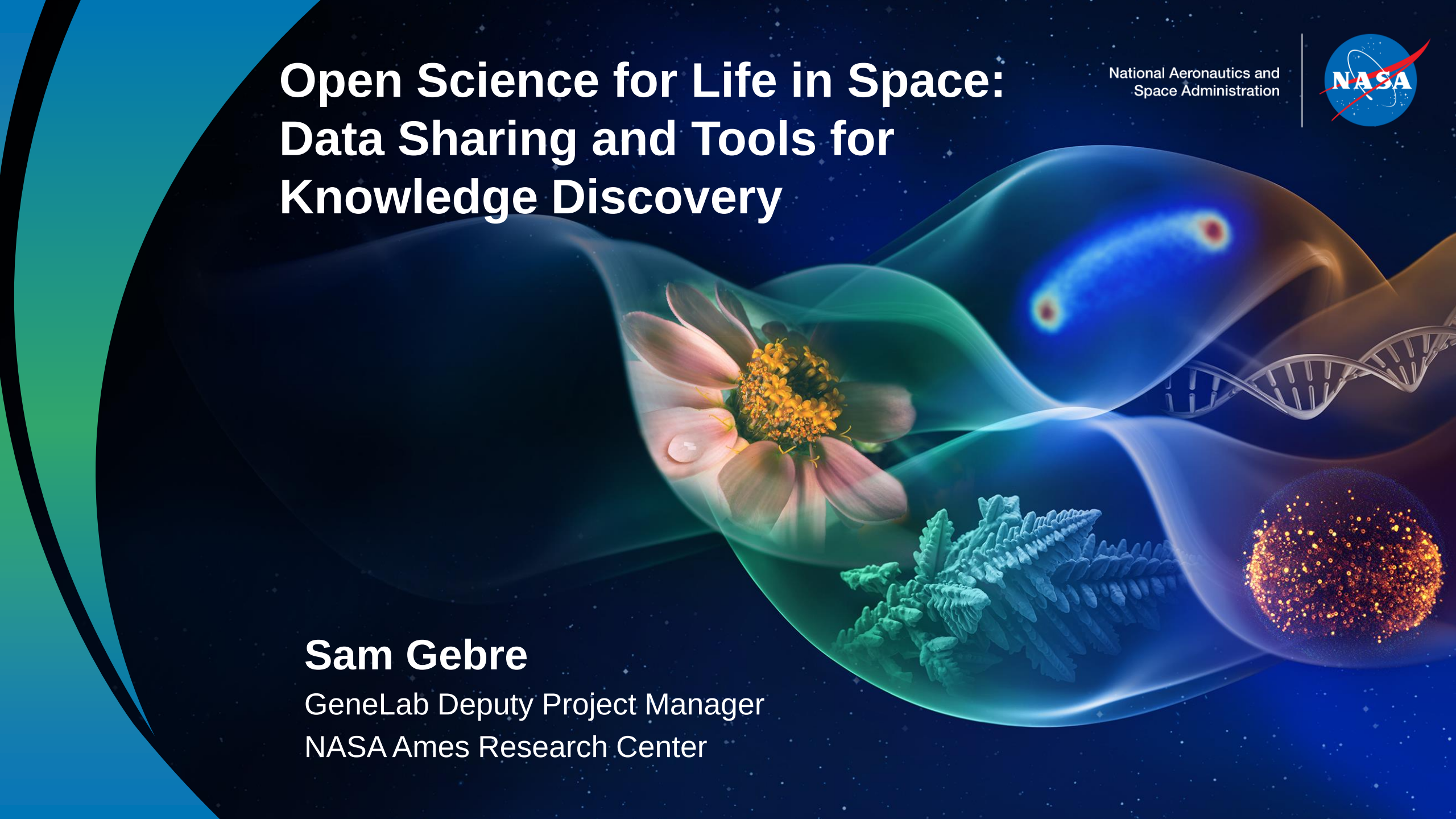
Open Science for Life in Space: Data Sharing and Tools for Knowledge Discovery

National Aeronautics and
Space Administration



Sam Gebre

GeneLab Deputy Project Manager
NASA Ames Research Center

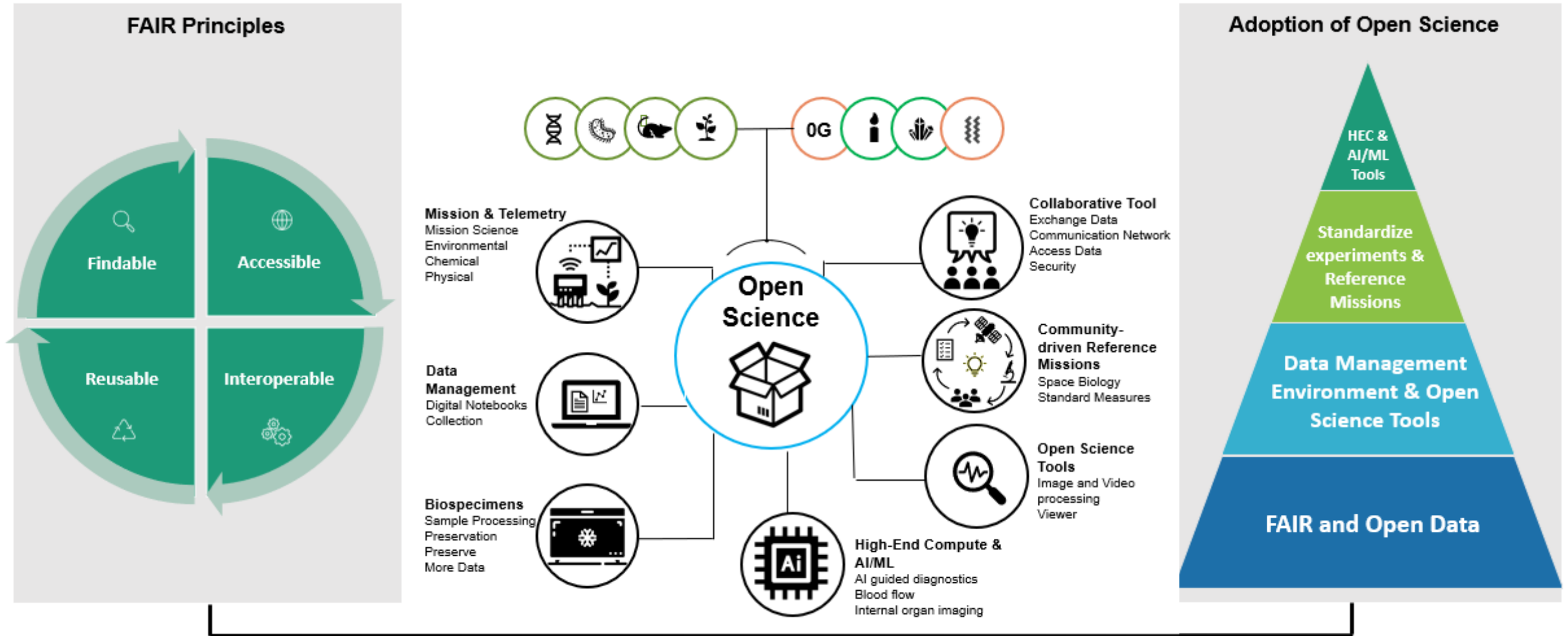


Open Science

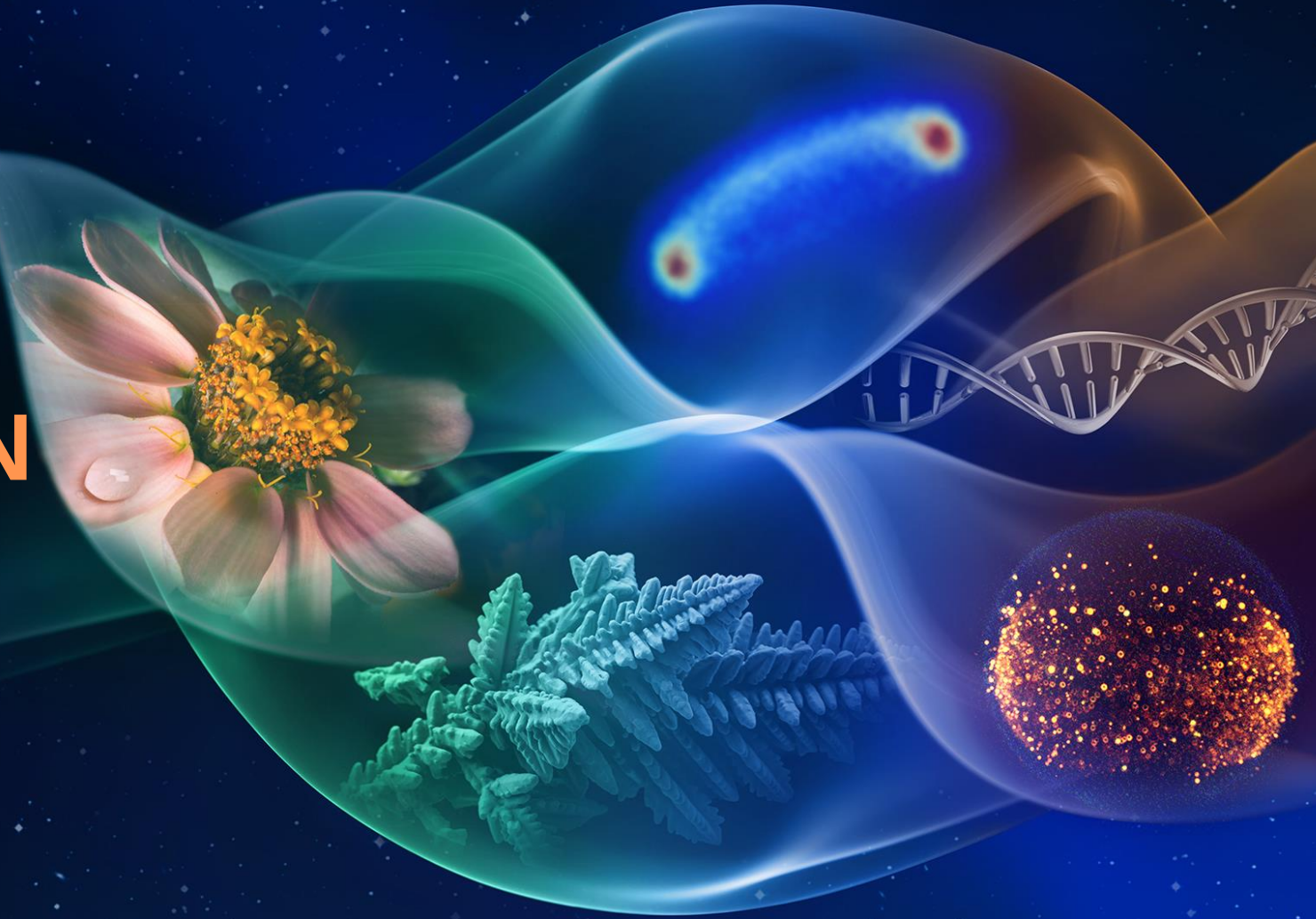
“We define open science as a collaborative culture enabled by technology that empowers the **open sharing of data, information, and knowledge** within the scientific community and the wider public to accelerate scientific research and understanding.”

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>

How will we get there?

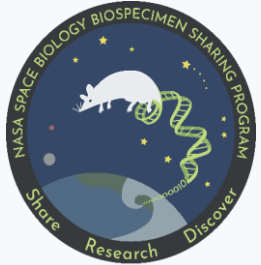


MAXIMIZING THE SCIENTIFIC RETURN OF SPACEFLIGHT EXPERIMENTS



NASA Biological Open Science Resources

Biospecimen Sharing Program (BSP)



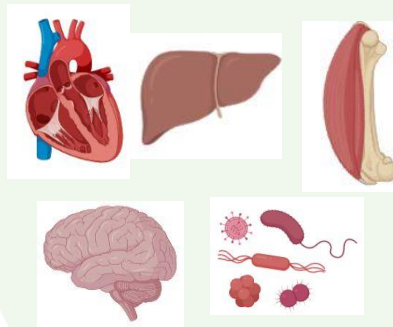
- *Dissects and preserves* rodent tissues from Flight and Ground investigations
- *Coordinates* internal tissue sharing



NASA Biological Institutional Scientific Collection (NBISC)



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



Ames Life Sciences Data Archive (ALSDA)



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



GeneLab (GL)



- *Collects and curates* omics data





NASA Biological Institutional Scientific Collection

Spaceflight

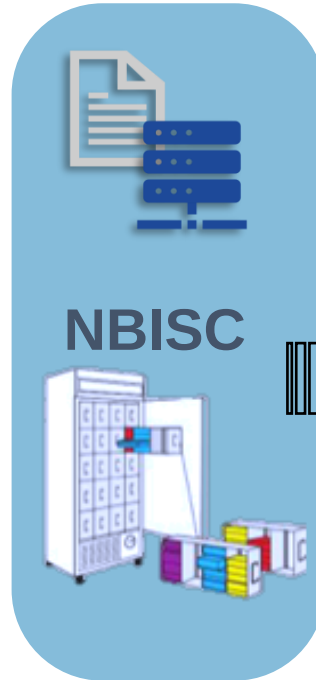


Ground Analogs



Metadata

Specimens



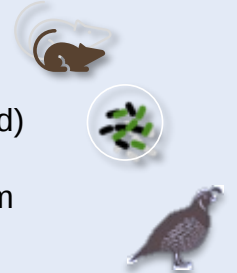
Sharing

~35,000 Non-Human Biospecimens*

- Spaceflight and Ground Analog
- Mouse, Rat, Quail, Microbial isolates (expected)
- Searchable catalog (LSDA Website)
- Requests via LSDA Biospecimen Request form

~50,000 Space Radiation Mouse Biospecimens

- Space Radiation Ground Analog
- Newly submitted, will be available soon

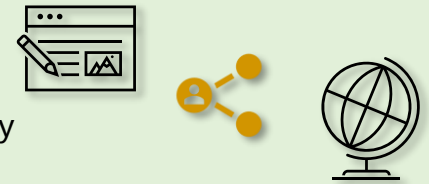


Resulted in:

- 33 publications since 2011
- 53 tissue requests since 2016
- 41 GeneLab data sets



- Fills spaceflight knowledge gaps
- Increases scientific return
- Publications for early career PIs
- Broadens our scientific community
- International collaboration



*Specimens mostly from Shuttle and ISS missions. Also stored are specimens from ground analog studies including centrifuge, hind-limb unloaded and partial weight bearing.

