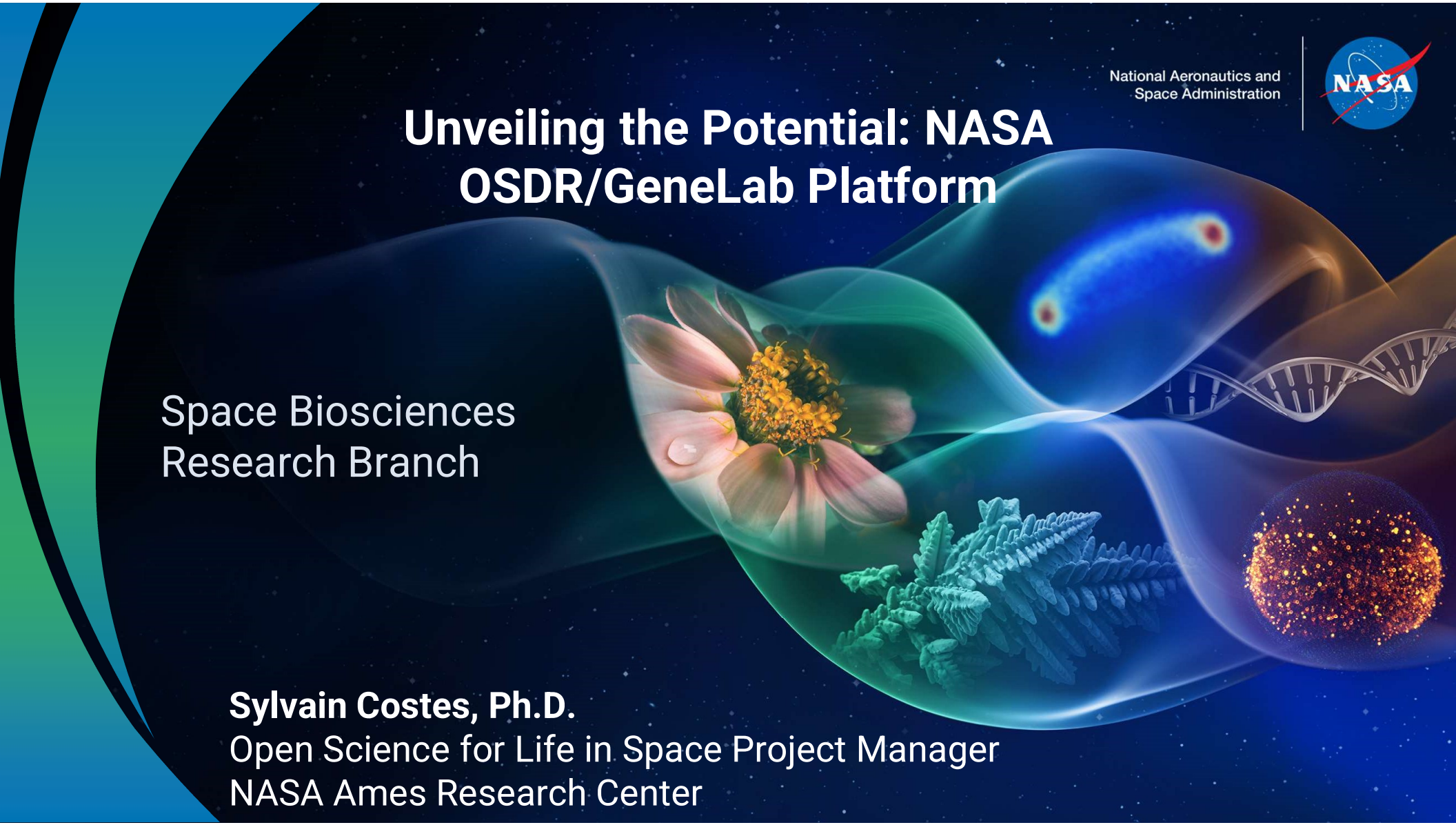
National Aeronautics and Space Administration

Unveiling the Potential: NASA OSDR/GeneLab Platform



Space Biosciences
Research Branch

Sylvain Costes, Ph.D.
Open Science for Life in Space Project Manager
NASA Ames Research Center

OSDR Collaborates with ESA SciSpace

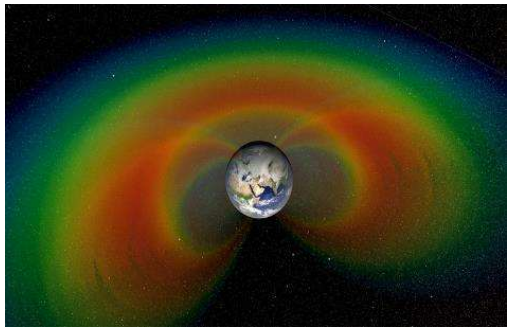
- ESA will host a 5-day virtual (Sept. 23-27) ESA Space Omics Hackathon to tackle space exploration challenges using NASA GeneLab omics datasets.
- GeneLab Scientist will serve as a judge and mentor to winning team.
- **Prizes:**
  - First-place prize will be an exciting opportunity – an online internship with NASA GeneLab!
  - The first, second and third place winners will have an opportunity to present at *NASA-GeneLab Multi-omics Working Group online*.

More info: https://scispace.esa.int/event/esa-space-omics-hackathon/

Biologically Relevant Environmental Factors Encountered in Spaceflight

Microgravity/Reduced Gravity

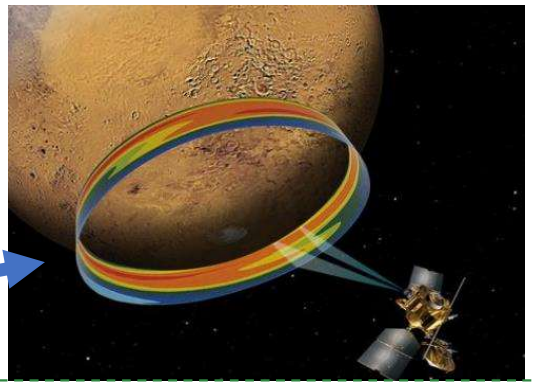
Ionizing Radiation



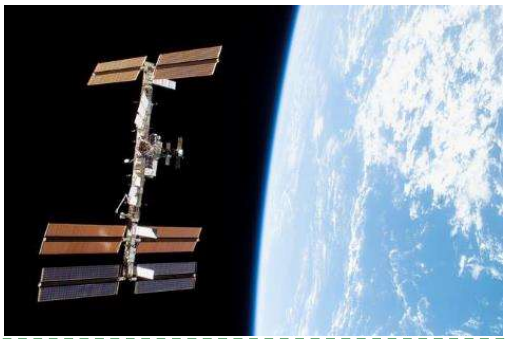
Altered Day/Night Cycles:
Circadian Rhythm Changes

Altered Temperature
and Atmosphere

- Elevated CO_2
- Reduced atmospheric pressure and elevated volumetric fraction of oxygen



Isolation



Thriving in Deep Space (TIDES)

Biological effects of multiple **deep-space stressors**

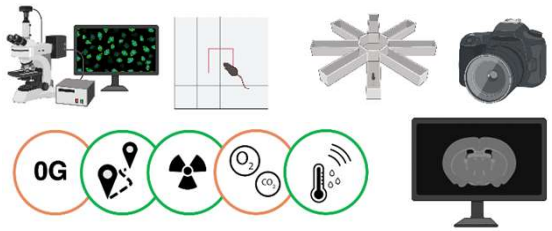
1. Radiation
2. Gravity
3. Temperature & Atmosphere
4. Day & Night Light Cycles
5. Isolation
6. Regolith/dust

Transformative biological science and exploration applications

1. Animal biology	Vertebrate and invertebrate models to probe analogous changes in humans
2. Plant Biology	From plant models to crops to sustain life for long-term human habitation
3. Microbiology	How it influences animals and plants in space

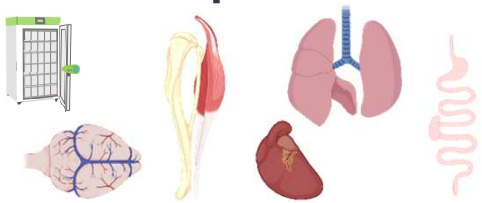
NASA's Open Science Data Repository

Physiological/Phenotypic/Imaging/ Environmental Telemetry Data



Molecular/Omics Data

Biospecimens



ALSDA

GeneLab

NBISC

NASA Open Science Data Repository (OSDR)
osdr.nasa.gov/bio

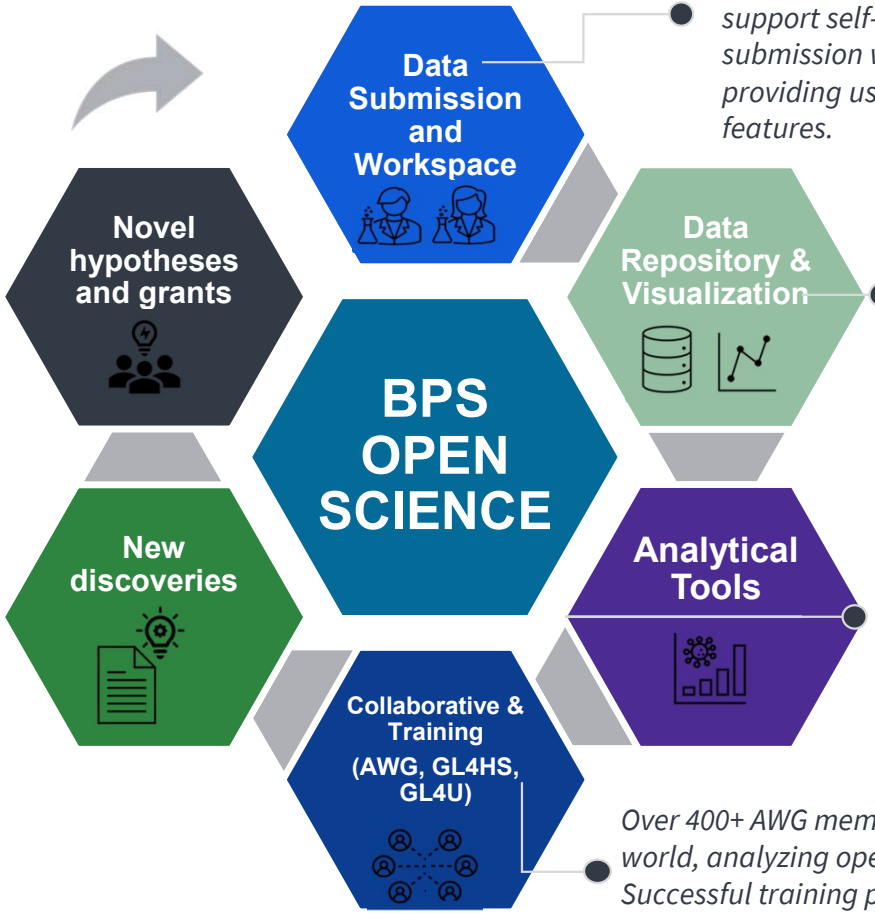
- **Single Submission Portal (BDME)**
- **User Interface/Website Tool for RDSAs (Research Data Submission Agreements)**
- **Workspace**
- **Maximally Open Access with Necessary Controls for Sensitive Data**
- **Data Maximally FAIR**

Open Science Data Repository ecosystem

Multi-project web-based submission portal to support self-service metadata curation and data submission within FAIR Guidelines. Workspace providing users file storage in S3 and sharing features.

