

Introduction to NASA Open Science and GeneLab

National Aeronautics and
Space Administration



GL4U: Introduction 2024
Lecture 2 of 4

Amanda M. Saravia-Butler, Ph.D.
NASA GeneLab Science Lead
Contractor: KBR

Biological & Physical Sciences

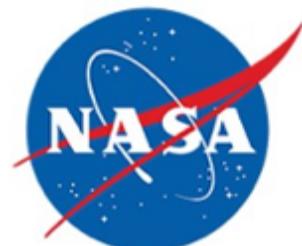


Open Science for Life in Space Teams



Open Science Analysis Working Group Members

Support

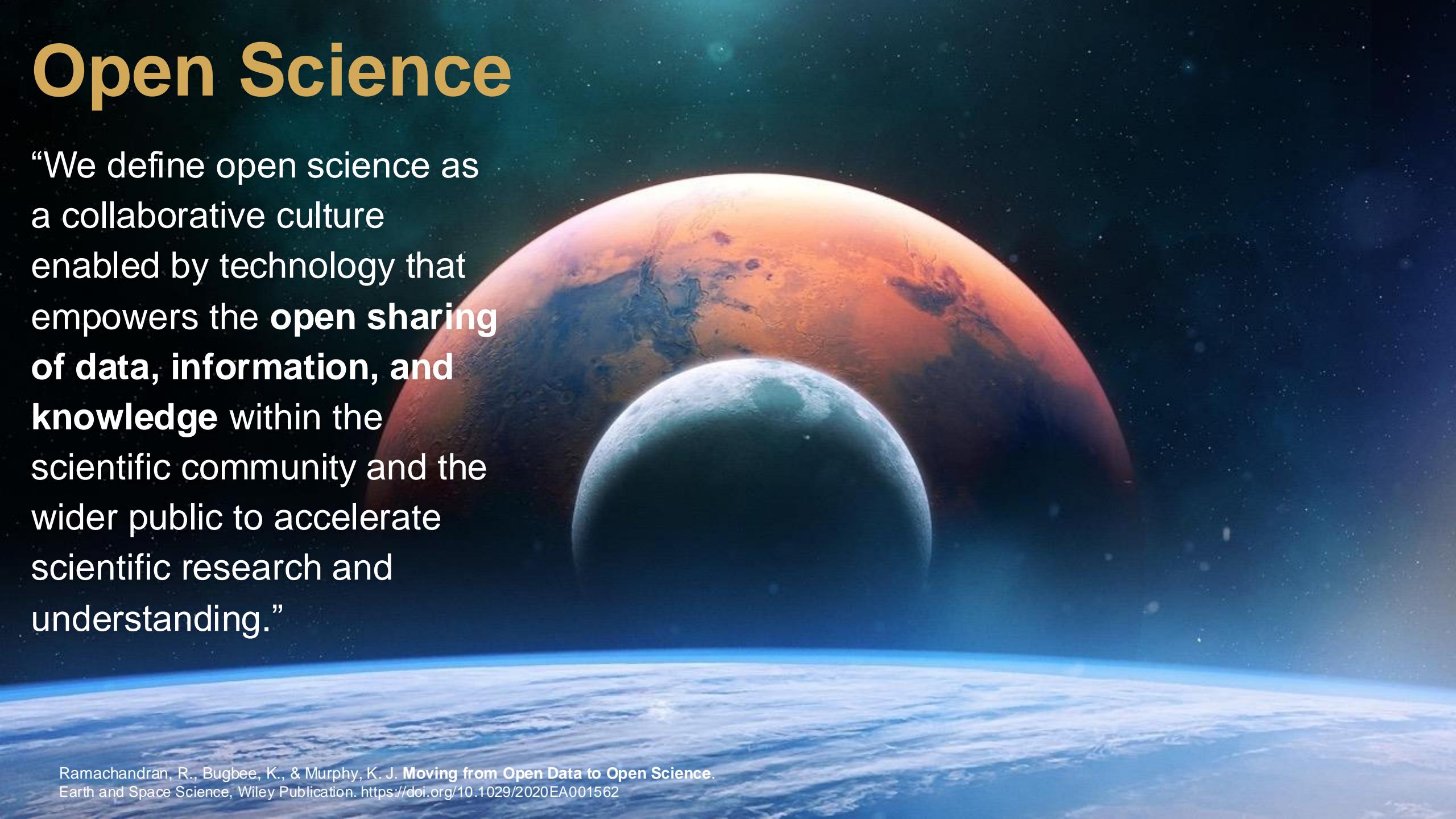


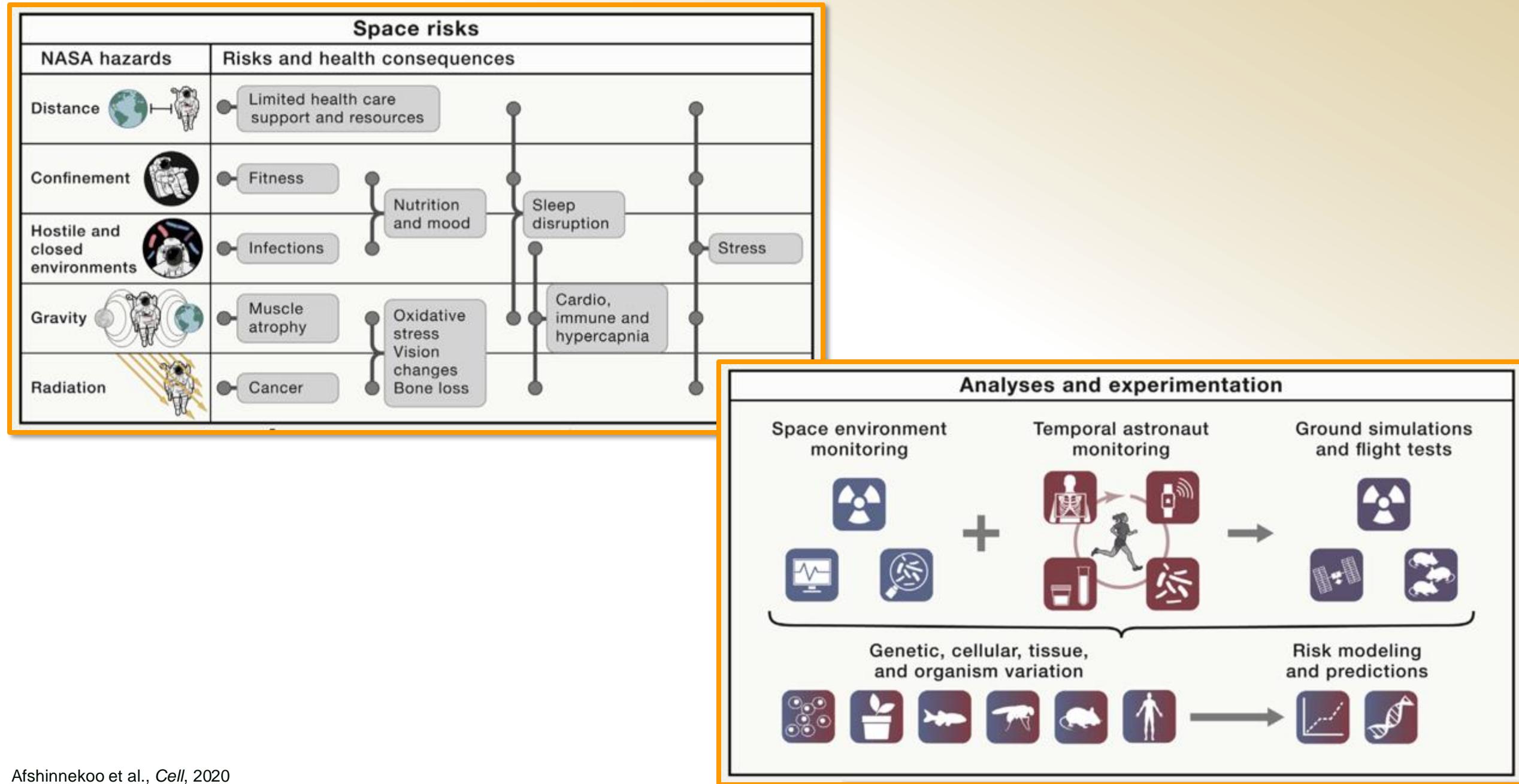
NASA Space Biology Program
NASA Science Mission Directorate
NASA Human Research Program
NASA Biological and Physical Sciences