Ames Life Science Data Archive



• Past, Present, Future

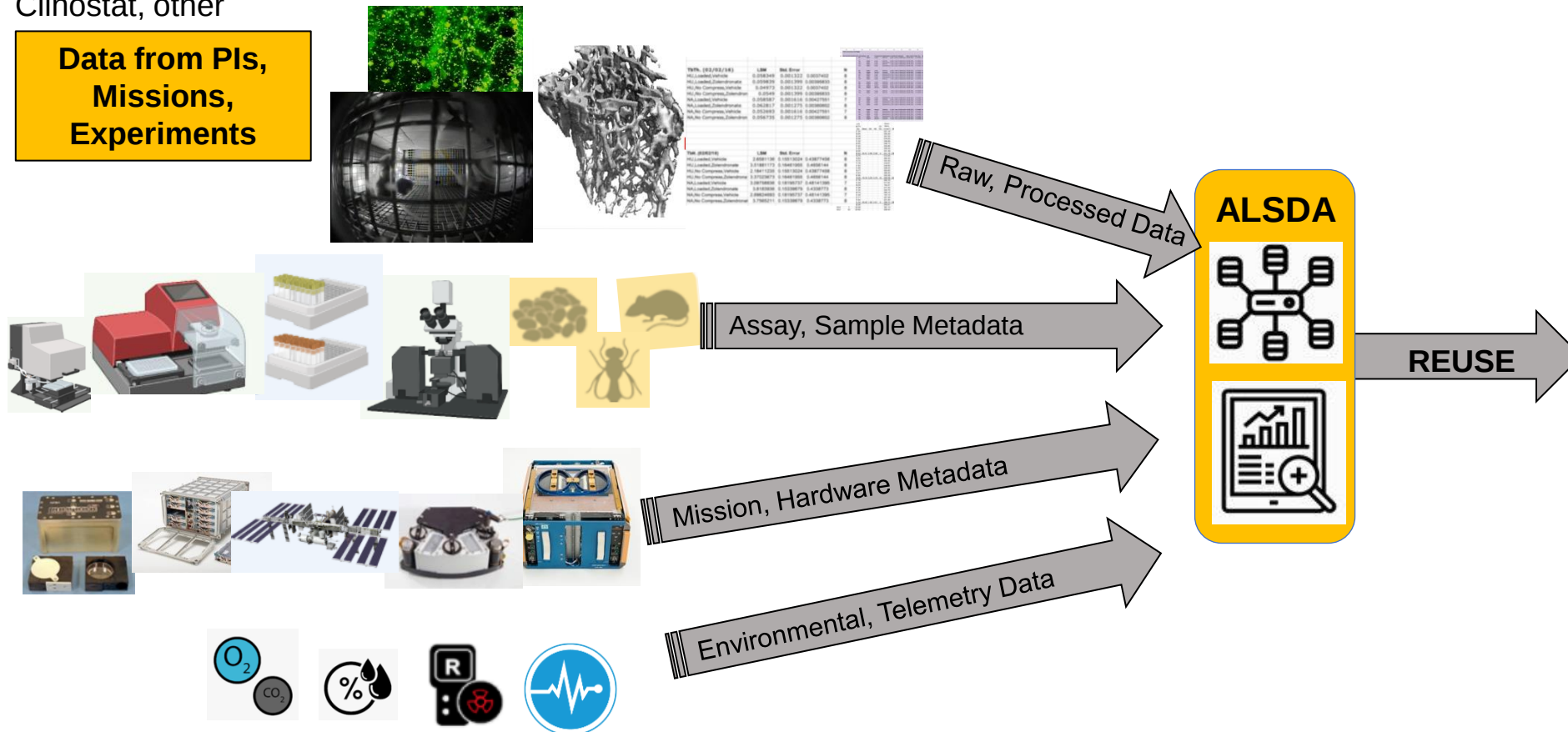
- International Space Station, Space Shuttle, Bion, Hindlimb Unloading, Radiation-Exposed, Partial Weight Bearing, Rotating Wall Vessel, Clinostat, other

Data from Pls, Missions, Experiments

• Basic & Applied

- μ Gravity, Radiation
- Behavioral
- Countermeasures

- **Mouse, Rat, Drosophila, Yeast, Bacteria, more soon!**



Open Science Data Repositories

- Phenotypic-Physiological-Behavioral-Imaging-Level Data
- **Single Submission Portal**
- **Data federated with GeneLab** (Molecular Omics-Level Data)
- Often Same Subjects

GeneLab

Open Science for Life in Space

Home About Data & Tools Research & Resources Working Groups Help

REGISTER NOW for the GeneLab AWG Symposium on June 30th from 10 AM to 2:30 PM PT.

Welcome to NASA GeneLab - the first comprehensive space-related omics database; users can upload, download, share, store, and analyze spaceflight and spaceflight-relevant data from experiments using model organisms.

Data Repository
Search and upload spaceflight datasets

Analyze Data
Perform large-scale analysis of biological omics data

Environmental Data
Radiation data collected during experiments conducted in space

Collaborative Workspace
Share, organize and store files

Submit Data
Have space-relevant data to submit to GeneLab?

Visualize Data
Interact with GeneLab processed data

LATEST NEWS

AWG Symposium 2022

Leveraging Advances in Modern Science to Revitalize Low-Dose Radiation

AWG member spotlight—Congratulations Dr Parul Gupta!

AWG charter gets a facelift!

TRANSCRIPTOMICS

GLDS-402: SOG1 activator and MYB3R repressors regulate a complex DNA damage network in Arabidopsis.

GLDS-467: Expression profiles in multiple tissues of VP and TOR1-knockout mouse under space flight

GLDS-432: Expression profiling reveals transcriptional regulation by Flow7/mTOR pathway in radiation-

GENOMICS (including metagenomics, epigenetics)

GLDS-498: The SOG1 transcriptional activator and the MYB3R family of repressors control a complex gene

GLDS-497: Fungal Diversity associated with space-craft assembly facility (SAF)

GLDS-486: Microbiome profiling of feces from mice flown on the RR-10 mission

PROTEOMICS / METABOLOMICS

GLDS-390: Effects of low dose space radiation exposures on the splenic metabolome

GLDS-462: Transcriptional profiling, protein expression profiling, and phosphoprotein profiling of kidneys from

GLDS-359: Proteome profiling of the exhaled breath condensate after long-term spaceflights

PROCESSED

GLDS-401: Alternative splicing regulates the physiological adaptation of the mouse hind limb postural and phasic

GLDS-370: Transcriptional profiling of livers from mice flown on Rodent Research Reference Mission-1 (RRRM-1)

GLDS-314: Adaptive response of Arabidopsis seedlings in microgravity and Mars reduced gravity environment is

www.genelab.nasa.gov

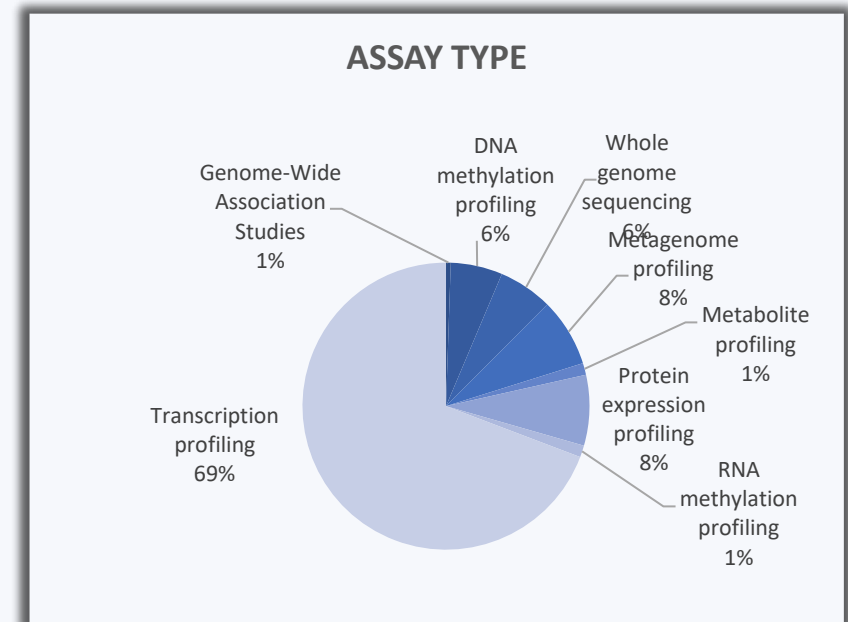
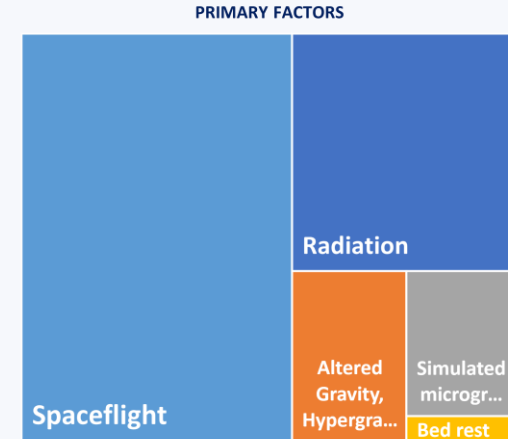
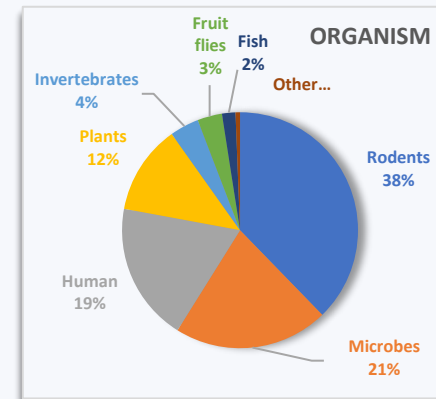
378
Studies

429
Datasets

45
Species

>10
Assays

>135 TB
Data





5,847

Biospecimens
available

159

Experimental
Payloads

1,260

Research
Records

<20

Species

164

Omics Studies

Re-use of Data and Enabling New Discoveries

40 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 Cahill¹, Henry Cope², Joseph J. Bass³, Eliah G. Overbey⁴, Rachel Gilbert^{5,6}, William Abraham da Silveira^{1,7}, Amber M. Paul^{5,8,9}, Tejaswini Mishra¹⁰, Raúl Herranz¹¹, Sigrid S. Reinsch⁵, Sylvain V. Costes⁵, Gary Hardiman^{1,12}, and Nathaniel J. Szewczyk

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

Cell Supports open access

ARTICLE | VOLUME 183, ISSUE 5, P1185-1201.E20, NOVEMBER 25, 2020

PDF [10 MB]

Comprehensive Multi-omics Analysis Reveals Mitochondrial Stress as a Central Biological Hub for Spaceflight Impact

Environmental and Molecular Mutagenesis

Environmental Mutagenesis and Genomics Society

BRIEF COMMUNICATION | Open Access

In vitro relationships of galactic cosmic radiation and epigenetic clocks in human bronchial epithelial cells

Jamaji C. Nwanaji-Enwerem, Philippe Boileau, Jonathan M. Galazka, Andres Cardenas

ELSEVIER

Computational and Structural Biotechnology Journal

Volume 19, 2021, Pages 2223-2235

Reanalysis of the Ma...
common gut microb...
astronauts induced l...