Open access data enables discovery of new hypotheses and new ideas for grant proposal. Data from those new research experiments are generated and deposited back into GeneLab.

Open access and well curated data repositories including data from various model organisms from microbes to plants to humans covering the full range of biological assay technologies. Interactive visualization portal to enable knowledge discovery.

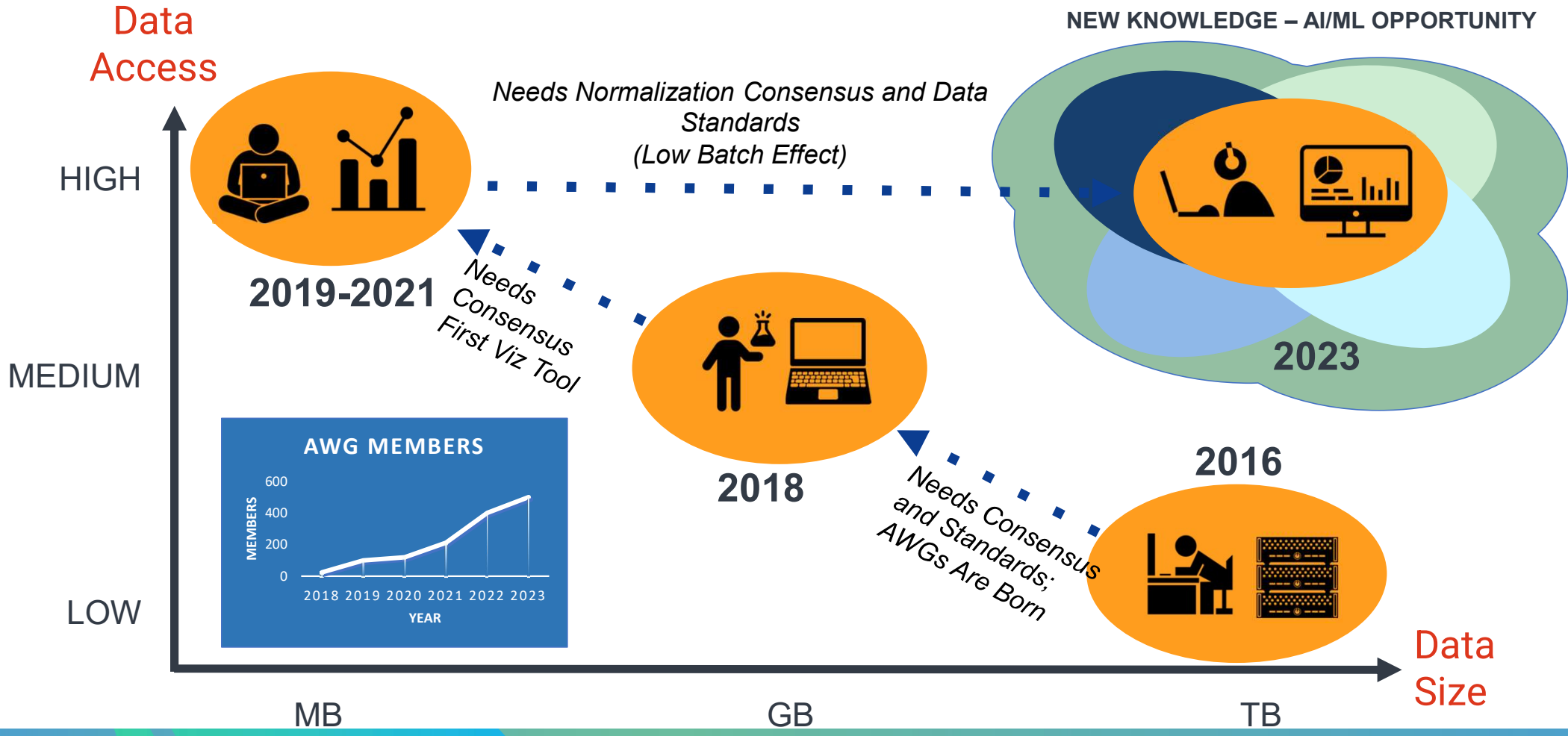


Enabling new publications and research through data reuse and performing meta-analysis.

Toolshed of bioinformatics tools with user-friendly interface for data analysis at any level (students, citizen scientists, PIs).

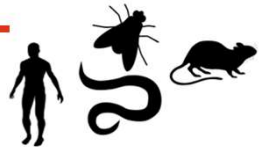
Over 400+ AWG members from around the world, analyzing open science data. Successful training programs developed for high school and university students.

A Brief History of the Open Science AWGs



Open Science Analysis Working Groups

ANIMAL



104 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

307 members

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

MICROBES

98 members

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

PLANTS

88 members

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

AI/ML

96 members

Focuses on developing data AI-readiness guidelines, algorithm and automation development, and developing ethical guidelines to increase trust and explainability surrounding AI in space biology.

ALSDA

138 members

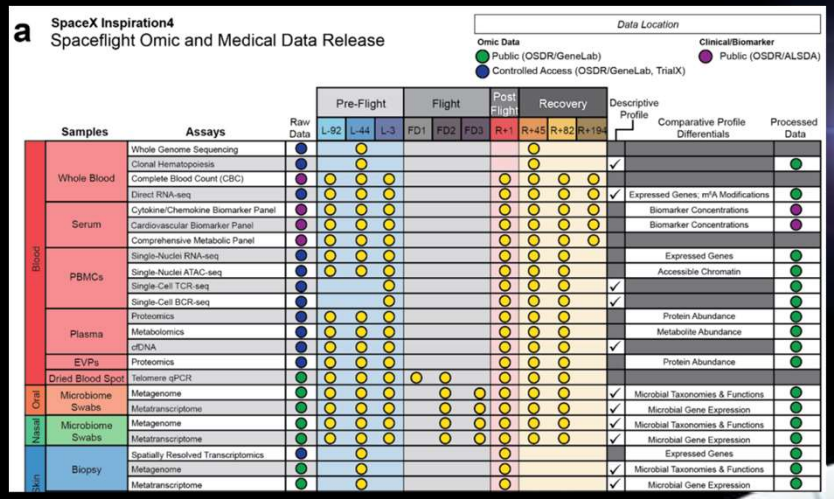
Feedback on science data and metadata standards for physiological, phenotypic, and behavioral datasets to be reusable. Datasets span from raw to processed-results data, and across tabular, bioimaging, and video formats.

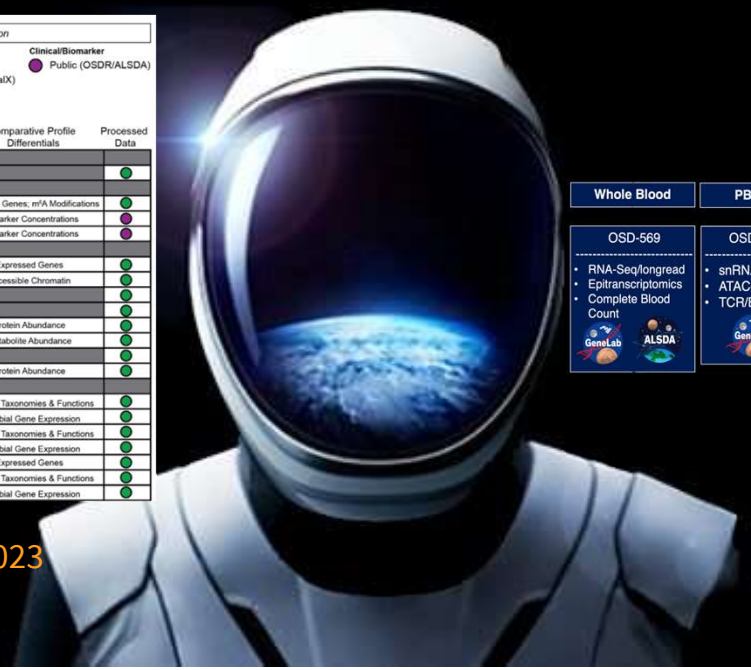
*Consist of **500+ scientists** from multiple space agencies, international institutions, and industry. Scientists meet monthly with each group to provide feedback, develop standards, and analyze data.*

We invite you to join - https://genelab.nasa.gov/awg/join!

Engaging with commercial entities: I4 Human Data into OSDR

INSPIRATION4





Sample Type

Whole Blood	PBMC	Blood Plasma	Skin Swabs	Environmental	Skin Biopsy	Blood Serum
OSD-569	OSD-579	OSD-571	OSD-572	OSD-573	OSD-574	OSD-575
• RNA-Seq/longread • Epitranscriptomics • Complete Blood Count	• snRNA-seq • ATAC-seq • TCR/BCR Seq	• Proteomics • Metabalome • Mass Spec • cfDNA	• Metagenomics • Metatranscriptomics	• Metagenomics • Metatranscriptomics	• Spatial Transcriptomics • Metagenomics • Metatranscriptomics	• Comprehensive Metabolic Panel • Cytokine Panel
GeneLab ALSDA	GeneLab	GeneLab	GeneLab	GeneLab	GeneLab	ALSDA

Assay Type

OSDR able to ingest and properly curate all data for reuse

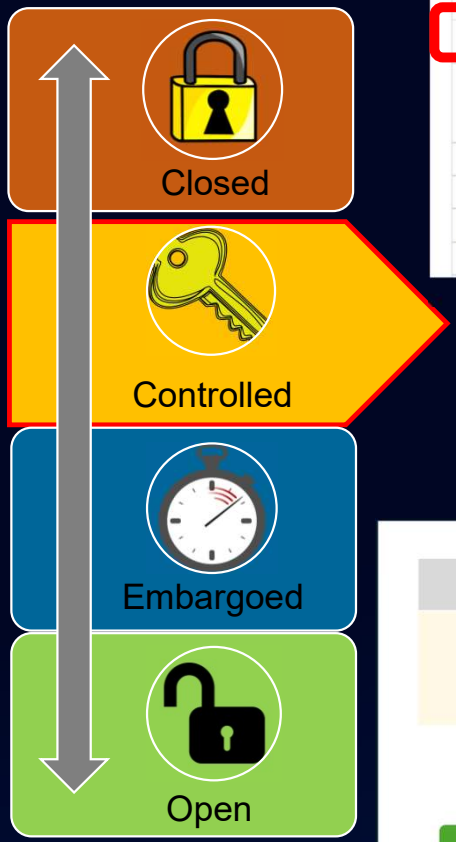
- Credit: Chris Mason's Lab; Overbey et al., 2023
- In Peer Review at *Nature*