NASA Open Science Overview

Open Science

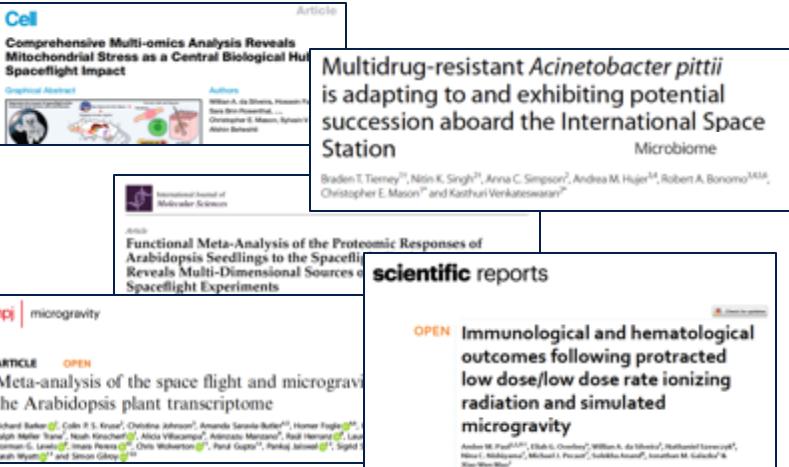
“We define open science as a collaborative culture enabled by technology that empowers the **open sharing of data, information, and knowledge** within the scientific community and the wider public to accelerate scientific research and understanding.”





Reusable Data is Essential for Life and Humans to Thrive in Deep Space

Knowledge Discovery

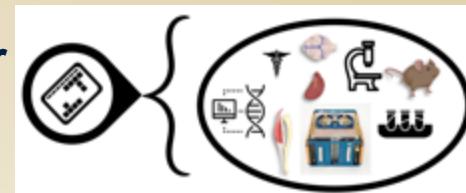


Advanced Computational Methods

Model development
Active and adaptive learning
Causal and network inference
Explainable AI
Federated learning
Few- or one-shot learning
Generative learning (synthetic datasets)
Knowledge graphs
Lifelong learning (living lab)
Out-of-distribution learning
Transfer learning