Cell Reports

Open access

REPORT | VOLUME 33, ISSUE 10, 108406, DECEMBER 08, 2020

PDF [1 MB] Figures

ic Aging and Leukocyte Analysis of Simulated
-500 Mission

Izoi Nwanaji-Enwerem, Lars Van Der Laan, Jonathan M. Galazka

iScience

Open access

ARTICLE | VOLUME 23, ISSUE 12, 101734, DECEMBER 18, 2020

PDF [2 MB]

Comparative Transcriptomics Identifies Neuronal and Metabolic Adaptations to Hypergravity and Microgravity in *Caenorhabditis elegans*

Craig R.G. Willis, Nathaniel J. Szewczyk, Sylvain V. Costes, ... Sigrid S. Reinsch, Timothy Etheridge, Catharine A. Conley, Show all authors, Show footnotes

Open Access Published: November 25, 2020 DOI: <https://doi.org/10.1016/j.isci.2020.101734>

Open Access

Article

GeneLab Database Analyses Suggest Long-Term Impact of Space Radiation on the Cardiovascular System by the Activation of *FYN* Through Reactive Oxygen Species

by Afshin Beheshti^{1,*}, J. Tyson McDonald², Jack Miller³, Peter Grabham⁴ and Sylvain V. Costes^{5,*}

CONTRIBUTE

StatusIn CurationRelease Date24-Nov-2021ValidationFailed

0 Reset Study

NASA GeneLab: interfaces for the exploration of space omics data.

Samrawit Gebre, Daniel C Berrios, Jonathan Galazka, Kirill Grigorev, Sylvain V Costes

+ Add person

The mission of NASA's GeneLab database (<https://genelab.nasa.gov/>) is to collect, curate, and provide access to the genomic, transcriptomic, proteomic and metabolomic (so-called 'omics') data from biospecimens flown in space or exposed to simulated space stressors, maximizing their utilization. This large collection of data enables the exploration of molecular network responses to space environments using a systems biology approach. We review here the various components of the GeneLab platform, including the new data repository web interface, and the GeneLab Online Data Entry (GEODE) web portal, which will support the expansion of the database in the future to include companion non-omics assay data. We discuss our design for GEODE, particularly how it promotes investigators providing more accurate metadata, reducing the curation effort required of GeneLab staff. We also introduce here a new GeneLab Application Programming Interface (API) specifically designed to support tools for the visualization of processed omics data. We review the outreach efforts by GeneLab to utilize the spaceflight data in the repository to generate novel discoveries and develop new hypotheses, including spearheading data analysis working groups, and a high school student training program. All these efforts are aimed ultimately at supporting precision risk management for human space exploration.

PUBLICATIONS

+ Add Publication

NASA GeneLab: interfaces for the exploration of space omics data.

▲ Berrios DC, Galazka J, Grigorev K, Gebre S, Costes SV.

+ Edit

Description

Samples

Assays

Protocols

Files

Related Studies

Study Validations

Study Design Descriptor

+ Add

✓ Growth Condition Design

DATA SUBMISSION:
ADD, EDIT, SAVE,
SUBMIT

StatusIn CurationRelease Date24-Nov-2021ValidationFailed

Reset Study

NASA GeneLab: interfaces for the exploration of space omics data.

Samrawit Gebre, Daniel C Berrios, Jonathan Galazka, Kirill Grigorev, Sylvain V Costes

+ Add person

The mission of NASA's GeneLab database (<https://genelab.nasa.gov/>) is to collect, curate, and provide access to the genomic, transcriptomic, proteomic and metabolomic (so-called 'omics') data from biospecimens flown in space or exposed to simulated space stressors, maximizing their utilization. This large collection of data enables the exploration of molecular network responses to space environments using a systems biology approach. We review here the various components of the GeneLab platform, including the new data repository web interface, and the GeneLab Online Data Entry (GEODE) web portal, which will support the expansion of the database in the future to include companion non-omics assay data. We discuss our design for GEODE, particularly how it promotes investigators providing more accurate metadata, reducing the curation effort required of GeneLab staff. We also introduce here a new GeneLab Application Programming Interface (API) specifically designed to support tools for the visualization of processed omics data. We review the outreach efforts by GeneLab to utilize the spaceflight data in the repository to generate novel discoveries and develop new hypotheses, including spearheading data analysis working groups, and a high school student training program. All these efforts are aimed ultimately at supporting precision risk management for human space exploration.

PUBLICATIONS

+ Add Publication

NASA GeneLab: interfaces for the exploration of space omics data.

👤 Berrios DC, Galazka J, Grigorev K, Gebre S, Costes SV.

Edit

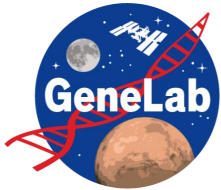
DescriptionSamplesAssaysProtocolsFilesRelated StudiesStudy Validations

Study Design Descriptor

+ Add

Growth Condition Design

12



Enhancing Data Submission Portal

Enhancing submission automation and “self-service”

Welcome to NASA GeneLab - the first comprehensive space-related omics database; users can upload, download, share, store, and analyze spaceflight and spaceflight-relevant data from experiments using model organisms.



Data Repository

Search and upload spaceflight datasets



Analyze Data

Perform large-scale analysis of biological omics data



Environmental Data

Radiation data collected during experiments conducted in space



Collaborative Workspace

Share, organize and store files



Submit Data

Have space-relevant data to submit to GeneLab?



Visualize Data

Interact with GeneLab processed data

Search Data

☐ All ☒ GeneLab ☐ NIH GEO ☐ EBI PRIDE ☐ ANL MG-FAST ☒ ALSDA

Search Filters (GeneLab Only)

Project Type Factors Organisms Assay Type

Show Only:

☐ Studies With Visualizations

Page 1 of 13 (Total Studies: 302) [Next >](#)

Studies Per Page:



GLDS-426

Effects of Spaceflight on Bone Microarchitecture in the Axial and Appendicular Skeleton in Growing Ovariectomized Rats

Organisms	Factors	Assay Types	Release Date	Description
Rattus norvegicus	Spaceflight Ovariectomized Age	Radiography	25-Feb-2020	This study investigated the effects of a 14-day spaceflight on bone mass, density and microarchitecture in weight bearing (femur and humerus) and non-weight bearing (2nd lumbar vertebra and calvarium)...



GLDS-401

Alternative splicing regulates the physiological adaptation of the mouse hind limb postural and phasic muscles to microgravity

Organisms	Factors	Assay Types	Release Date	Description
Mus musculus	Spaceflight Tissue	transcription profiling	25-Jun-2021	We sought to comprehensively elucidate the transcriptomic underpinnings of microgravity-induced muscle phenotypes in mice by evaluating both differential gene expression (DGE) and changes in alternati...

Description Samples **Assays** Protocols Files Study Validations

+ Add Assay

Micro-computed tomography

Novel Object Recognition

Environmental Gene Survey

Transcription Factor Binding Site Identification

Western Blot

Histomorphometry

Protein Identification

Protein-Protein Interaction Detection

Radiation Measurement

Ultrasonography

Radiography

Adding
Phenotypic
ALSDA
Assays to
Portal

Cancel

GEODE

NEWSLETTER SIGNUP

Email Address *

Questions and Feedback

SITE INFORMATION

NASA GeneLab Overview

Terms and Conditions

NASA Ames Space Biosciences Division

NASA Life Sciences Data Archive

NASA Physical Science Informatics

Center for the Advancement of Science

CONTRIBUTE

StatusIn CurationRelease Date24-Nov-2021ValidationFailed

0Reset Study

NASA GeneLab: interfaces for the exploration of space omics data.

Samrawit Gebre, Daniel C Berrios, Jonathan Galazka, Kirill Grigorev, Sylvain V Costes

+ Add person

The mission of NASA's GeneLab database (https://genelab.nasa.gov/) is to collect, curate, and provide access to the genomic, transcriptomic, proteomic and metabolomic (so-called 'omics') data from biospecimens flown in space or exposed to simulated space stressors, maximizing their utilization. This large collection of data enables the exploration of molecular network responses to space environments using a systems biology approach. We review here the various components of the GeneLab platform, including the new data repository web interface, and the GeneLab Online Data Entry (GEODE) web portal, which will support the expansion of the database in the future to include companion non-omics assay data. We discuss our design for GEODE, particularly how it promotes investigators providing more accurate metadata, reducing the curation effort required of GeneLab staff. We also introduce here a new GeneLab Application Programming Interface (API) specifically designed to support tools for the visualization of processed omics data. We review the outreach efforts by GeneLab to utilize the spaceflight data in the repository to generate novel discoveries and develop new hypotheses, including spearheading data analysis working groups, and a high school student training program. All these efforts are aimed ultimately at supporting precision risk management for human space exploration.

PUBLICATIONS

+ Add Publication

NASA GeneLab: interfaces for the exploration of space omics data.

▲ Berrios DC, Galazka J, Grigorev K, Gebre S, Costes SV.

✎ Edit

DescriptionSamplesAssaysProtocolsFilesRelated StudiesStudy Validations

Study Design Descriptor

+ Add

✓ Growth Condition Design

DISCOVER

Search Data

x

Q

☐ All☒ GeneLab☐ NIH GEO☐ EBI PRIDE☐ ANL MG-RAST

Search Filters (GeneLab Only)

Project TypeFactorsOrganismsAssay TypeClear

☐ Ground☐ Spaceflight☐ Parabolic Flight Study

bions

☐ Acinetobacter pittii☐ Acinetobacter pittii☐ Acinetobacter radioresistens☐ Arabidopsis thaliana☐ Arabidopsis thaliana☐ Arancicola proteolyticus☐ Aspergillus fumigatus☐ Aspergillus nidulans☐ Aspergillus niger☐ Aspergillus niger☐ Aspergillus terreus

Arabisidopsis th

Organisms

Arabisidopsis thaliand

GLDS-436

GLDS-434

Page 1 of 14 (Total Studies: 339) Next >

Studies Per Page: 25

n C-9 parabolic flight, 2015

Assay Types	Release Date	Description
transcription profiling	01-Apr-2021	The increasing availability of flights on suborbital rockets creates new avenues for the study of spaceflight effects on biological systems, in particular the transitions between hypergravity and micr...

4-3-3k:GFP overexpression line on the F-104 Starfighter, first flight

Organisms	Factors	Assay Types	Release Date	Description
Arabidopsis thaliana	Altered Gravity Time	transcription profiling	01-Apr-2021	The increasing availability of flights on suborbital rockets creates new avenues for the study of spaceflight effects on biological systems, in particular the transitions between hypergravity and micr...

DATA REPOSITORY

Search Data x 

☐ All ☒ GeneLab ☐ NIH GEO ☐ EBI PRIDE ☐ ANL MG-RAST

Search Filters (GeneLab Only)

Project Type  Factors  Organisms  Assay Type  Clear

☐ Ground
☐ Spaceflight
☐ Parabolic Flight Study

☐ Acinetobacter pittii
☐ Acinetobacter pittii
☐ Acinetobacter radioresistens
☐ Arabidopsis thaliana
☐ Arabidopsis thaliana
☐ Aranicola proteolyticus
☐ Aspergillus fumigatus
☐ Aspergillus nidulans
☐ Aspergillus niger
☐ Aspergillus niger
☐ Aspergillus terreus

Page 1 of 14 (Total Studies: 339) [Next >](#)

Studies Per Page: 25 



GLDS-436

Arabidopsis th

Organisms

Arabidopsis thaliana

on C-9 parabolic flight, 2015

Arabidopsis th

4-3-3k:GFP overexpression line on the F-104 Starfighter, first flight



GLDS-434

Organisms	Factors	Assay Types	Release Date	Description
Arabidopsis thaliana	Altered Gravity Time	transcription profiling	01-Apr-2021	The increasing availability of flights on suborbital rockets creates new avenues for the study of spaceflight effects on biological systems, in particular the transitions between hypergravity and micr...

CONTRIBUTE

StatusIn CurationRelease Date24-Nov-2021ValidationFailed

0

Reset Study

NASA GeneLab: interfaces for the exploration of space omics data.

Samrawit Gebre, Daniel C Berrios, Jonathan Galazka, Kirill Grigorev, Sylvain V Costes

+ Add person

The mission of NASA's GeneLab database (<https://genelab.nasa.gov/>) is to collect, curate, and provide access to the genomic, transcriptomic, proteomic and metabolomic (so-called "omics") data from biospecimens flown in space or exposed to simulated space stressors, maximizing their utilization. This large collection of data enables the exploration of molecular network responses to space environments using a systems biology approach. We review here the various components of the GeneLab platform, including the new data repository web interface, and the GeneLab Online Data Entry (GEODE) web portal, which will support the expansion of the database in the future to include companion non-omics assay data. We discuss our design for GEODE, particularly how it promotes investigators providing more accurate metadata, reducing the curation effort required of GeneLab staff. We also introduce here a new GeneLab Application Programming Interface (API) specifically designed to support tools for the visualization of processed omics data. We review the outreach efforts by GeneLab to utilize the spaceflight data in the repository to generate novel discoveries and develop new hypotheses, including spearheading data analysis working groups, and a high school student training program. All these efforts are aimed ultimately at supporting precision risk management for human space exploration.