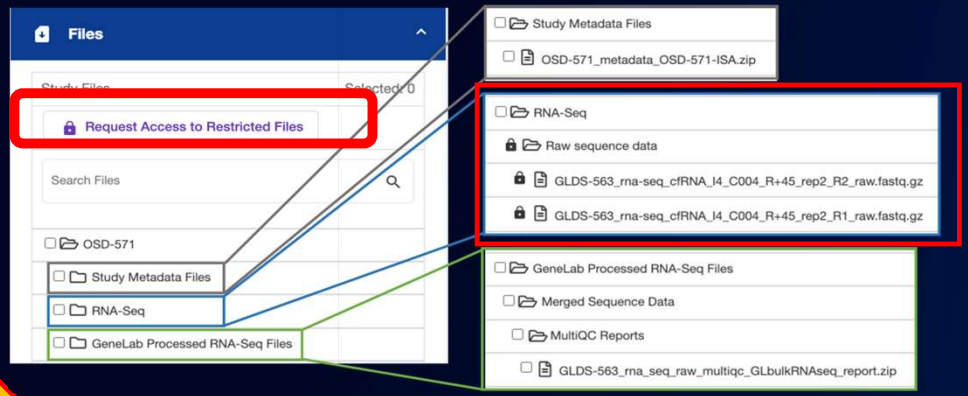
Request Process inspired by NIH's Database of Genotypes and Phenotypes (dbGaP)

1. Request button In OSDR
2. Data Request Package w/ Data Use Agreement
3. PI Verification
4. Review Board by Stakeholders
5. Approved/Denied
6. Access granted
7. Access removed at end of IRB

Controlled Access Data

The Continuum





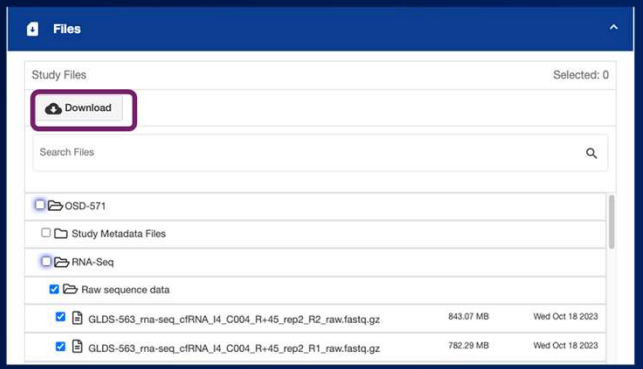
If Approved

Access Request Process for Private Data

1	2	3	4	5	6	7	8
Identify the Dataset	Submit Request	Verify Affiliation	Review by OSDR	Review by Committee	Data Access	Progress Report	Close Out
Request Button clicked in OSD dataset	OSDR Data Request Package sent w/ Data Use Agreement	OSDR Follow-Up with your institution	Data Access Committee review with Specific Stakeholders (OSDR, PIs, etc)		Approved / Denied Secure Access	Annual update	Access removed at end of access period, upon participant request, or if conditions of Data Use Agreement are violated

Process inspired by NIH's Database of Genotypes and Phenotypes (dbGaP)

Data Access Info Available On DAR Portal

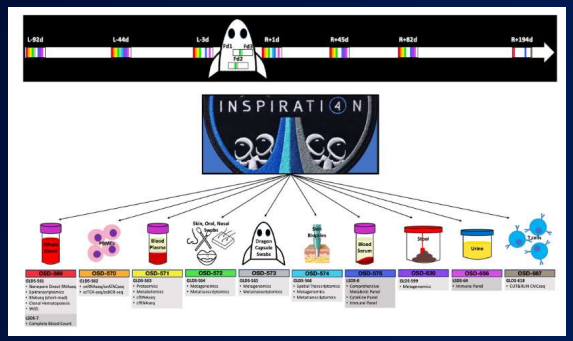


AWGs Accelerate Data Mining & Publications

Nature Portfolio Collection 2024:

Space Omics and Medical Atlas across orbits (SOMA)

OSDR hosts the first curated set of commercial astronaut data

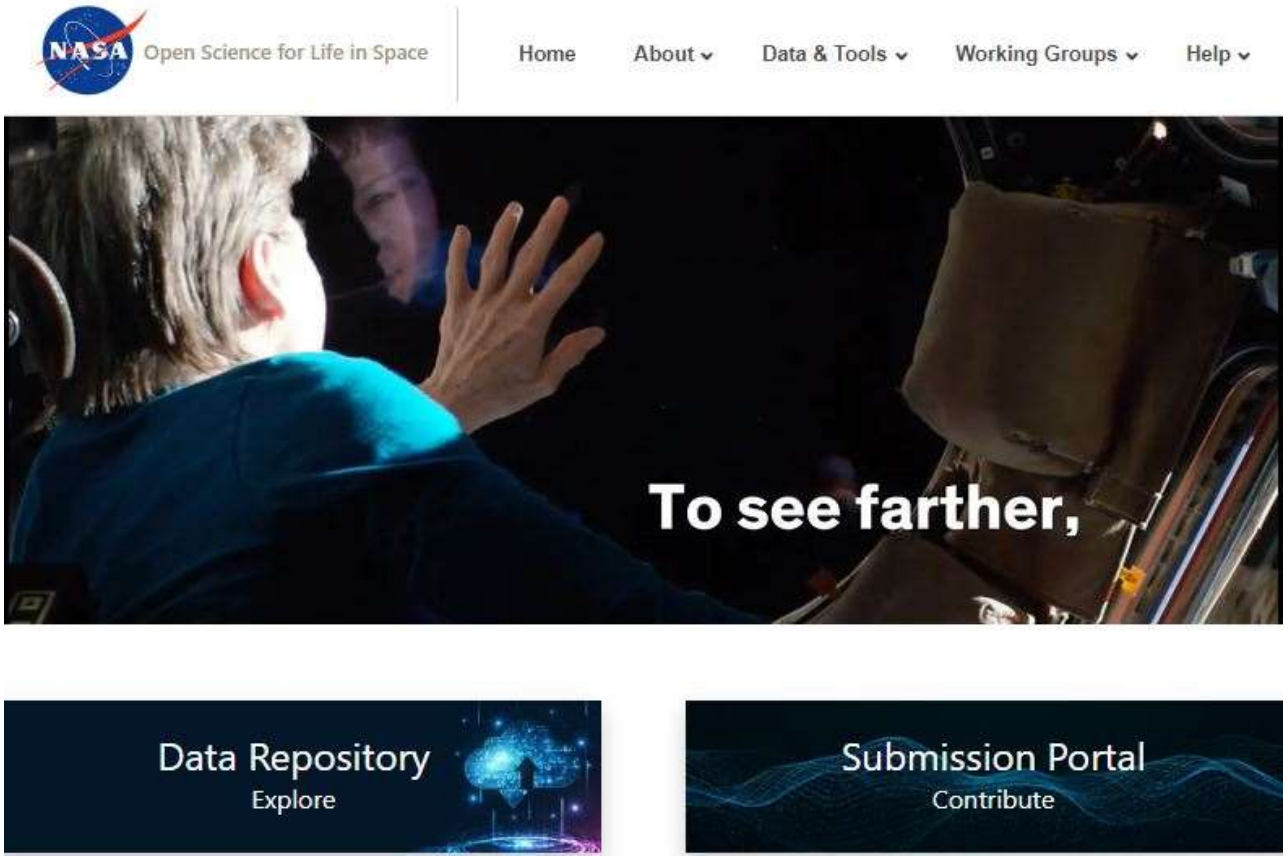


111 AWG Members participated in the Nature Press Package across **43 publications**

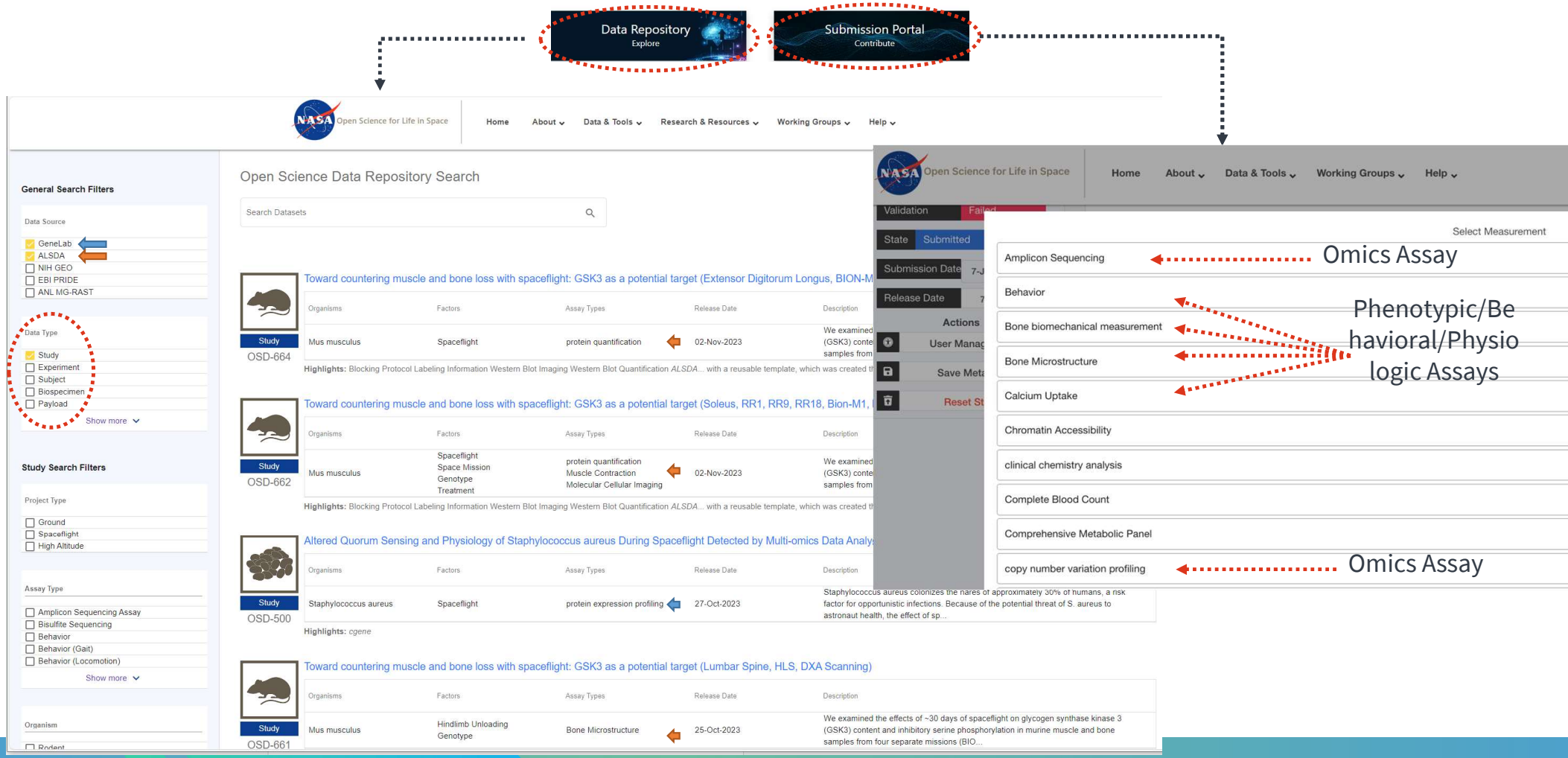
13 OSDR Members participated in the Nature Press Package across **17publications**