Telemetry for Biomedical Context



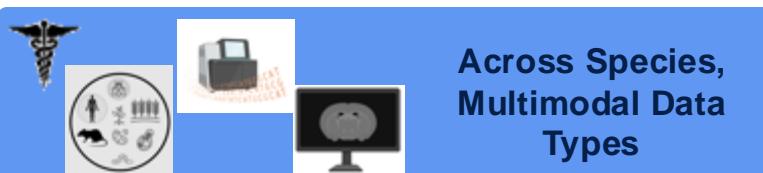
Analysis Across Missions



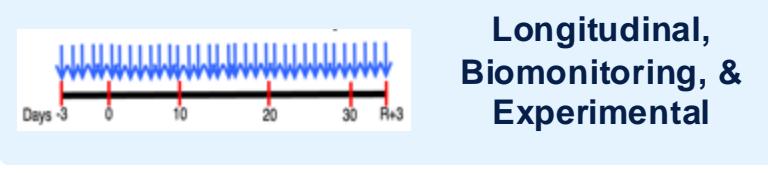
Across Hazards



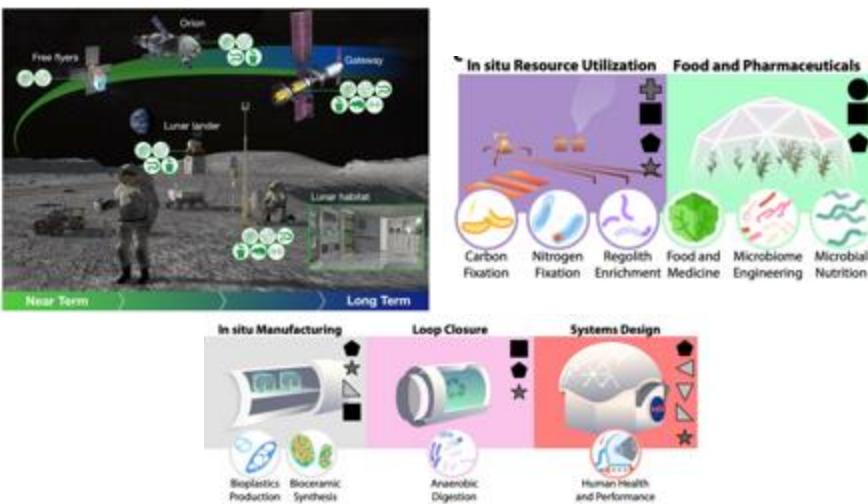
Across Species, Multimodal Data Types



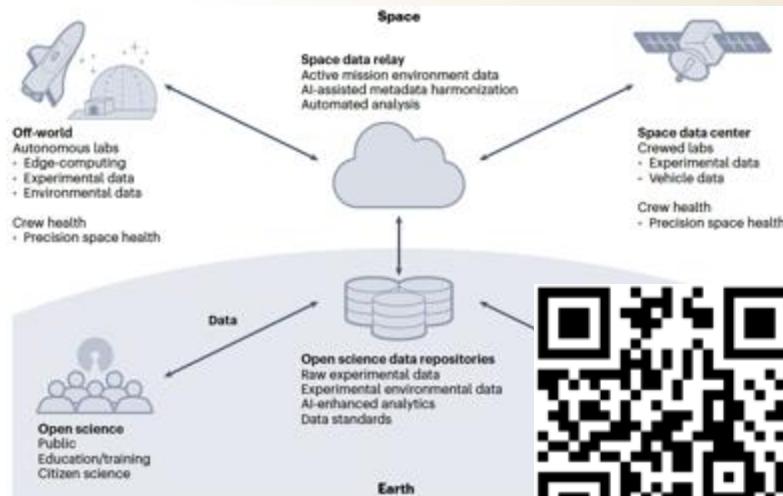
Longitudinal, Biomonitoring, & Experimental



Novel Ecosystem, Platform, Countermeasure Development, Health Risk Quantification



Involve the World in Spaceflight



Biospecimen Sharing Program (BSP)



Dissection and preservation of rodent tissues from Flight and Ground investigations

Coordination of internal tissue sharing



NASA Biological Institutional Scientific Collection (NBISC)



Collection of non-human specimens and space microbial culture



NASA GeneLab (GL)



Collection and curation of omics data



Ames Life Sciences Data Archive (ALSDA)