PUBLICATIONS

+ Add Publication

NASA GeneLab: interfaces for the exploration of space omics data.

Berrios DC, Galazka J, Grigorev K, Gebre S, Costes SV.

Edit

Description

SamplesAssaysProtocolsFilesRelated StudiesStudy Validations

Study Design Descriptor

+ Add

Growth Condition Design

ANALYZE

GeneLab

Open Science for Life in Space

Analyze DataWorkflowVisualizeShared DataAdminHelpUser

Using 29.1 GB

Tools

search tools

from tastq tiles

Volcano Plot create a volcano plot

Annotate DESeq2/DEXSeq output tables Append annotation from GTF to differential expression tool outputs

edgeR Perform differential expression of count data

annotateMyIDs annotate a generic set of identifiers

htseq-count - Count aligned reads in a BAM file that overlap features in a GFF file

featureCounts Measure gene expression in RNA-Seq experiments from SAM or BAM files.

htseq-count - Count aligned reads in a BAM file that overlap features in a GFF file (Galaxy Version 0.9.1+galaxy1)

FavoriteOptions

Aligned SAM/BAM File

No sam or bam dataset available.

GFF File

No gff dataset available.

Mode

Union

Mode to handle reads overlapping more than one feature. (--mode)

Stranded

Yes

Specify whether the data is from a strand-specific assay. **Be sure to choose the correct value** (see help for more information). (--stranded)

History

search datasets

New

7 shown, 2 deleted

780.32 MB

9: Mus_musculus.GRCm38.dna.primary_assembly.fa.gz

6: fgsea on data 2 and data 1: Plots

5: fgsea on data 2 and data 1: Ranked category list

4: fgsea on data 2 and data 1: Plots

DISCOVER

Search Data

GeneLab

AllGeneLabNIH GEONEBI PRIDEANL MG-RAST

Search Filters (GeneLab Only)

Project TypeFactorsOrganismsAssay TypeClear

GroundSpaceflightParabolic Flight Study

Acinetobacter pittiiAcinetobacter pittiiAcinetobacter radioresistensArabidopsis thalianaArabidopsis thalianaAranicola proteolyticusAspergillus fumigatusAspergillus nidulansAspergillus nigerAspergillus nigerAspergillus terreus

Page 1 of 14 (Total Studies: 339) Next >

Studies Per Page: 25

GLDS-436

GLDS-434

Arabis C-9 parabolic flight, 2015

Assay Types	Release Date	Description
transcription profiling	01-Apr-2021	The increasing availability of flights on suborbital rockets creates new avenues for the study of spaceflight effects on biological systems, in particular the transitions between hypergravity and microgravity.

Arabis 4-3-3k:GFP overexpression line on the F-104 Starfighter, first flight

Organisms	Factors	Assay Types	Release Date	Description
Arabidopsis thaliana	Altered Gravity Time	transcription profiling	01-Apr-2021	The increasing availability of flights on suborbital rockets creates new avenues for the study of spaceflight effects on biological systems, in particular the transitions between hypergravity and microgravity.

16

ANALYSIS PLATFORM



Open Science for Life in Space

Analyze Data

Workflow

Visualize ▾

Shared Data ▾

Admin

Help ▾

User ▾



Using 29.1 GB

Tools



search tools



from fastq files

Volcano Plot create a volcano plot

Annotate DESeq2/DEXSeq output tables Append annotation from GTF to differential expression tool outputs

edgeR Perform differential expression of count data

annotateMyIDs annotate a generic set of identifiers

htseq-count - Count aligned reads in a BAM file that overlap features in a GFF file

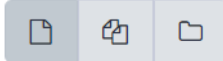
featureCounts Measure gene expression in RNA-Seq experiments from SAM or BAM files.

htseq-count - Count aligned reads in a BAM file that overlap features in a GFF file (Galaxy Version 0.9.1+galaxy1)

☆ Favorite

▾ Options

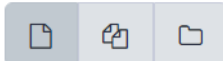
Aligned SAM/BAM File



No sam or bam dataset available.



GFF File



No gff dataset available.



Mode

Union



Mode to handle reads overlapping more than one feature. (--mode)

Stranded

Yes



Specify whether the data is from a strand-specific assay. **Be sure to choose the correct value** (see help for more information). (--stranded)

History



search datasets



New

7 shown, 2 deleted

780.32 MB



9: Mus_musculus.GRCm38.dna.primary_assembly.fa.gz



6: fgsea on data 2 and data 1: Plots



5: fgsea on data 2 and data 1: Ranked category list



4: fgsea on data 2 and data 1: Plots



CONTRIBUTE

Status

In Curation

Released Date

24-Nov-2021

Validation

Failed

Reset Study

NASA GeneLab: interfaces for the exploration of space omics data.

Samravitt Gebre, Daniel C Berrios, Jonathan Galazka, Kiril Grigorev, Sylvain V Costes

+ Add person

The mission of NASA's GeneLab database (<https://genelab.nasa.gov/>) is to collect, curate, and provide access to the genomic, transcriptomic, proteomic and metabolomic (so-called "omics") data from bioprecursors flown in space or exposed to simulated space stressors, maximizing their utilization. This website provides a collection of data enables the exploration of molecular network responses to space environments using a systems biology approach. We review here the various components of the GeneLab platform, including the new data repository web interface, and the GeneLab Online Data Entry (GOEDE) web portal, which will support the expansion of the database in the future to include companion non-omics assay data. We discuss our design for GOEDE, particularly how it promotes investigators providing more accurate metadata, reducing the curation effort required of GeneLab staff. We also introduce here a new GeneLab Application Programming Interface (API) specifically designed to support tools for the visualization of processed omics data. We review the outreach efforts by GeneLab to utilize the spacecraft data in the repository to generate novel discoveries and develop new hypotheses, including spearheading data analysis working groups, and a high school student training program. All these efforts are aimed ultimately at supporting precision risk management for human space exploration.

+ Add Publication

NASA GeneLab: interfaces for the exploration of space omics data.

✚

Berrios DC, Galazka J, Grigorev K, Gebre S, Costes SV.

✎

Edit

Description

Samples

Assays

Protocols

Files

Related Studies

Study Validations

Study Design Descriptor

+ Add

✓

Growth Condition Design

DISCOVER

Search Data

☐ All
 ☒ GeneLab
 ☐ NIH GEO
 ☐ EBI PRIDE
 ☐ ANL MG-RAST

Search Filters (GeneLab Only)

Project Type ▾

Factors ▾

Organisms ▾

Assay Type ▾

Clear

☐ Ground

☐ Spaceflight

☐ Parabolic Flight Study

bions

Arabidopsis thaliana
☐ Arabidopsis thaliana
☐ Aracnicola proteolyticus
☐ Aspergillus fumigatus
☐ Aspergillus nidulans
☐ Aspergillus niger
☐ Aspergillus niger
☐ Aspergillus terreus

Page 1 of 14 (Total Studies: 339) [Next >](#)
 Studies Per Page: 25 ▾

Study Types	Release Date	Description
transcription profiling	01-Apr-2021	The increasing availability of flights on suborbital rockets creates new avenues for the study of spaceflight effects on biological systems, in particular the transitions between hypergravity and microgravity.

Arabidopsis thaliana
☐ Arabidopsis thaliana
☐ Aracnicola proteolyticus
☐ Aspergillus fumigatus
☐ Aspergillus nidulans
☐ Aspergillus niger
☐ Aspergillus niger
☐ Aspergillus terreus

Arabidopsis thaliana C-9 parabolic flight, 2015

Study Types	Release Date	Description
transcription profiling	01-Apr-2021	The increasing availability of flights on suborbital rockets creates new avenues for the study of spaceflight effects on biological systems, in particular the transitions between hypergravity and microgravity.

Arabidopsis thaliana
☐ Arabidopsis thaliana
☐ Aracnicola proteolyticus
☐ Aspergillus fumigatus
☐ Aspergillus nidulans
☐ Aspergillus niger
☐ Aspergillus niger
☐ Aspergillus terreus

Arabidopsis thaliana C-4-3-3k:GFP overexpression line on the F-104 Starfighter, first flight

Organisms	Factors	Assay Types	Release Date	Description
Arabidopsis thaliana	Altered Gravity Time	transcription profiling	01-Apr-2021	The increasing availability of flights on suborbital rockets creates new avenues for the study of spaceflight effects on biological systems, in particular the transitions between hypergravity and microgravity.

ANALYZE


Open Science for Life in Space

Analyze Data
Workflow
Visualize
Shared Data
Admin
Help
User

Using 29.1 GB

Tools

search tools
from tastq tiles

Volcano Plot create a volcano plot
Annotate DESeq2/DEXSeq output tables Append annotation from GTF to differential expression tool outputs
edgeR Perform differential expression of count data
annotateMyIDs annotate a generic set of identifiers
htseq-count Count aligned reads in a BAM file that overlap features in a GFF file
featureCounts Measure gene expression in RNA-Seq experiments from SAM or BAM files.

htseq-count - Count aligned reads in a BAM file that overlap features in a GFF file (Galaxy Version 0.9.1+galaxy1)

Favorite
Options

Aligned SAM/BAM File

No sam or bam dataset available.

GFF File

No gff dataset available.

Mode

Union

Mode to handle reads overlapping more than one feature. (--mode)

Stranded

Yes

Specify whether the data is from a strand-specific assay. **Be sure to choose the correct value** (see help for more information). (--stranded)

History

search datasets

New

7 shown, 2 deleted
780.32 MB

9: Mus_musculus.GRCm38.dna.primary_assembly.fa.gz

6: fgsea on data 2 and data 1: Plots

5: fgsea on data 2 and data 1: Ranked category 1 list

4: fgsea on data 2 and data 1: Ranked category 1 list

EXPLORE

GeneLab Visualization

[+ Add Visualization](#)

GLDS-4

Title: Microarray Analysis of Space-Flown Murine Thymus Tissue

Organism(s): Mus musculus

Number of Genes: 20217

Number of Samples: 7

GLDS-4 PCA

GLDS-4 Volcano

GLDS-4 DGE

Group Selection

Group 1: Ground Control

Group 2: Space Flight

[Update](#)

Copy CSV Excel PDF Print

Maximum p-value:

Search:

Symbol	LOG2FC	PVAL	ADJ.P
Search Symbol	Search LOG2FC	Search PVAL	Search ADJ.P
0610009B22Rik	0.15407506639823598	0.375179028040943	0.71010000817246831
0610009L18Rik	-0.0314533496186336	0.8000506355583309	0.9247611817918641

DATA VISUALIZATION PORTAL

GeneLab Visualization

+ Add Visualization

 GLDS-4

Title: Microarray Analysis of Space-flown Murine Thymus Tissue

Organism(s): Mus musculus

Number of Genes: 20217

Number of Samples: 7

Group Selection

Group 1:

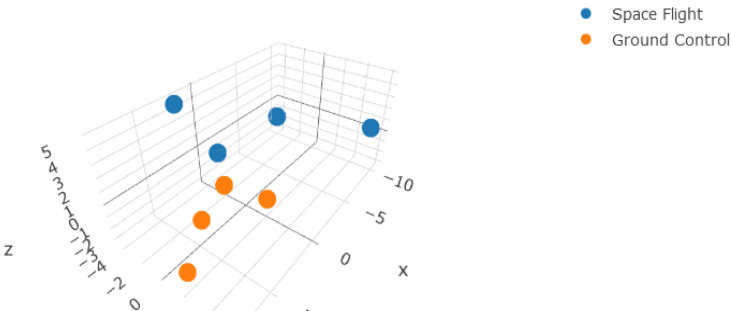
Ground Control

Group 2:

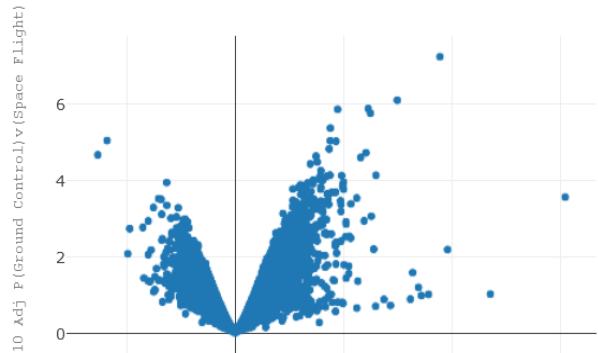
Space Flight

Update

GLDS-4 PCA



GLDS-4 Volcano



GLDS-4 DGE

Copy CSV Excel PDF Print

Maximum p-value:

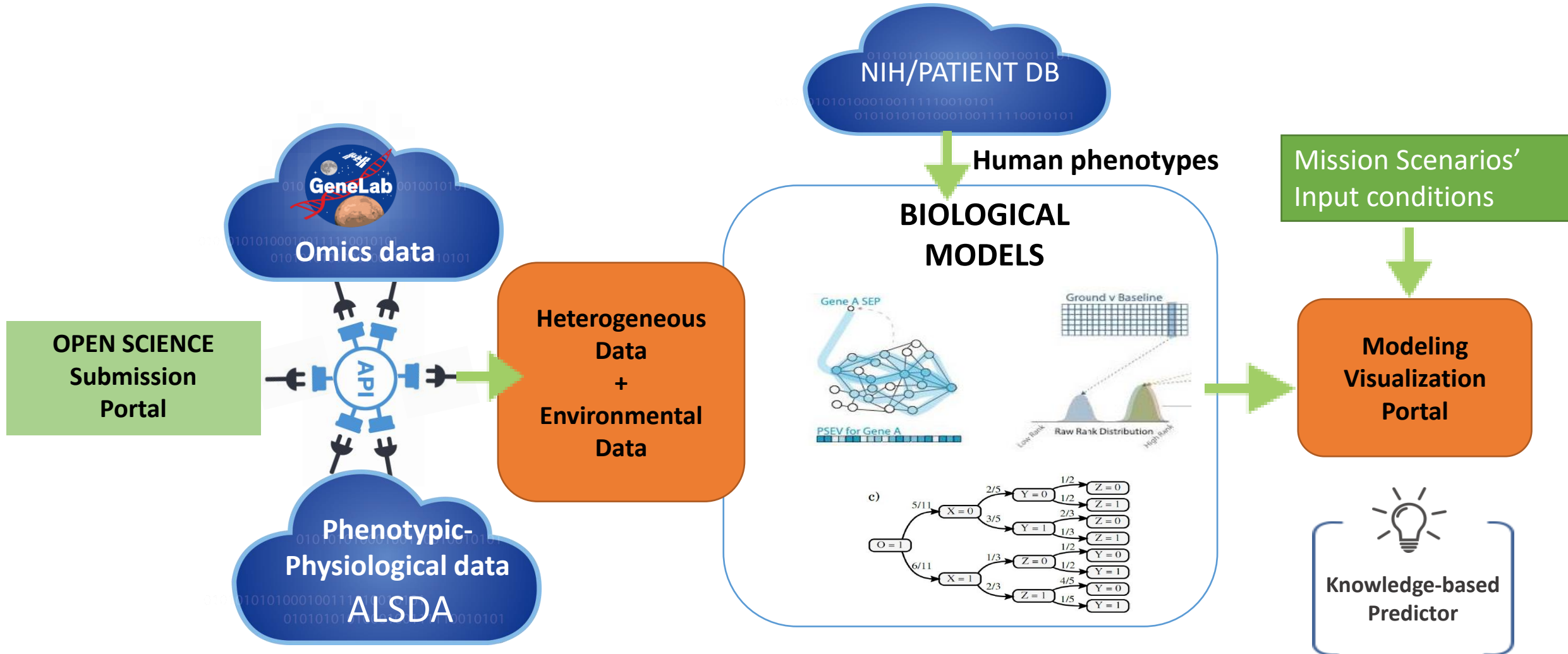
Search:

Symbol	LOG2FC	PVAL	ADJP
Search Symbol	Search LOG2FC	Search PVAL	Search ADJP
0610009B22Rik	0.15407506639823598	0.375179028040943	0.7101000817246831
0610009L18Rik	-0.0314533496186336	0.8000506355583309	0.9247611817918641
0610010F05Rik	0.124122725512421	0.208560405600702	0.662850561557224

19

Data Reuse Example: Data Modeling

AI ready FAIR data for modeling and risk assessment

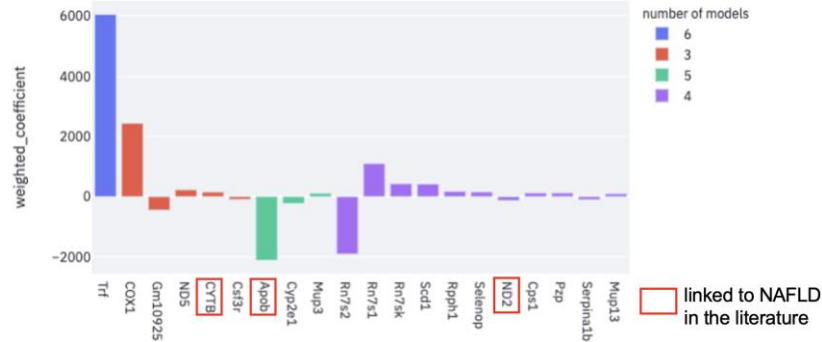


AI for Life in Space

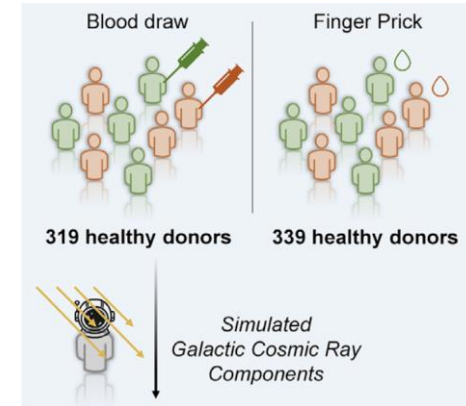
Causal Inference

Genes with a causal relationship to liver disease phenotype in spaceflown rodents

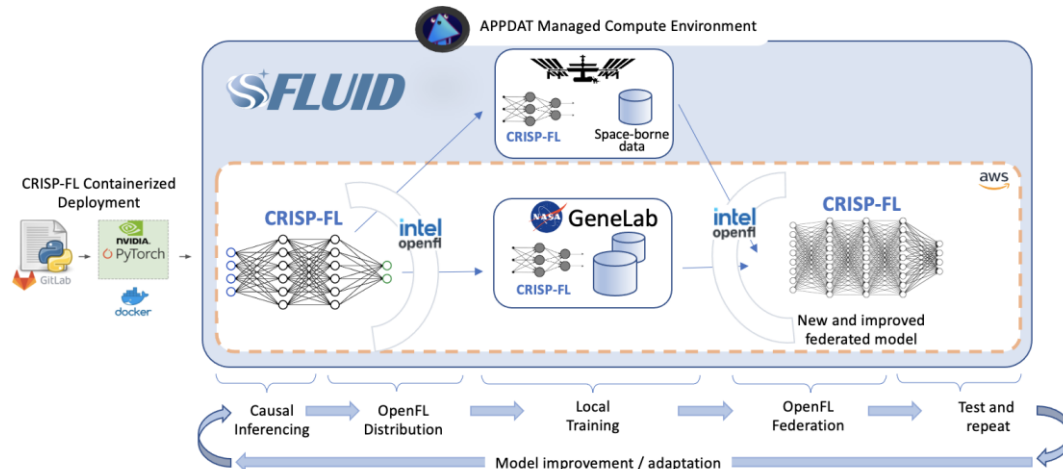
GLDS-47 GLDS-137
GLDS-48 GLDS-168



Genetic variants with a causal relationship to DNA damage and radiation sensitivity in human irradiated immune cells



Federated causal inference between Earth data and spaceborne data



2021 FDL Astronaut Challenge

A novel paradigm for generating human-relevant biomedical insights from observational datasets with limited size by leveraging and augmenting with animal model data

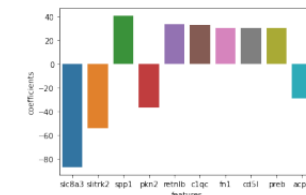


Figure 1: Top-10 IRM features according to feature coefficients with all human and mice gamma-radiation data.

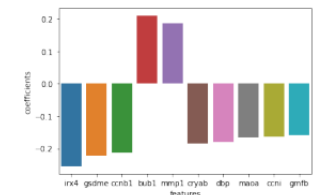


Figure 2: Top-10 IRM features according to feature coefficients with all human and mice heavy-ion radiation data.

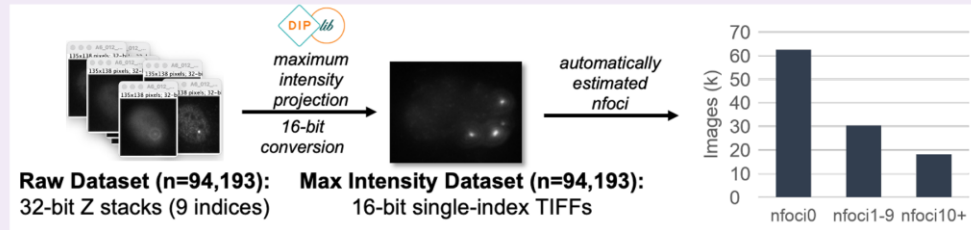
AI for Life in Space

Benchmark Microscopy Dataset

-Huge amounts of microscopy data in space biology, time consuming, manual annotate

Approach:

-Radiation exposure microscopy DNA damage dataset, distinct foci (PI: Costes)

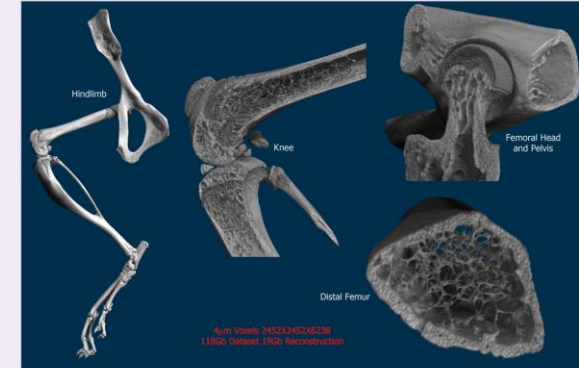


Benchmark microCT Dataset

-CT is key to assess space biological bone impacts

Approach:

-Mouse mCT dataset, both spaceflight and group (PI: Almeida)

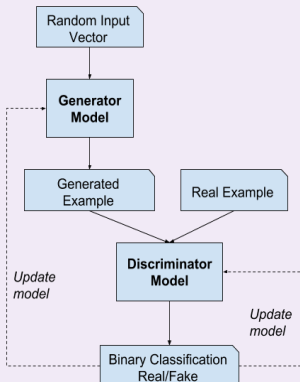


Benchmark RNA sequencing dataset

-RNA-seq data quantifies all expressed genes in a sample
-Classic example of high-D (~50k genes), low-N dataset

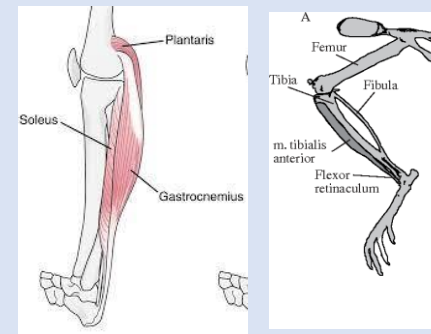
Approach:

1. Start with real RNA-seq data from GeneLab
2. Generate synthetic RNA-seq data using generative adversarial network (GAN)
3. Benchmark dataset will be high-D and high-N (synthetic)