List of 80+ publications enabled by OSDR
https://osdr.nasa.gov/bio/data/publications.html

Enhancing Search and Data Submission



Enhancing Search and Data Submission



OSDR Database (GeneLab and ALSDA)

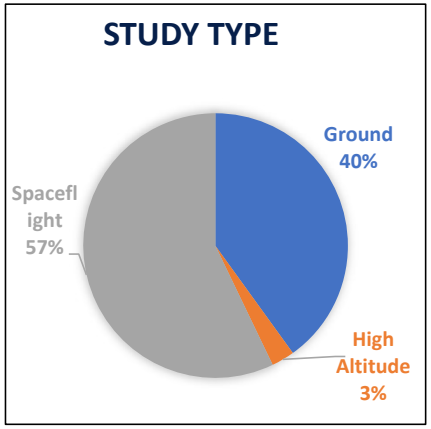
506
Studies

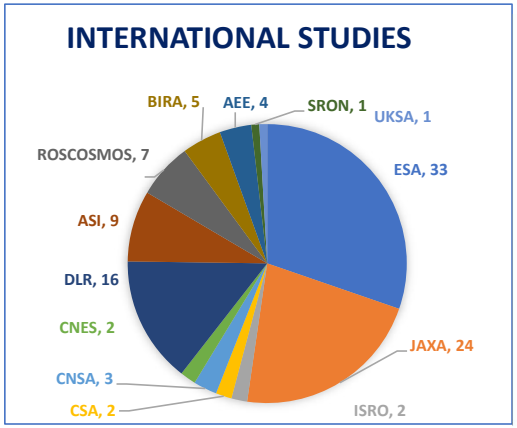
970
Datasets

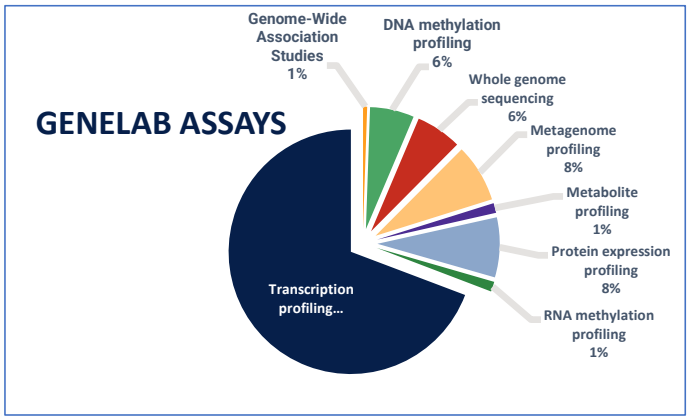
45
Species

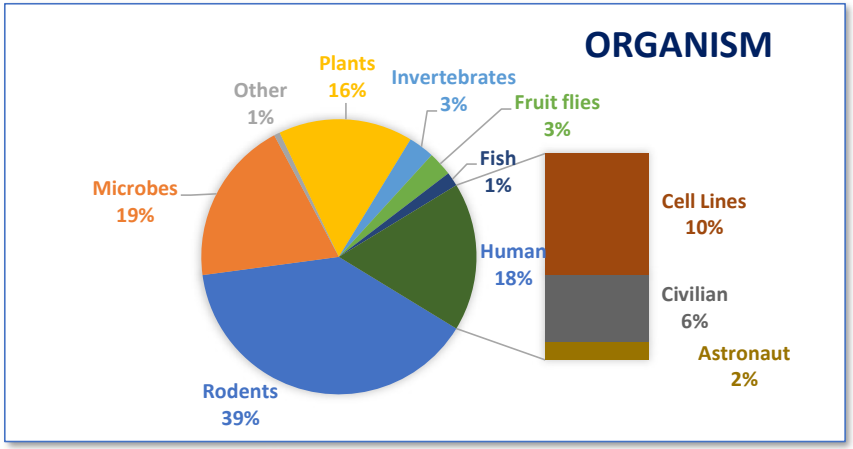
>80
Assays

>160TB
Data

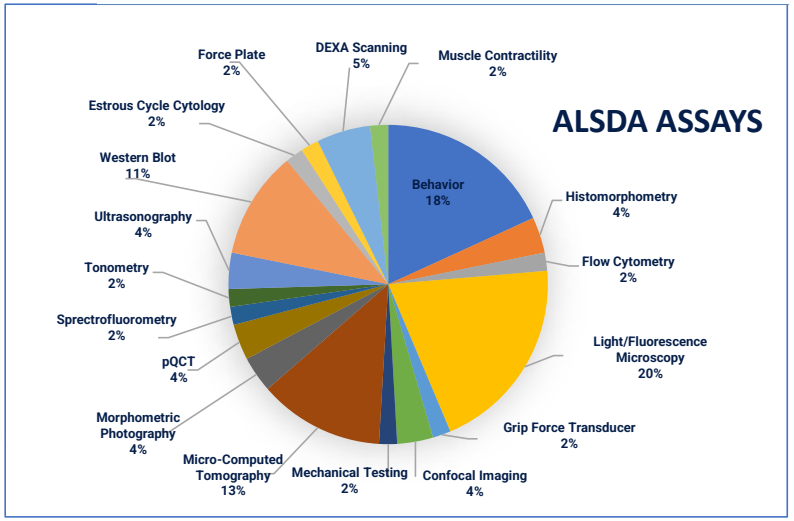








Civilian and Astronaut	Bed Rest, Spaceflight, Mars simulation
Cell Lines	Radiation (Ground), Simulated uG, Spaceflight, Parabolic Flight



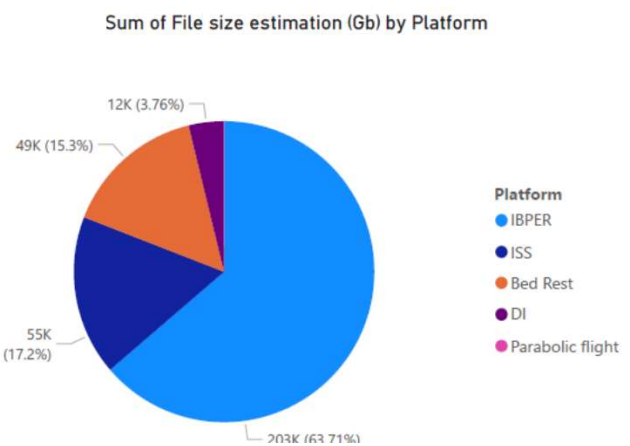
ESA Omics Data

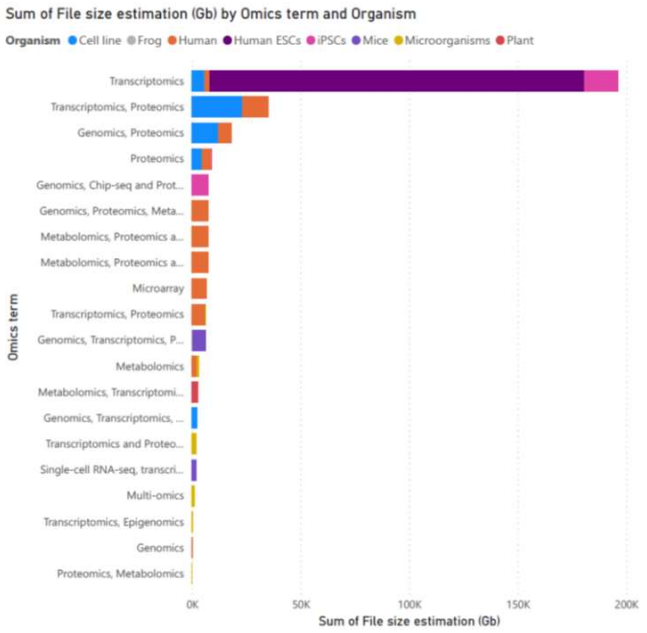
In the next 5 years, ESA estimates to collect 300 TB of omics data.

Majority of the omics projects are transcriptomics across several platforms using various organisms.

To utilize the current infrastructure, ESA SciSpacE team has requested all omics data be submitted and hosted on GeneLab (via OSDR).

Formal agreement for collaboration is currently in progress.

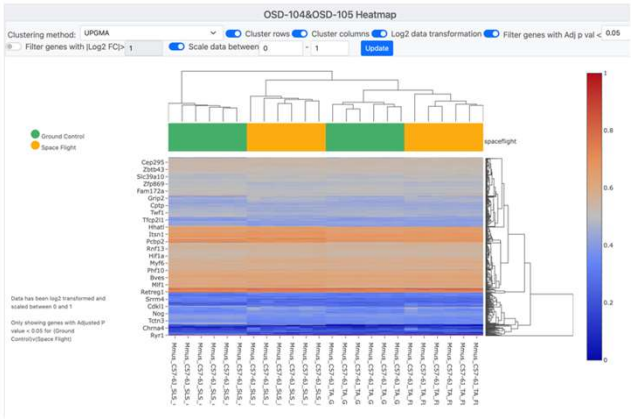


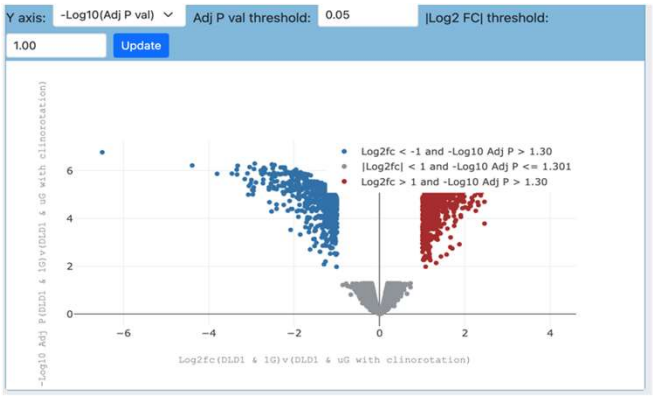


This initiative aims to centralize space omics data for the scientific community, facilitating cross-analysis and interoperability (FAIR), and leading to cost savings by avoiding infrastructure duplication.