Collection and curation of mission, project, and imaging data



NASA Internal Program

Open Source Science Programs – Available Globally

Physiological/Phenotypic/Imaging/ Environmental Telemetry Data



Tabular, text,
imaging,
video, code



Molecular/Omics Data

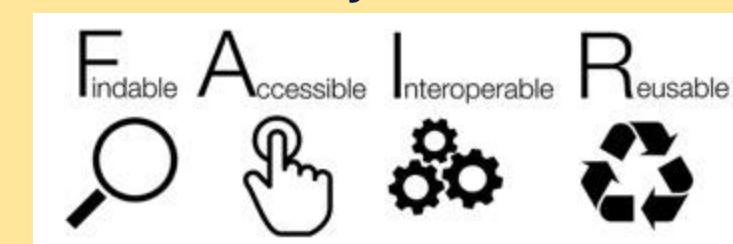


Biospecimens



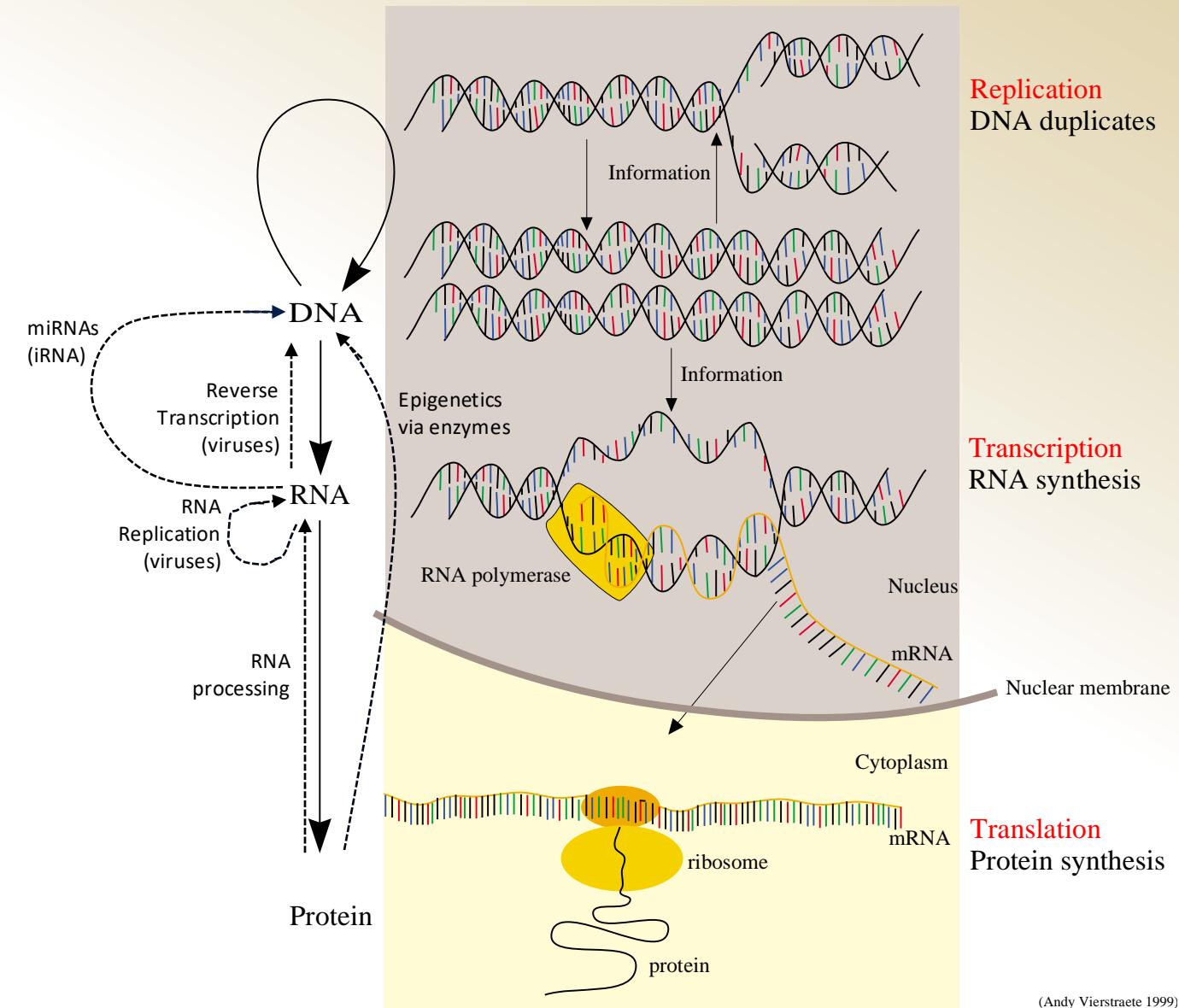
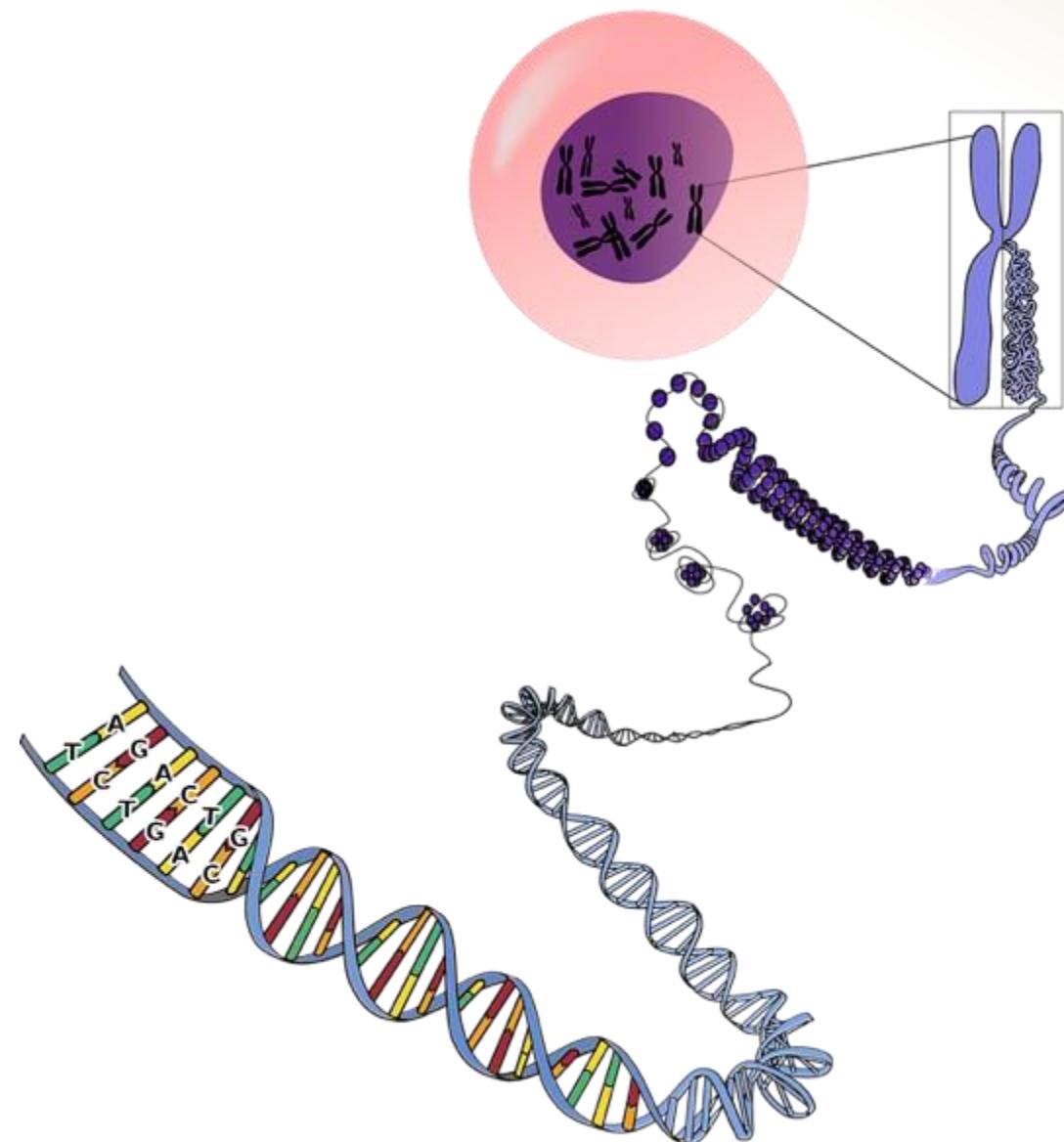
NASA Open Science
Data Repository (OSDR)
osdr.nasa.gov/bio