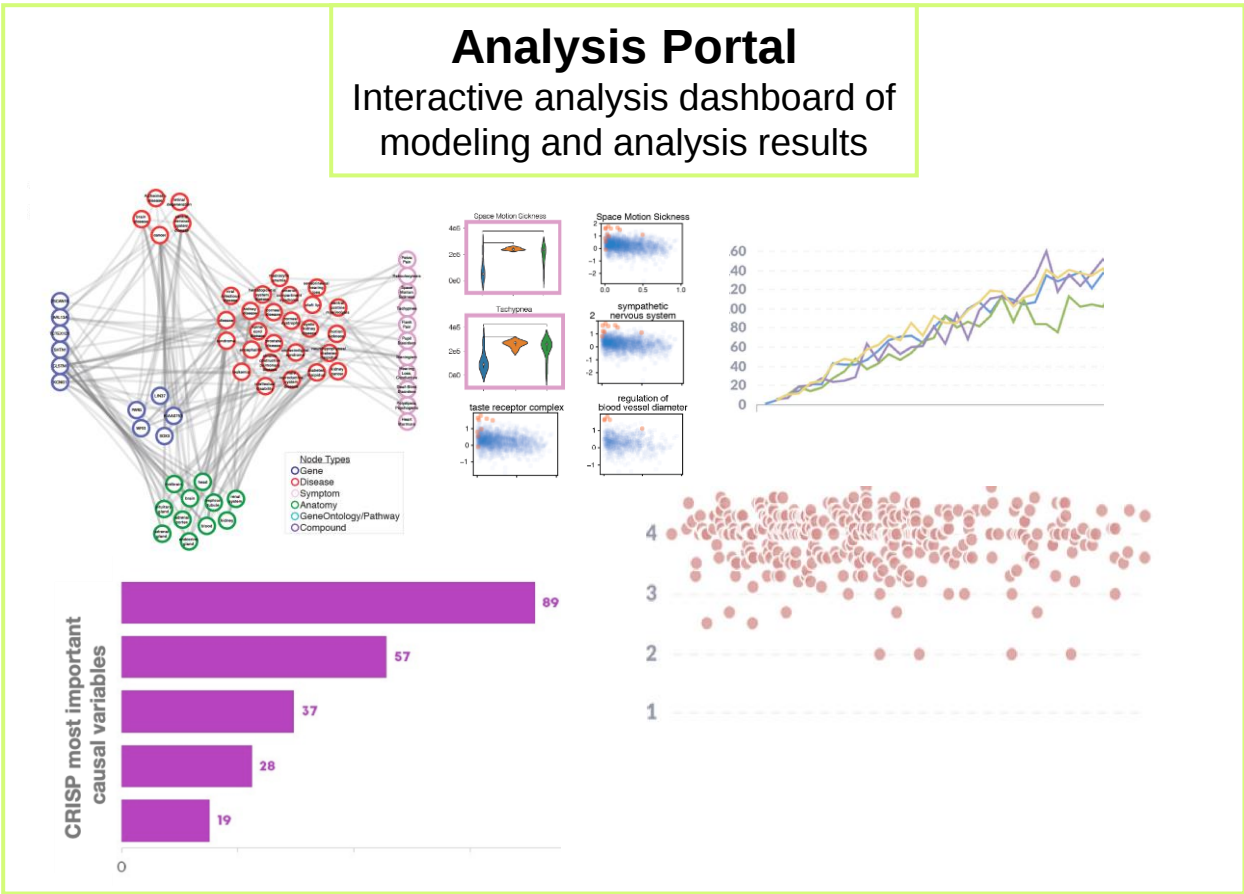
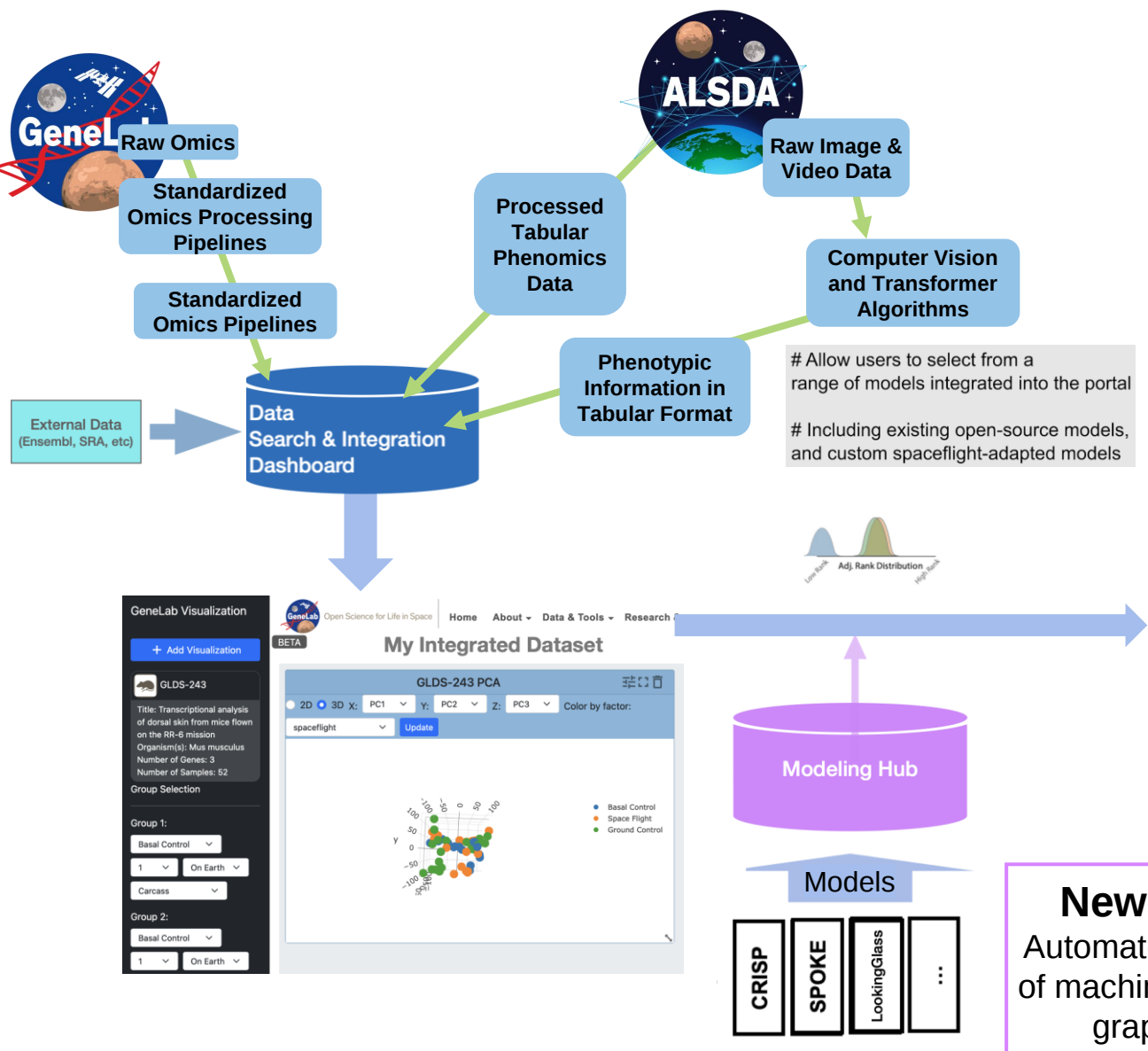
ML applications: e.g. causal inference to identify causal genes with relation to a pathogenic target variable

Omics and Phenomic Muscle Data using explainable xAI (Qlattice) to identify drivers of fast vs slow-twitch space muscle impacts



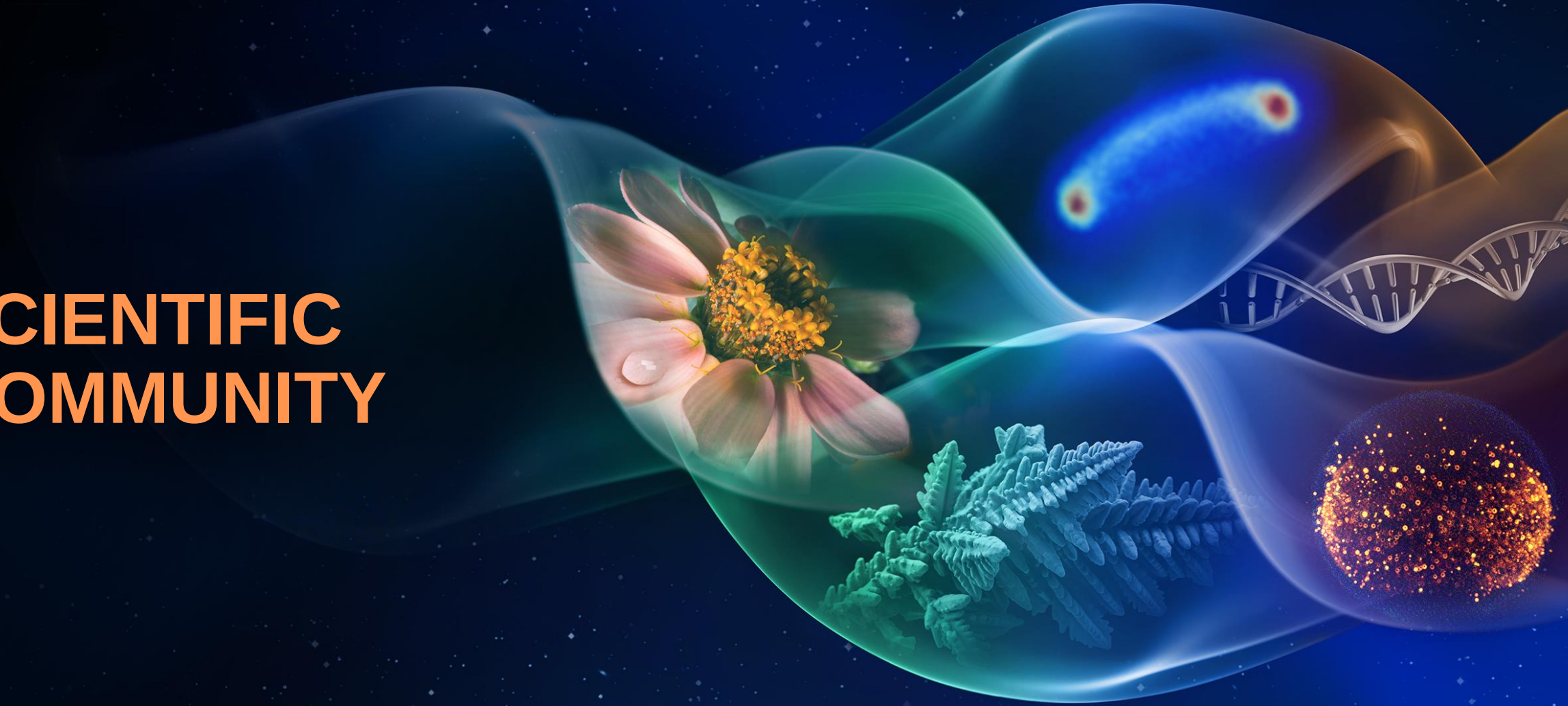
- Transcriptomics (RNA-seq)
- Methylation
- Proteomics
- Calcium uptake

Future: Streamlined Data Integration, Multimodal, Multihierarchical



New Modeling Hub
Automated pipelines for a suite of machine learning, knowledge graph, modeling tools

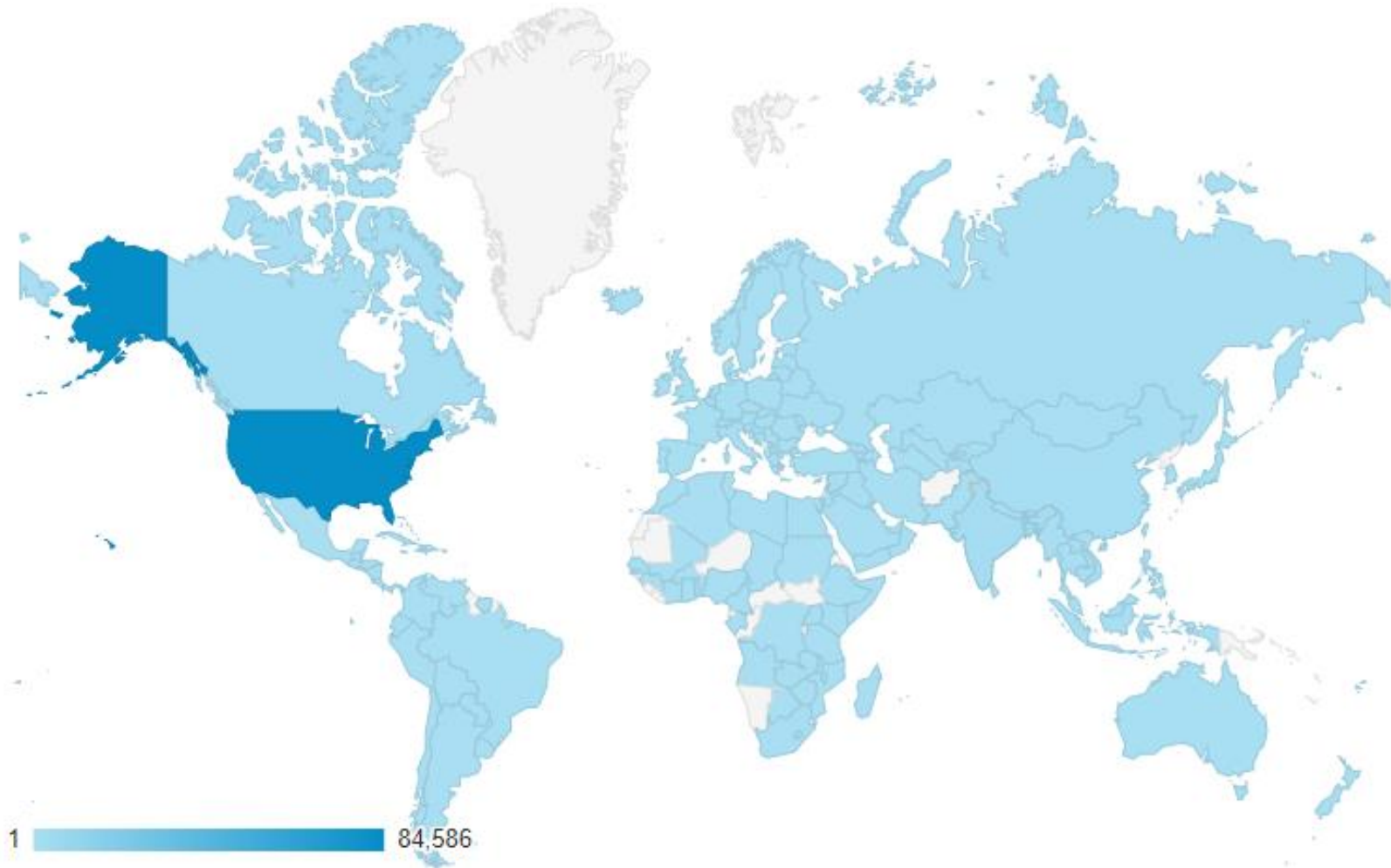
SCIENTIFIC COMMUNITY



Making data accessible worldwide

Top 20 Countries utilizing GeneLab:

1.  United States
2.  India
3.  Germany
4.  United Kingdom
5.  China
6.  Canada
7.  Singapore
8.  Japan
9.  Italy
10.  Spain
11.  France
12.  Turkey
13.  Australia
14.  Brazil
15.  South Korea
16.  Ghana
17.  Netherlands
18.  Mexico
19.  Russia
20.  Philippines



Analysis Working Groups

Analysis Working Groups (AWGs) consist of **400+ scientists** from multiple space agencies, international institutions, and industry. Scientists meet monthly with each group to provide feedback, develop standards, and analyze data. We invite you to join - <https://genelab.nasa.gov/awg/join!>

ANIMAL

68 members

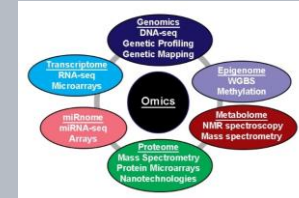
Facilitates the use of omics in understanding basic mechanisms by which animals and constituent tissues and cells adapt to the spaceflight environment.