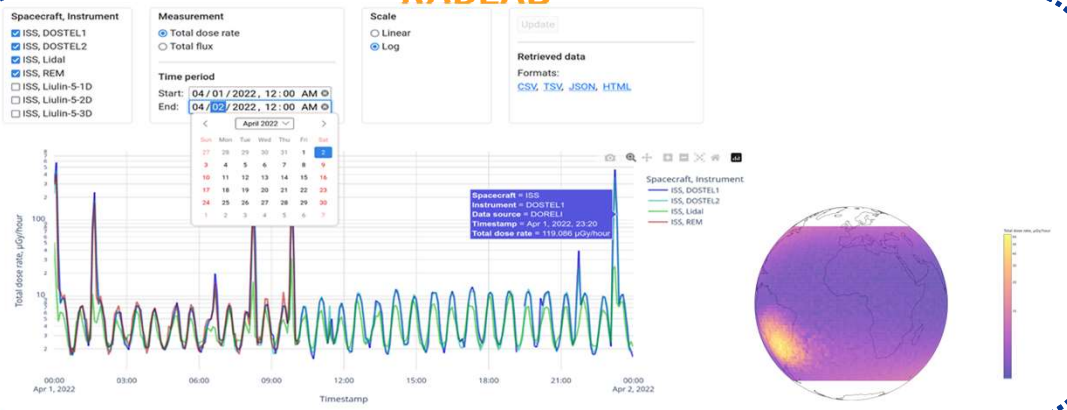
Interactive Visualization

MULTI-STUDY VISUALIZATION APP

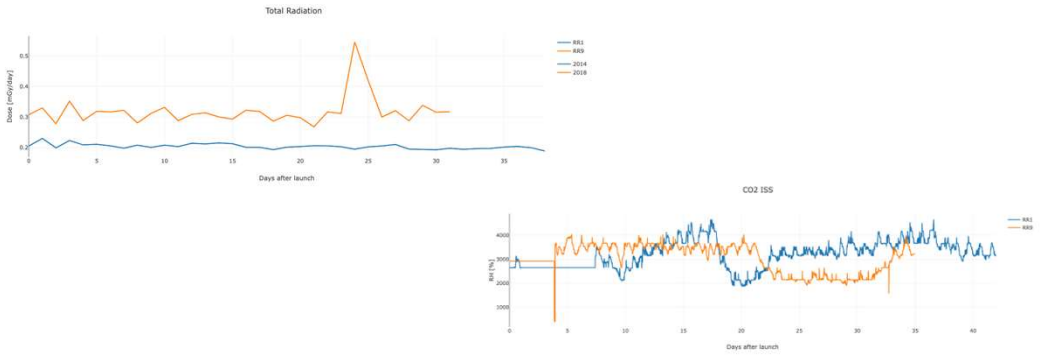




RADLAB



ENVIRONMENTAL DATA APP



GL4U – Enabling the next generation

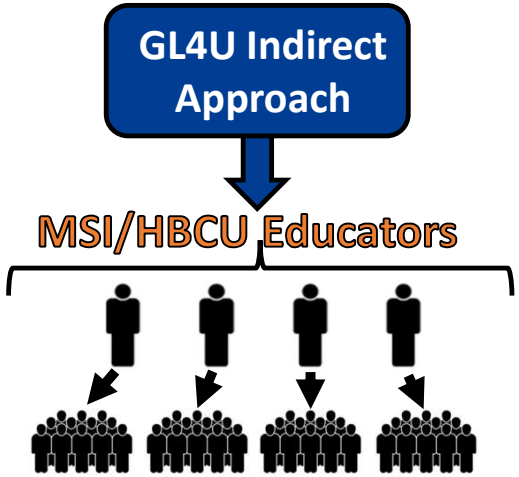
JPL Preparing for Research in Space Microbiology (PRISM) Workshop Series

2023 JPL PRISM Workshops:

- In-person/virtual 4-day workshop at 4 HBCUs/MSI
  - CalState Northridge – 25 students
  - Alabama A&M – 16 students
  - Howard University – 15 students
  - CalPoly Pomona – 20 students

- *Wetlab (1 day, JPL-led)*: Students extracted & quantified DNA from microbial isolates recovered from spacecraft-assembly clean rooms to be used for sequencing
- *Drylab (2 days, GeneLab-led)*: Students learned basic Unix coding and performed assembly, characterization, and annotation of these newly sequenced isolate genomes

GeneLab for Colleges and Universities (GL4U)



Direct GL4U – Student Bootcamps:

2021 – Intro & RNAseq Modules

- Virtual 1-week long bootcamp with 17 **SJSU** students
- Collaboration with USRA/SJSU; SJSU compute system
- Space biology- and RNAseq-specific lectures and hands-on instruction using Jupyter Notebooks (JNs)

2023 – Intro & Amplicon Seq Modules

- In-person 4-day bootcamp with 24 **CSULA** students
- NSF ACCESS compute resources
- Space biology- and Amplicon Seq-specific lectures/JNs

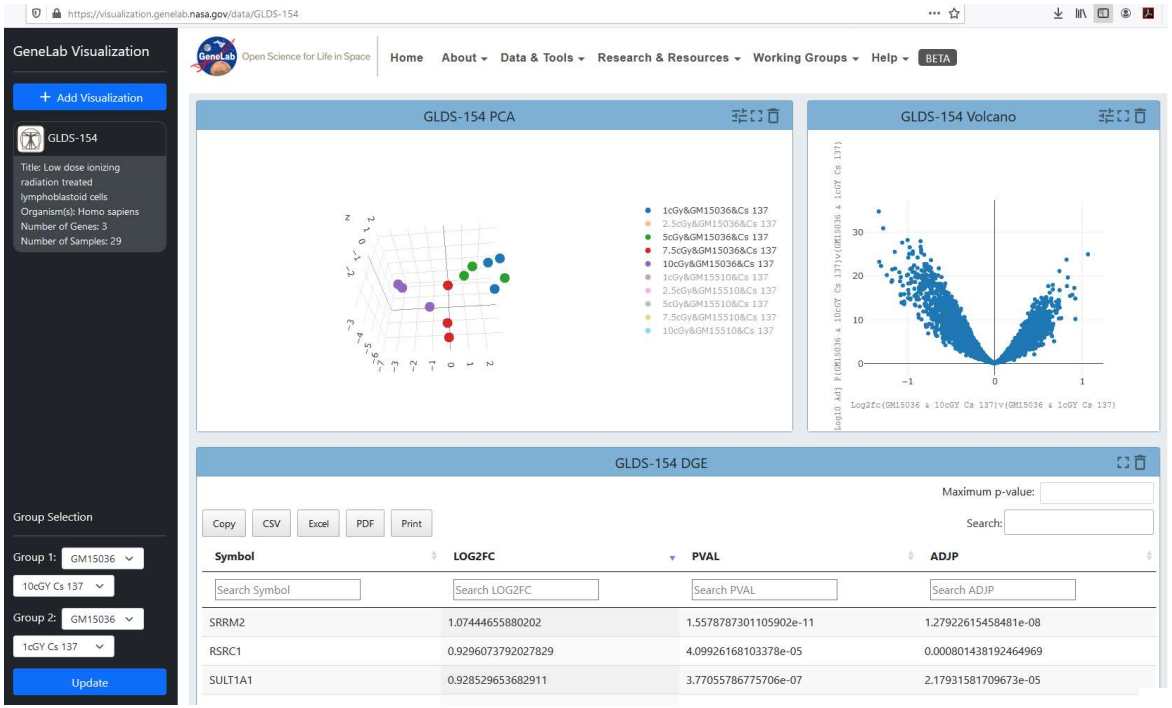
Indirect GL4U – Educator Bootcamps:

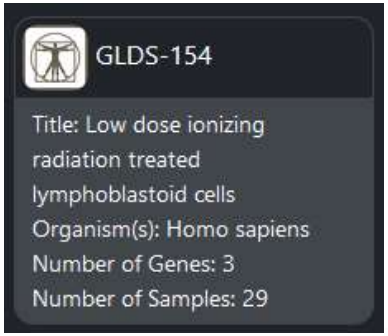
2022 – Intro & RNAseq Modules

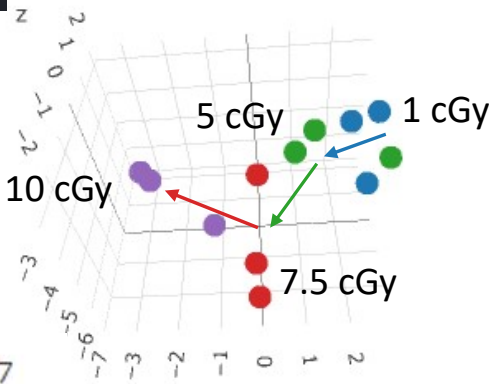
- Virtual ~1.5-week long bootcamp
- SMCE compute resources
- 6 professors and 4 graduate students from **4 HBCUs/MSIs**
- Space biology- and RNAseq-specific lectures and hands-on instruction using JNs
- Resources to teach at home institution
- Dr. Wei-Jen Lin taught content at CalPoly Pomona during their Fall 2022 semester