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



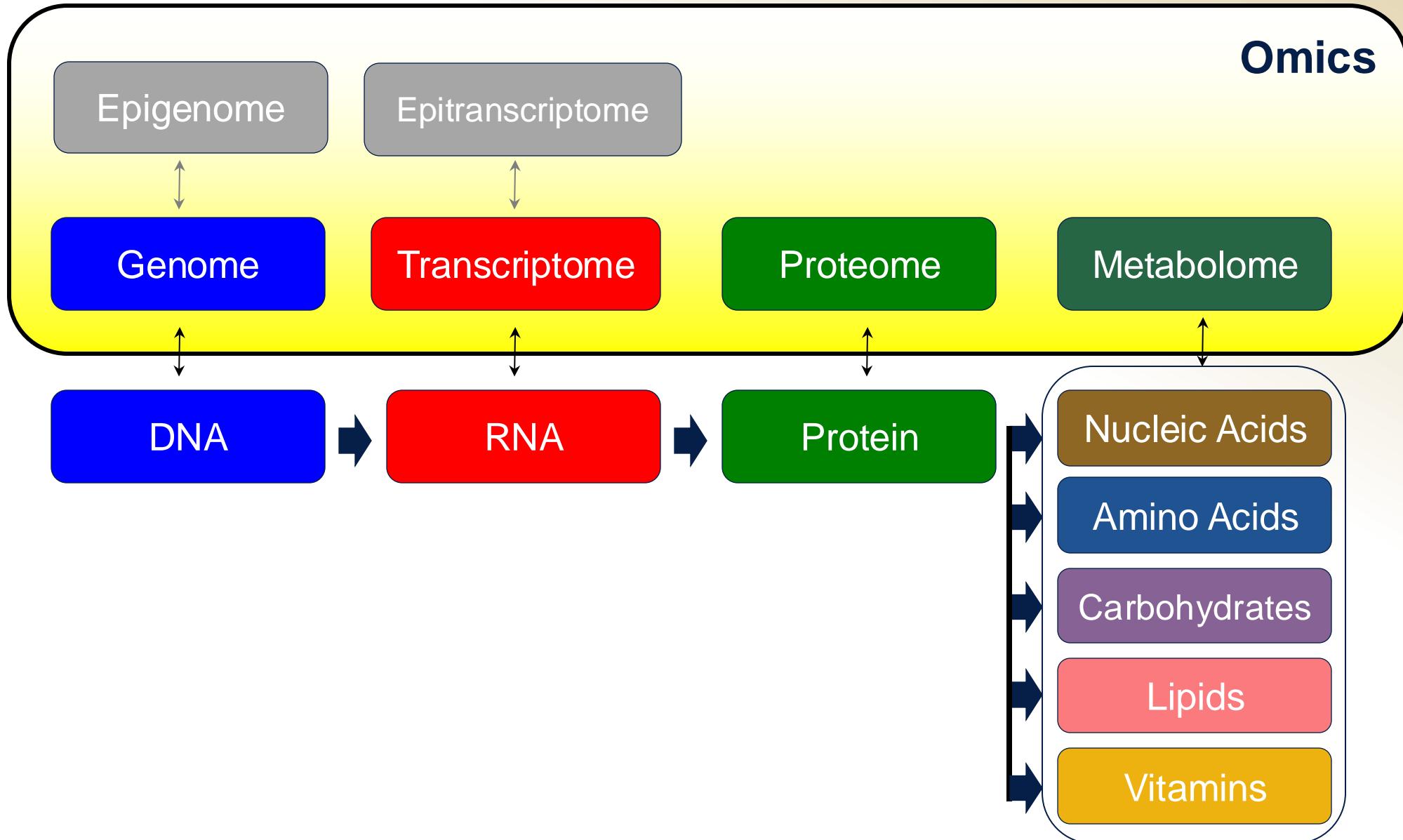
What are 'omics'?