MULTI-OMICS

181 members

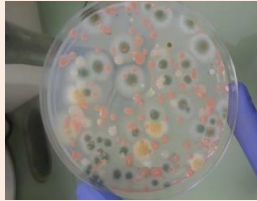
Interactions between the different omics to provide complete understanding of the entire system being studied.



MICROBES

50 members

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



PLANTS

58 members

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



ALSDA

150 members

- Recruiting members to support data standards and novel methods to analyze cross mission and platform data
- Establish standards for various assays' metadata
- Provide feedback on database and tools

SCIENTIFIC REPORTS

OPEN Microgravity induces proteomics changes involved in endoplasmic reticulum stress and mitochondrial protection

Byron J. Pagan¹, J. W. Thompson², Laura G. Soltis³, Roshly R. Khamidullayeva⁴, Matthew M. Foster⁵, Kagan M. M. Foster⁶, Lydia K. Soltis⁷, Nicholas Soltis⁸, Yarek N. Kildanov⁹, M. Arthur Moser¹⁰, Lisa S. Carroll¹¹ & Dawn E. Bowles¹²

TSMT (02/02/14)	LM	RM	RM Error	N
4 HU_Landmark_Vehicle	0.058349	0.003322	0.00374622	8
6 HU_Landmark_ZincOxide	0.059529	0.003322	0.00374622	8
2 HU_Au_Complex_Vehicle	0.04973	0.003322	0.00374622	8
7 HU_Au_Complex_ZincOxide	0.05049	0.003322	0.00374622	8
3 HU_Landmark_Vehicle	0.050507	0.003322	0.00374622	8
5 HU_Au_Complex_ZincOxide	0.049217	0.003322	0.00374622	8
8 HU_Au_Complex_Vehicle	0.050507	0.003322	0.00374622	8
9 HU_Au_Complex_ZincOxide	0.050507	0.003322	0.00374622	8
10 HU_Au_Complex_ZincOxide	0.050507	0.003322	0.00374622	8
11 HU_Au_Complex_ZincOxide	0.050507	0.003322	0.00374622	8
12 HU_Au_Complex_ZincOxide	0.050507	0.003322	0.00374622	8
13 HU_Au_Complex_ZincOxide	0.050507	0.003322	0.00374622	8
14 HU_Au_Complex_ZincOxide	0.050507	0.003322	0.00374622	8
15 HU_Au_Complex_ZincOxide	0.050507	0.003322	0.00374622	8
16 HU_Au_Complex_ZincOxide	0.050507	0.003322	0.00374622	8
17 HU_Au_Complex_ZincOxide	0.050507	0.003322	0.00374622	8
18 HU_Au_Complex_ZincOxide	0.050507	0.003322	0.00374622	8
19 HU_Au_Complex_ZincOxide	0.050507	0.003322	0.00374622	8
20 HU_Au_Complex_ZincOxide	0.050507	0.003322	0.00374622	8
21 HU_Au_Complex_ZincOxide	0.050507	0.003322	0.00374622	8
22 HU_Au_Complex_ZincOxide	0.050507	0.003322	0.00374622	8
23 HU_Au_Complex_ZincOxide	0.050507	0.003322	0.00374622	8
24 HU_Au_Complex_ZincOxide	0.050507	0.003322	0.00374622	8
25 HU_Au_Complex_ZincOxide	0.050507	0.003322	0.00374622	8
26 HU_Au_Complex_ZincOxide	0.050507	0.003322	0.00374622	8
27 HU_Au_Complex_ZincOxide	0.050507	0.003322	0.00374622	8
28 HU_Au_Complex_ZincOxide	0.050507	0.003322	0.00374622	8
29 HU_Au_Complex_ZincOxide	0.050507	0.003322	0.00374622	8
30 HU_Au_Complex_ZincOxide	0.050507	0.003322	0.00374622	8
31 HU_Au_Complex_ZincOxide	0.050507	0.003322	0.00374622	8
32 HU_Au_Complex_ZincOxide	0.050507	0.003322	0.00374622	8
33 HU_Au_Complex_ZincOxide	0.050507	0.003322	0.00374622	8
34 HU_Au_Complex_ZincOxide	0.050507	0.003322	0.00374622	8
35 HU_Au_Complex_ZincOxide	0.050507	0.003322	0.00374622	8
36 HU_Au_Complex_ZincOxide	0.050507	0.003322	0.00374622	8
37 HU_Au_Complex_ZincOxide	0.050507	0.003322	0.00374622	8
38 HU_Au_Complex_ZincOxide	0.050507	0.003322	0.00374622	8
39 HU_Au_Complex_ZincOxide	0.050507	0.003322	0.00374622	8
40 HU_Au_Complex_ZincOxide	0.050507	0.003322	0.00374622	8
41 HU_Au_Complex_ZincOxide	0.050507	0.003322	0.00374622	8
42 HU_Au_Complex_ZincOxide	0.050507	0.003322	0.00374622	8
43 HU_Au_Complex_ZincOxide	0.050507	0.003322	0.00374622	8
44 HU_Au_Complex_ZincOxide	0.050507	0.003322	0.00374622	8
45 HU_Au_Complex_ZincOxide	0.050507	0.003322	0.00374622	8
46 HU_Au_Complex_ZincOxide	0.050507	0.003322	0.00374622	8
47 HU_Au_Complex_ZincOxide	0.050507	0.003322	0.00374622	8
48 HU_Au_Complex_ZincOxide	0.050507	0.003322	0.00374622	8
49 HU_Au_Complex_ZincOxide	0.050507	0.003322	0.00374622	8
50 HU_Au_Complex_ZincOxide	0.050507	0.003322	0.00374622	8



Training the Next Generation of Scientists



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>



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>



Space Life Sciences Training Program (SLSTP): Provides undergraduate students entering their junior or senior years, and entering graduate students, with professional experience in space life science disciplines.

Learn more and apply at: <https://www.nasa.gov/ames/research/space-life-sciences-training-program>



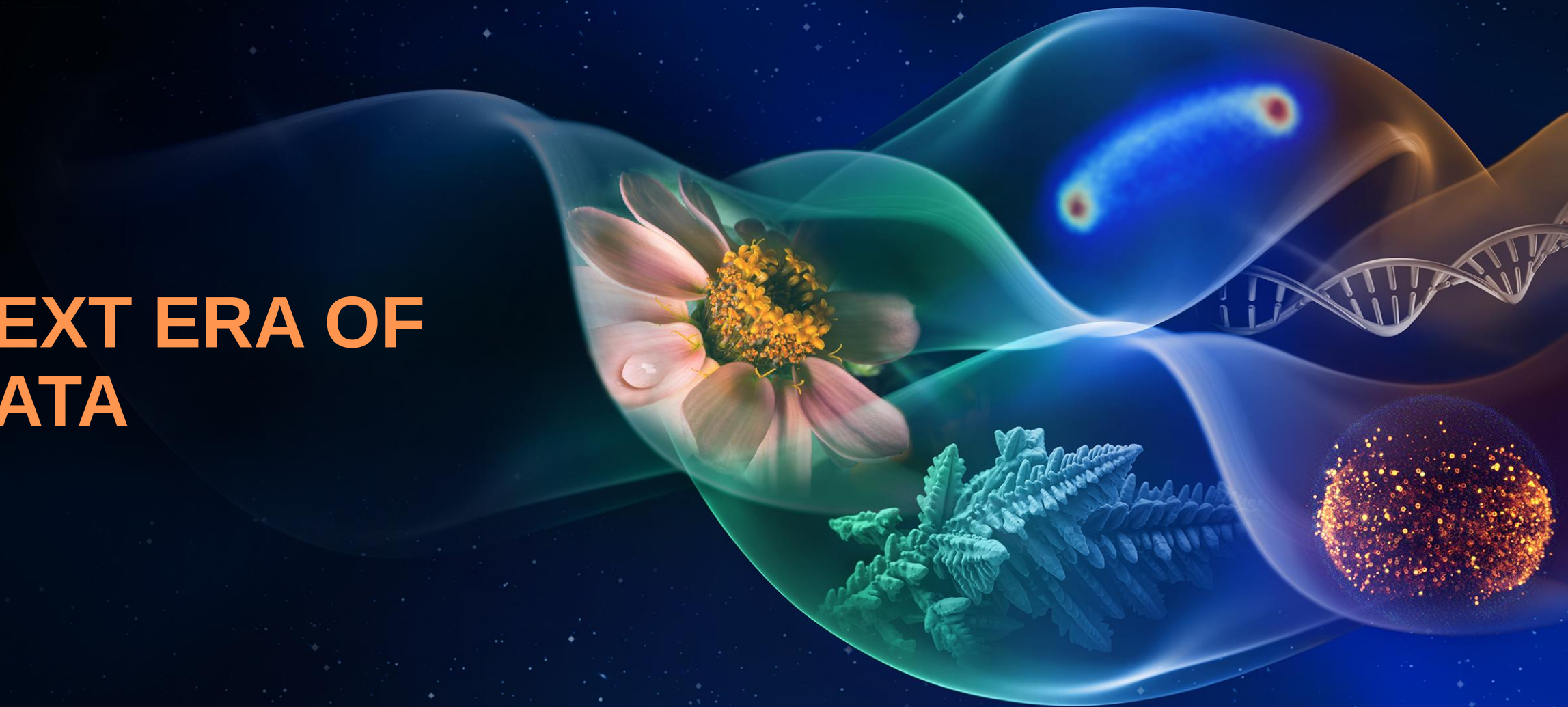
Spaceflight Technology, Applications, and Research (STAR): The annual course targets principal investigators (PIs), senior research scientists, and postdoctoral scholars and aims to facilitate their entry to space biology and preparation for conducting spaceflight experiments using NASA and commercial platforms.

Applications open until May 30th: <https://science.nasa.gov/biological-physical/programs/star>



NASA Postdoctoral Program: If you're an early-career or senior scientist, apply to the NASA Postdoctoral Program to help pursue NASA's mission and experience the world's most diverse technology and expertise.

NEXT ERA OF DATA



Commercial and International Partners

Open access data in GeneLab, **accelerates the pace of scientific discovery** and thus **enables exploration and development of countermeasures** by using systems biology approaches to identify cross-species relevance, mechanisms, biomarkers and potential health risks for LEO and beyond LEO (Lunar and Mars Exploration).

To increase usability and accessibility, space biological research should be available through a common portal. It's important to work closely with our commercial and international partners to ensure data across agencies and platforms are harmonized and comparable. In addition, systems already developed can be utilize not duplicating efforts and costs for data sharing.