Multi-Study Visualization Platform

Low dose radiation transcriptomic data can be visualized





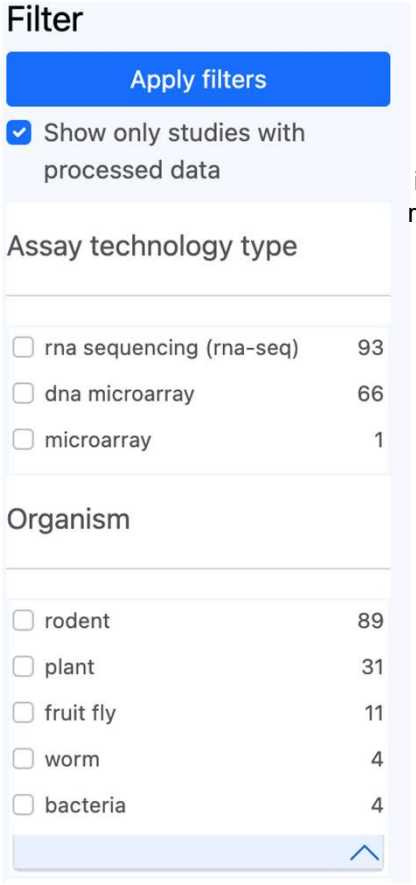


Application

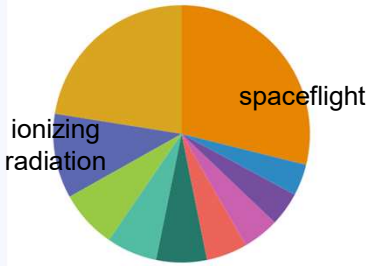
Visualization Platform

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

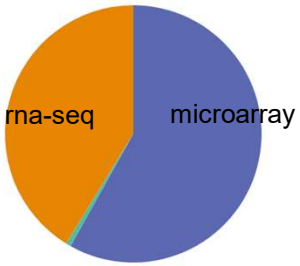
View and select RNA-seq studies



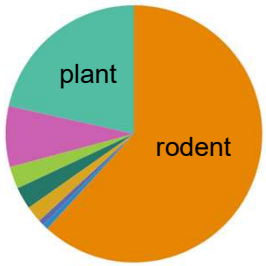
Factor



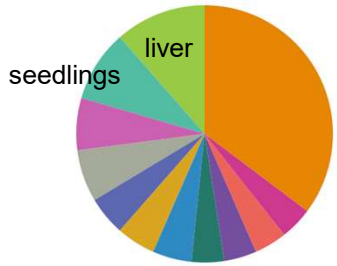
Assay technology type



Organism



Tissue



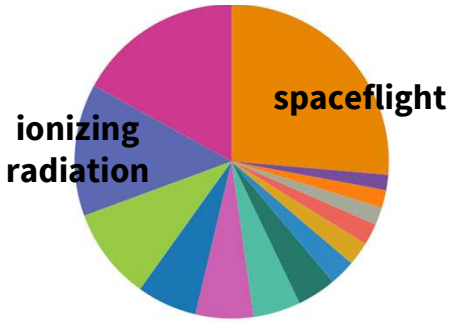
Copy CSV Excel PDF Print Search: RNA-seq

	OSD	Title	Assay	Organism	Tissue	Factor
☐	OSD-37	Comparison of the spaceflight transcriptome of four commonly used Arabidopsis thaliana ecotypes	RNA Sequencing (RNA-Seq)	Arabidopsis thaliana	Seedlings	ecotype,ecotype.term accession number,ecotype.term source ref,spaceflight,spaceflight.term accession number,spaceflight.term source ref,
☐	OSD-38	Proteomics and Transcriptomics analysis of Arabidopsis Seedlings in Microgravity	RNA Sequencing (RNA-Seq)	Arabidopsis thaliana	Seedling	sample preservation method,sample preservation method.term accession number,spaceflight,spaceflight.term accession number,spaceflight.term source ref,

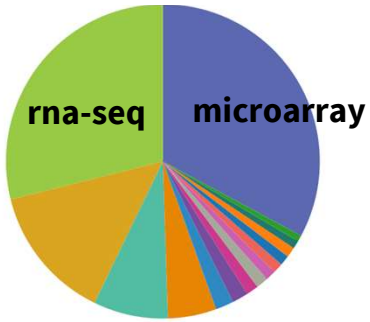
Visualization and Analysis Platform

1. Filter Data

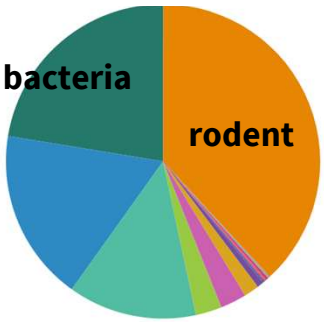
Experimental Factor



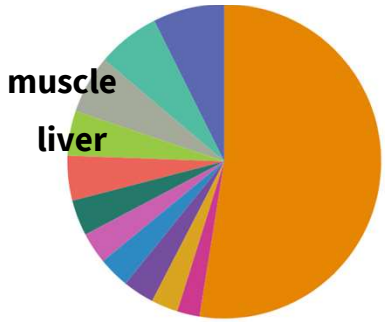
Experimental Assay



Organism



Tissue



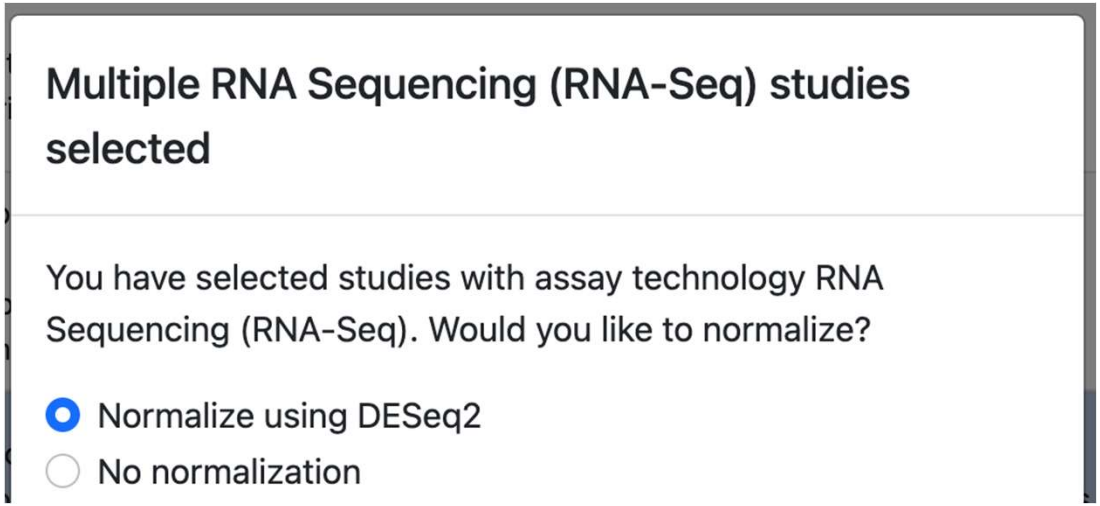
2. Select Datasets to Analyze

	OSD	Title	Assay	Organism	Tissue	Factor
☐	OSD-47	Rodent Research-1 (RR1) National Lab Validation Flight: Mouse liver transcriptomic, proteomic, epigenomic and histology data	RNA Sequencing (RNA-Seq)	Mus musculus	Liver	spaceflight,spaceflight.term accession number,spaceflight.term source ref,
☐	OSD-48	Rodent Research-1 (RR1) NASA Validation Flight: Mouse liver transcriptomic, proteomic, epigenomic and histology data	RNA Sequencing (RNA-Seq)	Mus musculus	Liver	dissection condition,spaceflight,spaceflight.term accession number,spaceflight.term source ref,

Visualize Study

Multi-Study Normalization

- Multiple RNA-seq studies can be *combined* and *normalized*
- DESeq2 “Median of Ratios” normalization
  - Ensures that all samples are directly comparable across experiments



Multi study PCA

OSD-37&OSD-38&OSD-120&OSD-217

Multi-Study PCA

Counts were normalized using DESeq2

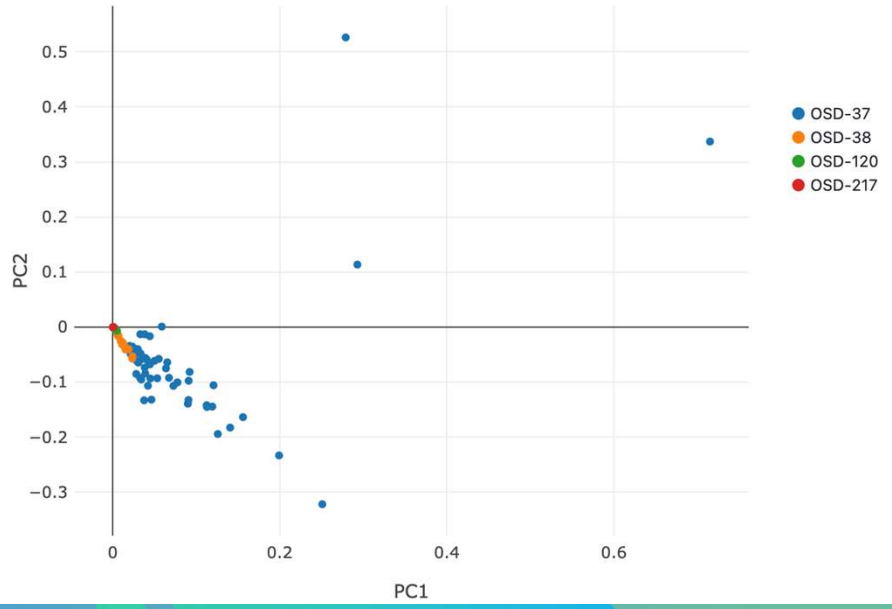
Change normalization method

Show Unnormalized Counts PCA

Normalized

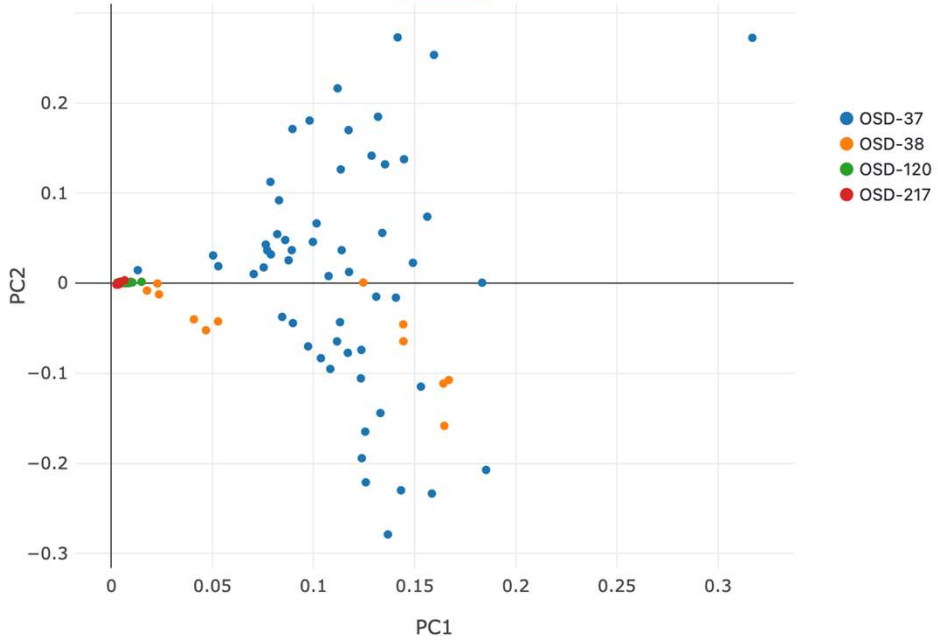
2D 3D X: PC1 Y: PC2 Color: dataset accession Shape: Size: Update

To add more options to Color, Shape, and Size, you can add more variables to the Factor table on the right.