The Central Dogma of Molecular Biology



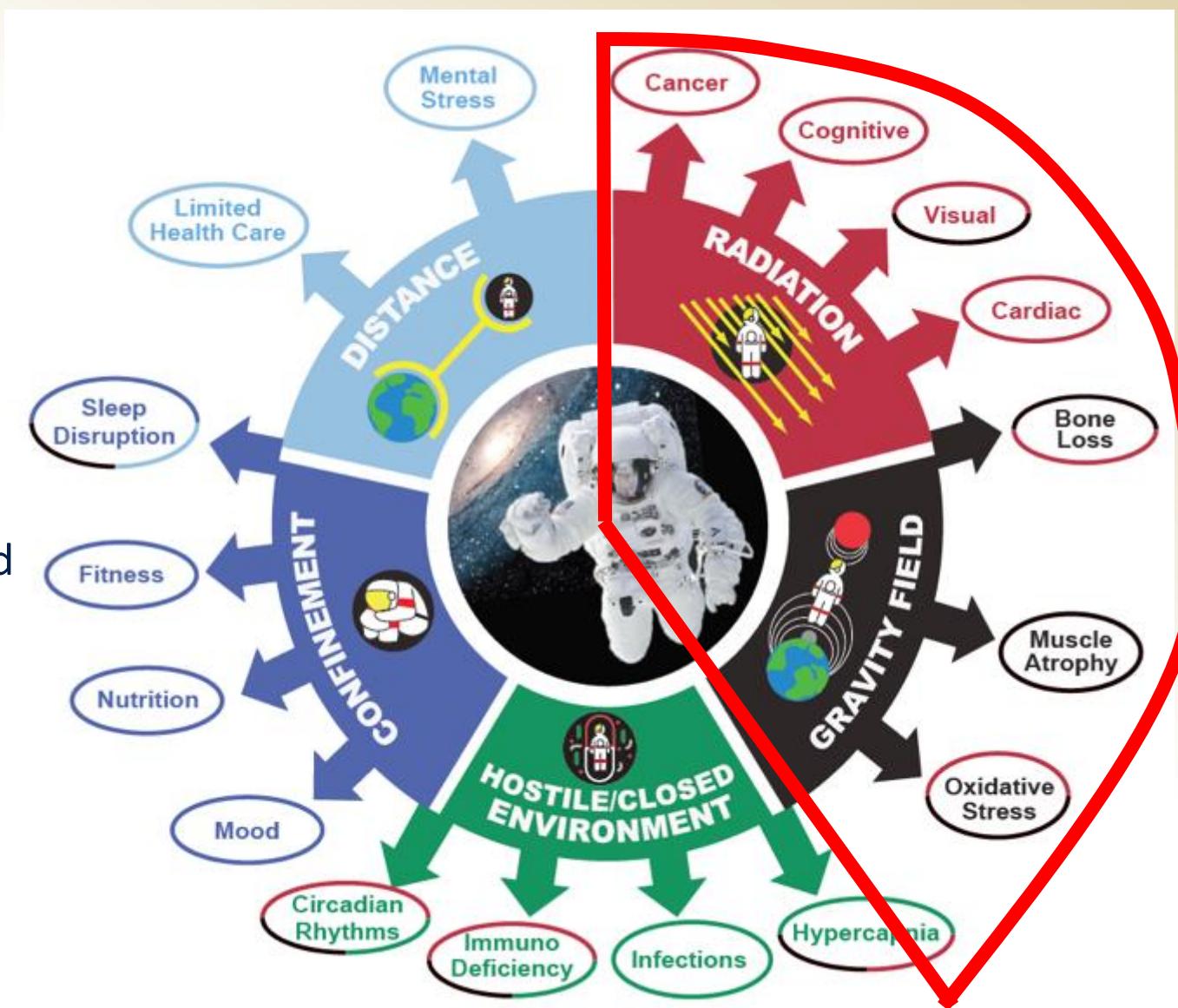
(Andy Vierstraete 1999)

What are ‘omics?