For example, **GeneLab already hosts datasets from multiple space agencies (ESA, JAXA, CSA) and from CASIS funded experiments.**

In order to continue maximizing the return of spaceflight experiments, **we encourage commercial and international partners to utilize the GeneLab and ALSDA Data Systems for their research data, both for open and proprietary data.**



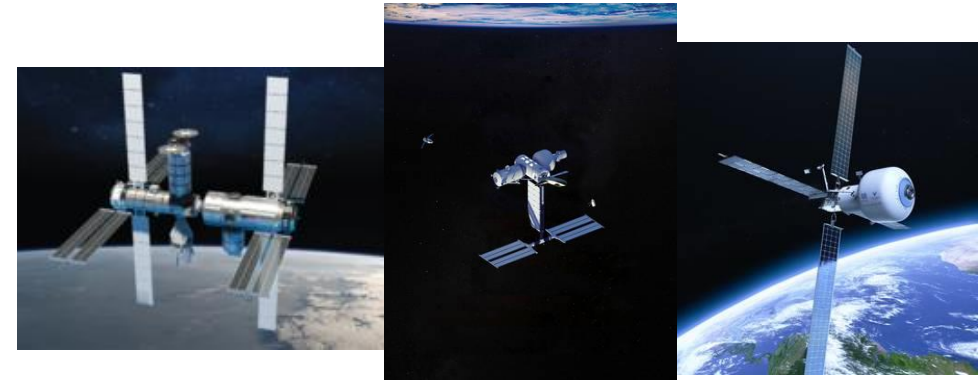
Proprietary data

Currently all the data that is published, is openly accessible worldwide. GeneLab/ALSDA has already in place a mechanism for standardize data curation and account-based access. This capability can be expanded in the future for:



Astronaut/Human space-relevant data

- Develop a controlled access data portal
- Create an automatic data access request workflow
- Scientists will be able to translate their findings from model organisms to human outcomes



Commercial LEO Destinations

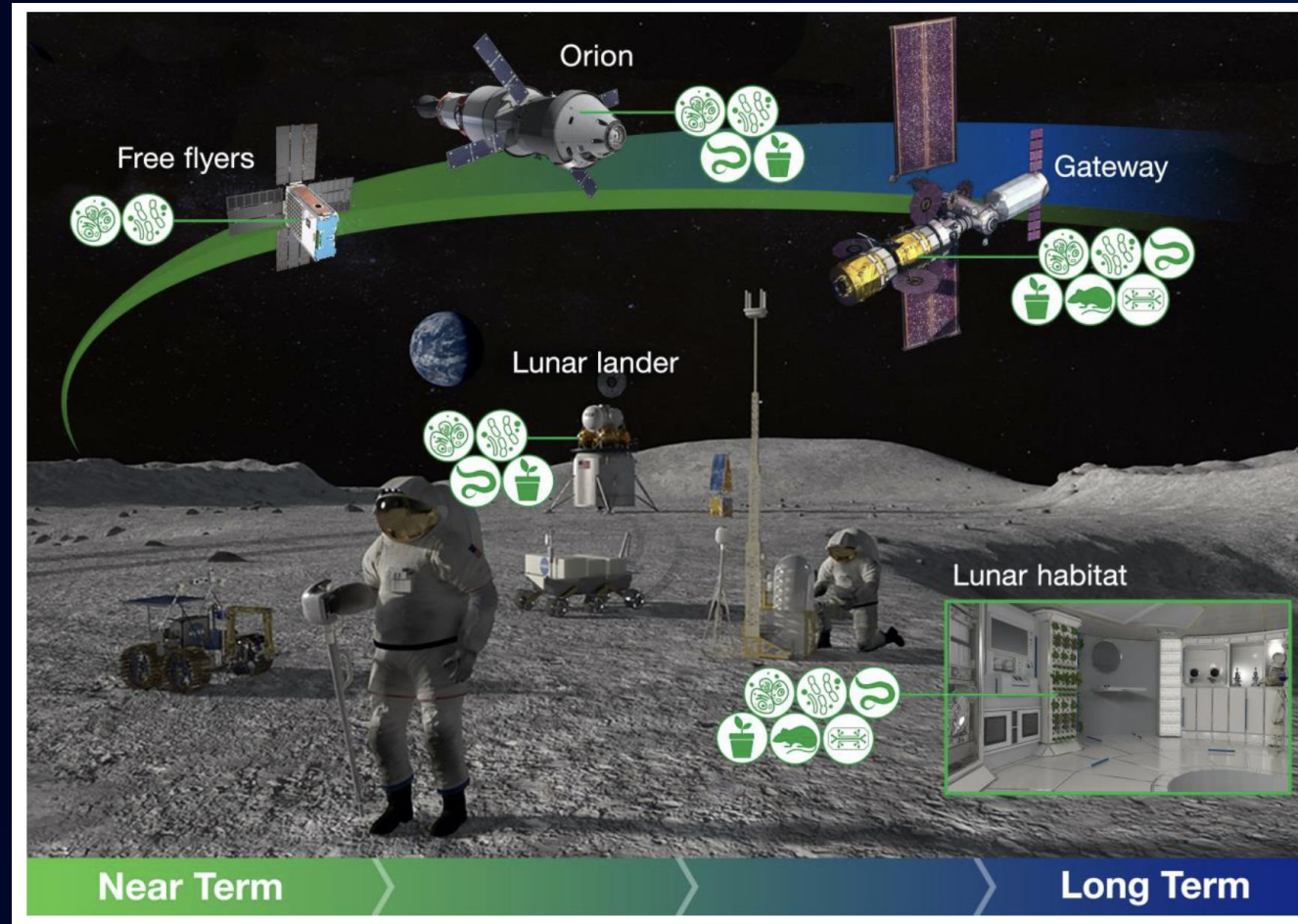
- Develop a controlled access data portal and an automatic data access request workflow for “closed” data
- Enable access to all spaceflight research a centralized data repository

Data should be as open as possible, as closed
as necessary.

Metadata can always be open.

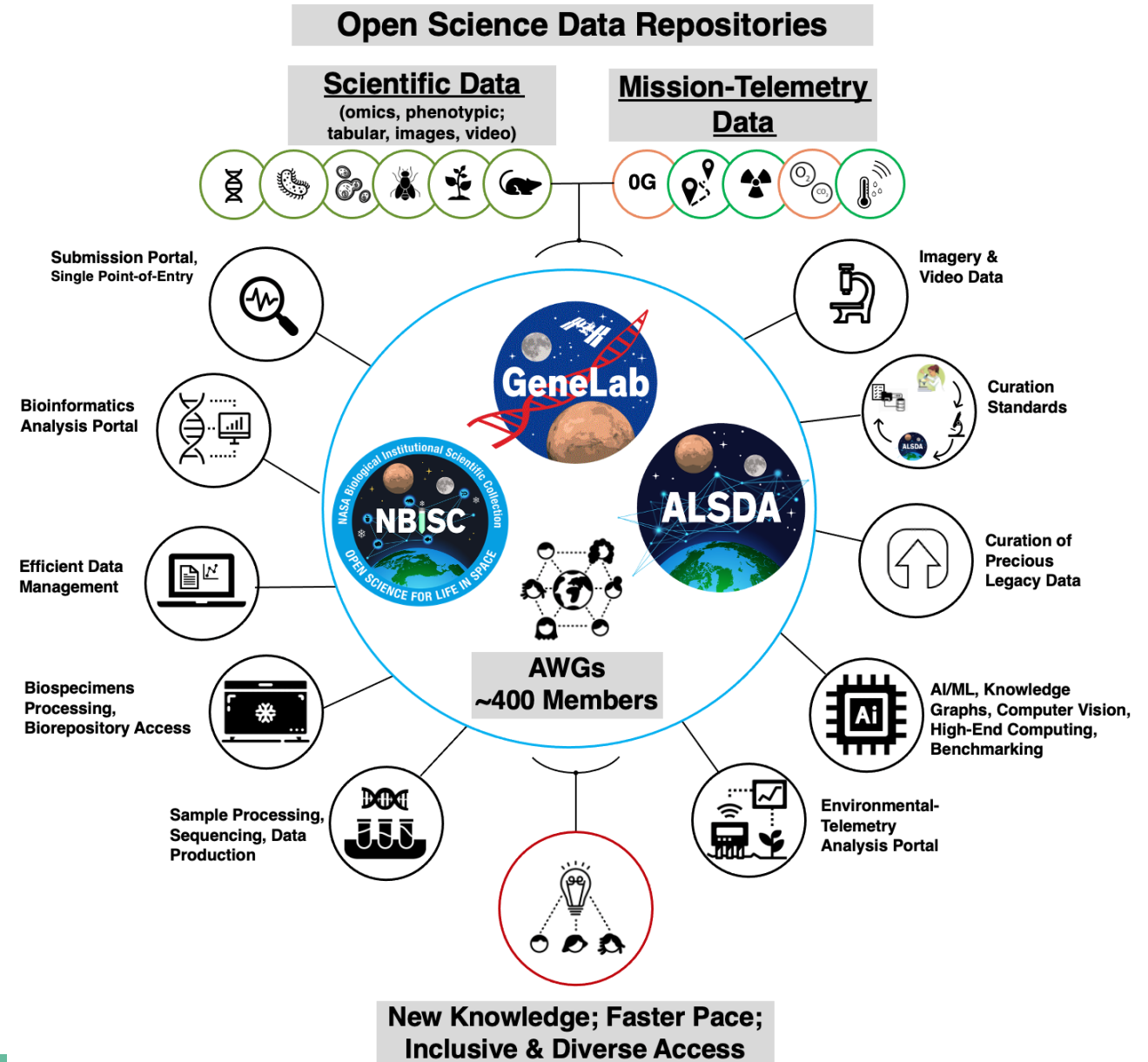
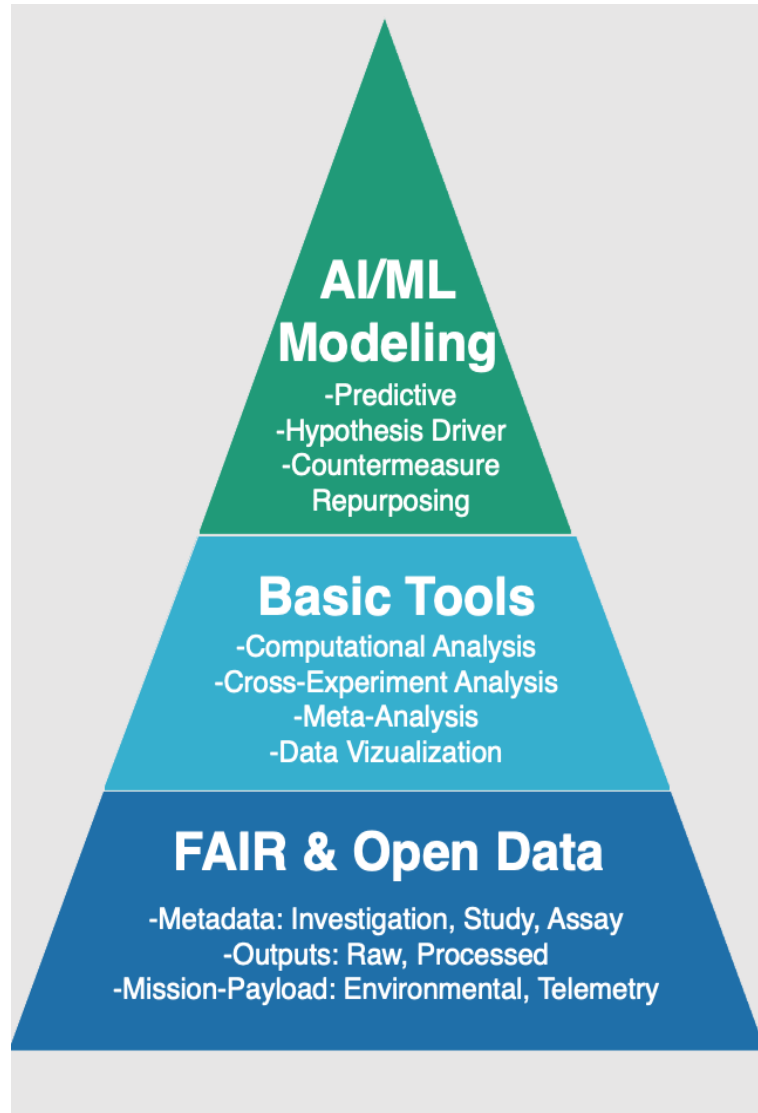
-Shelly Stall, TEM April 2020

Beyond LEO



With limited sample return for various platforms, we are preparing new metadata templates and data streams to host and publish data from future lunar experiments and beyond.

Big Picture for Open Science for Life in Space



THANK YOU!



Open Science for Life in Space projects are funded by the NASA Space Biology program and NASA Human Research Program.



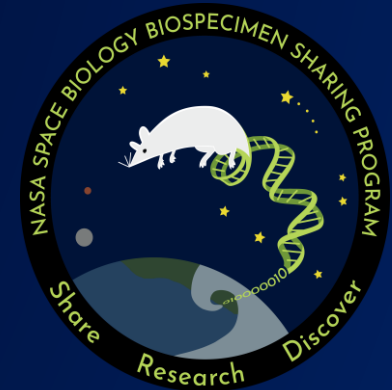
 <https://www.nasa.gov/ames/research/space-biosciences/alsda>



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



 <https://www.nasa.gov/ames/research/space-biosciences/nbisc>



 <https://science.nasa.gov/biological-physical/space-biology-biospecimen-sharing-program>

Feel free to email me at samrawit.g.gebre@nasa.gov with any questions or links from this presentation.