Unnormalized

2D 3D X: PC1 Y: PC2 Z: PC3 Color: dataset accession Shape: Size: Update



Select new factors to visualize

Factor selection

Select variables and click 'Add' to add variables to the factors table. The variables selected will be as used as factors for Differential Gene Expression Analysis. If no variables are added as factors, OSD will be used as a factor.

Adding a variable to the table will automatically add it as a color option on the PCA.

Factors | sample preservation meth | Add

OSD	ecotype	spaceflight
OSD-37	Col-0,Cvi-0,Ler-0,Ws-2	Ground Control,Space Flight
OSD-38		Ground Control,Space Flight
OSD-120	Col-0,Col-0 PhyD,Wassilewskija ecotype	Ground Control,Space Flight
OSD-217	Wassilewskija ecotype	Ground Control,Space Flight

Normalize using only selected samples and/or factors

Expand table

The expanded table will allow you to select the samples you want to use for the Differential Gene Expression Analysis. All samples are selected by default.

Select new factors to visualize

Factor selection

Select variables and click 'Add' to add variables to the factors table. The variables selected will be as used as factors for Differential Gene Expression Analysis. If no variables are added as factors, OSD will be used as a factor.

Adding a variable to the table will automatically add it as a color option on the PCA.

Factors | sample preservation meth | Add

OSD	ecotype	spaceflight
OSD-37	Col-0,Cvi-0,Ler-0,Ws-2	Ground Control,Space Flight
OSD-38		Ground Control,Space Flight
OSD-120	Col-0,Col-0 PhyD,Wassilewskija ecotype	Ground Control,Space Flight
OSD-217	Wassilewskija ecotype	Ground Control,Space Flight

Normalize using only selected samples and/or factors

Expand table

The expanded table will allow you to select the samples you want to use for the Differential Gene Expression Analysis. All samples are selected by default.

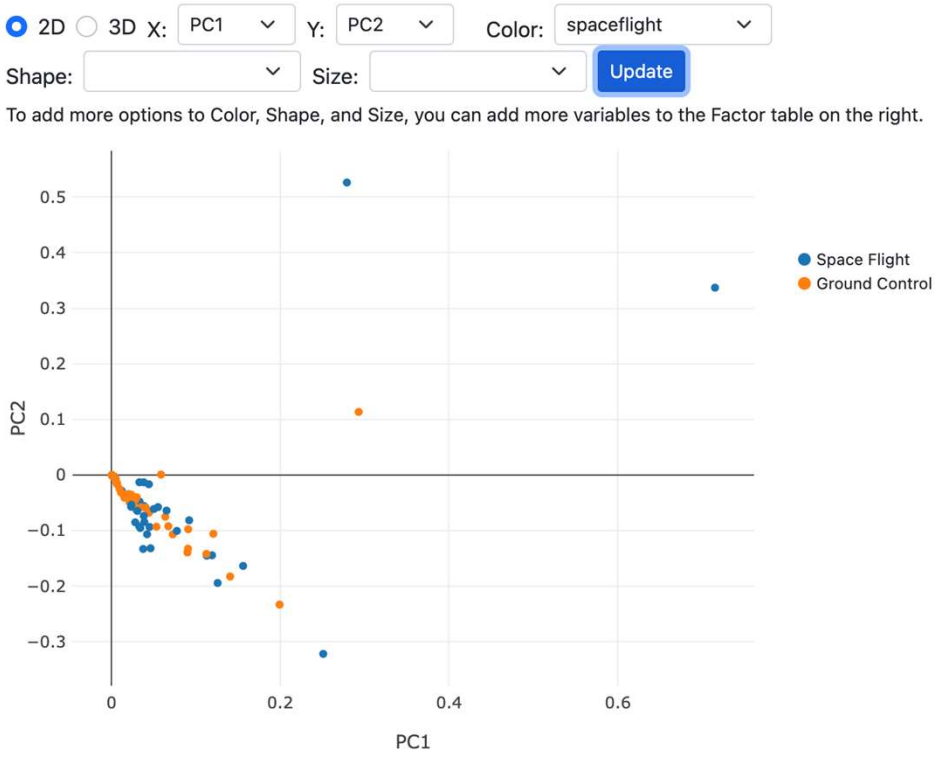
Hide expanded table

The expanded table will allow you to select the samples you want to use for the Differential Gene Expression Analysis. All samples are selected by default.

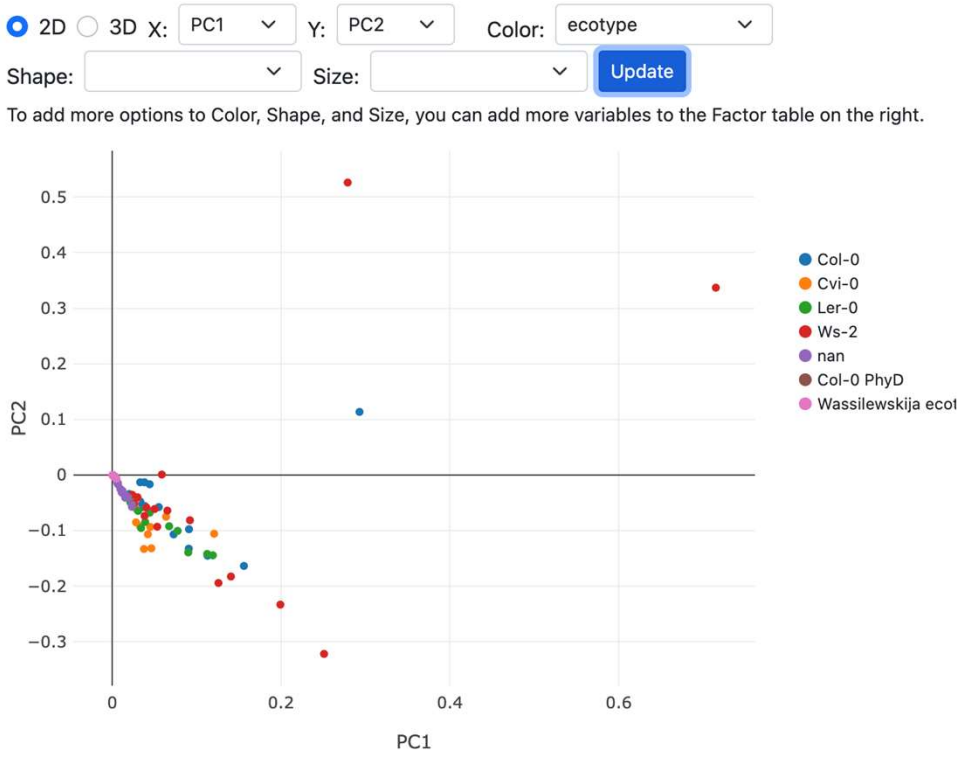
OSD	Sample	ecotype	spaceflight
Searc	Search	Search	Search
OSD-37	Atha_Col-0_sl-pool_FLT_Rep1_R1-FL-A1	Col-0	Space Flight
OSD-37	Atha_Col-0_sl-pool_FLT_Rep2_R1-FL-A4	Col-0	Space Flight
OSD-37	Atha_Col-0_sl-pool_FLT_Rep3_R1-FL-B1	Col-0	Space Flight
OSD-37	Atha_Col-0_sl-pool_FLT_Rep4_R1-FL-B4	Col-0	Space Flight
OSD-37	Atha_Col-0_sl-pool_FLT_Rep5_R2-FL-A1	Col-0	Space Flight

PCA with new factors

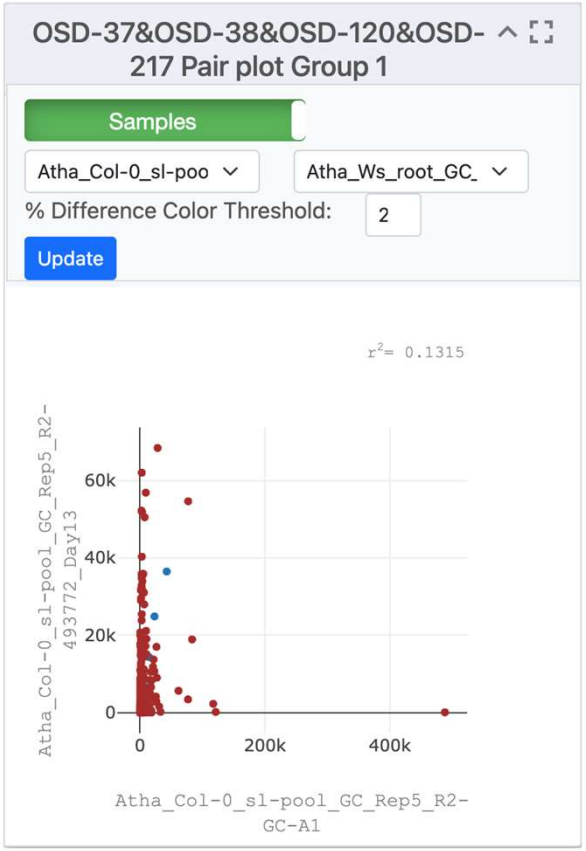
Spaceflight

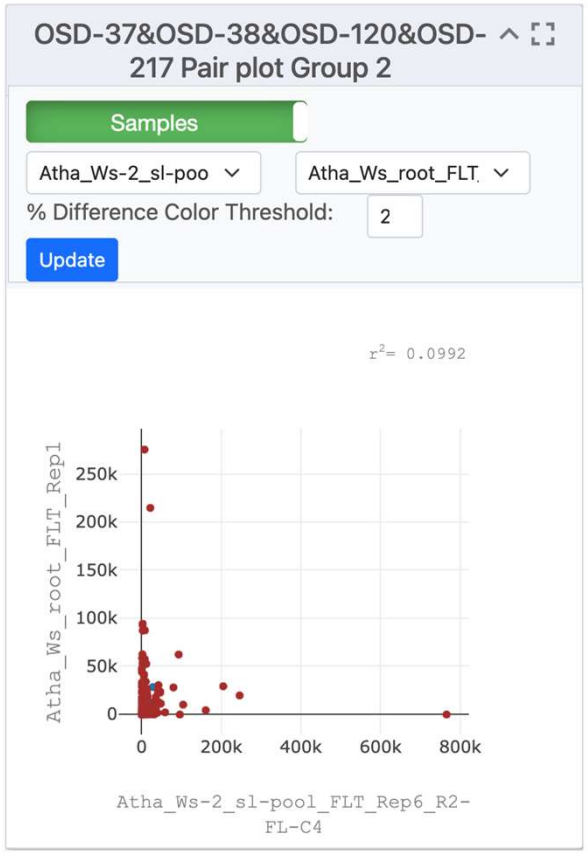


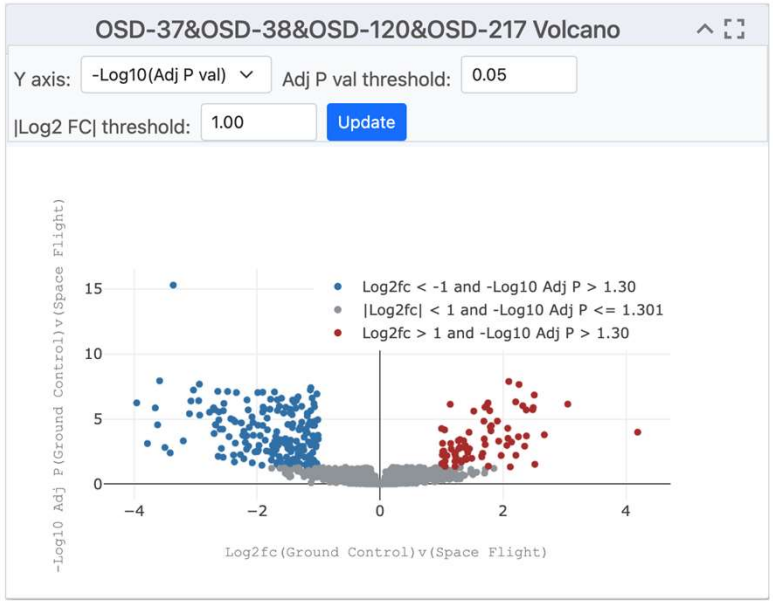
Ecotype



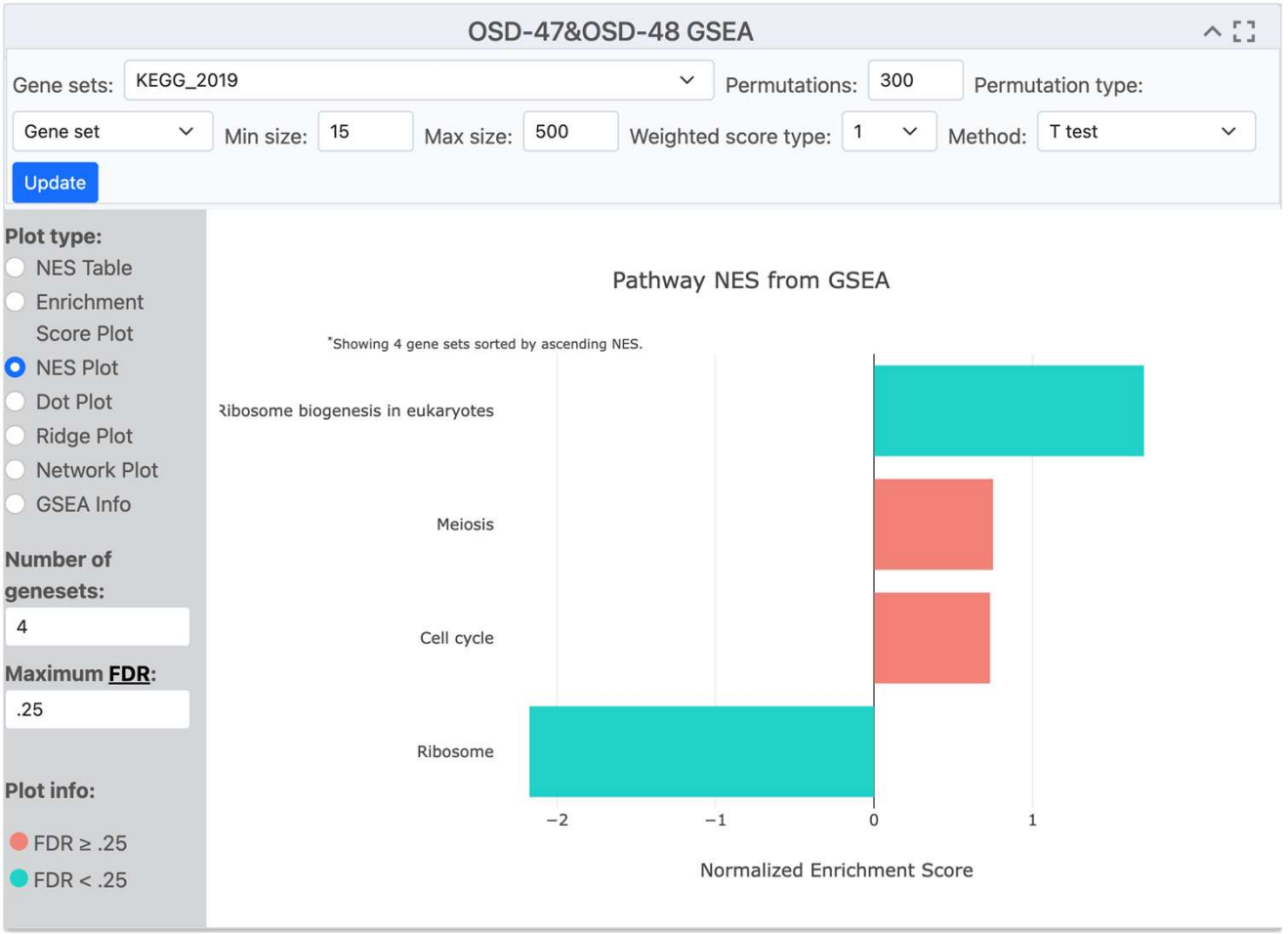
Pair and volcano plots



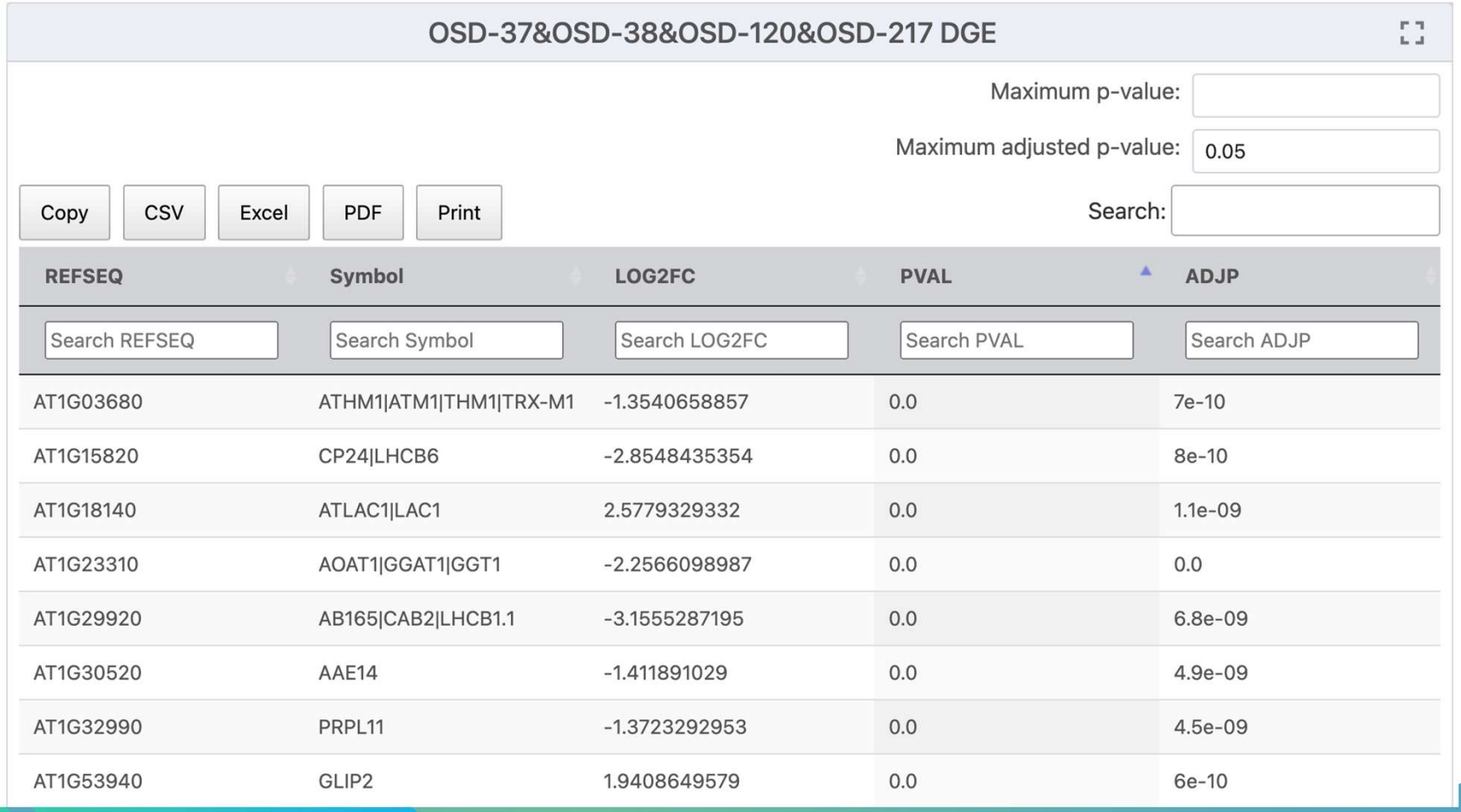




Gene Set Enrichment Analysis



View and download the DGE table



OSD-37&OSD-38&OSD-120&OSD-217 DGE

Maximum p-value:

Maximum adjusted p-value: 0.05

Copy | CSV | Excel | PDF | Print

Search:

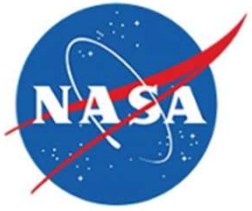
REFSEQ	Symbol	LOG2FC	PVAL	ADJP
Search REFSEQ	Search Symbol	Search LOG2FC	Search PVAL	Search ADJP
AT1G03680	ATHM1|ATM1|THM1|TRX-M1	-1.3540658857	0.0	7e-10
AT1G15820	CP24|LHCB6	-2.8548435354	0.0	8e-10
AT1G18140	ATLAC1|LAC1	2.5779329332	0.0	1.1e-09
AT1G23310	AOAT1|GGAT1|GGT1	-2.2566098987	0.0	0.0
AT1G29920	AB165|CAB2|LHCB1.1	-3.1555287195	0.0	6.8e-09
AT1G30520	AAE14	-1.411891029	0.0	4.9e-09
AT1G32990	PRPL11	-1.3723292953	0.0	4.5e-09
AT1G53940	GLIP2	1.9408649579	0.0	6e-10

OSDR Acknowledgements

- Ana Uriarte Acuna
- Kirill Grigorev
- Jamie Bales
- Lauren Sanders
- Amanda Saravia-Butler
- Ryan Scott
- Samrawit Gebre
- Danielle Lopez

Compute

NASA Science Managed Compute Environment



Support

- NASA Space Biology Program
- NASA Human Research Program