- What, when, and where genes are expressed allow for cell type diversity and enable living organisms to respond and adapt to surroundings
- Gene expression is primarily regulated by environmental factors both micro (cell's micro-environment) and macro (organism's external stimuli or stressors)
- Spaceflight alters the transcriptional patterns and molecular signaling networks within our cells, which in turn causes physiological changes
- Understanding such changes will enable development of mitigation strategies to better withstand the rigors of long-duration spaceflight

Primary Stressors of Spaceflight



Physiological/Phenotypic/Imaging/ Environmental Telemetry Data

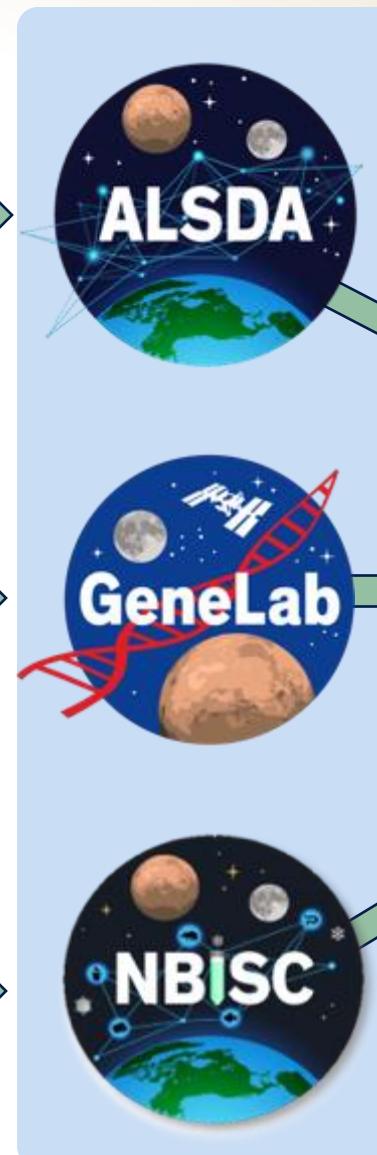


Tabular, text,
imaging,
video, code

Molecular/Omics Data

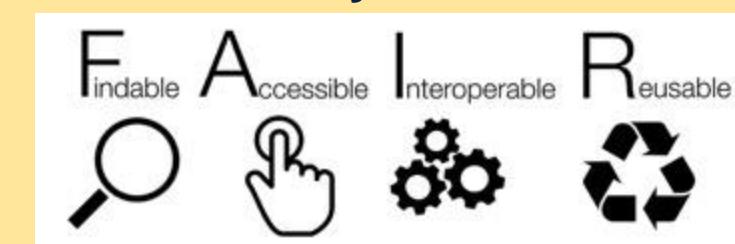


Biospecimens



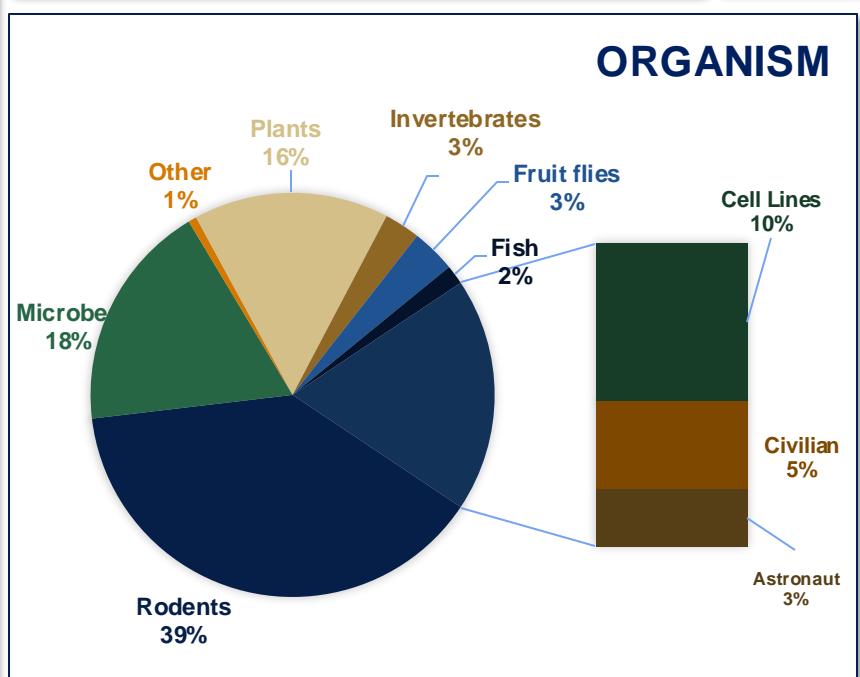
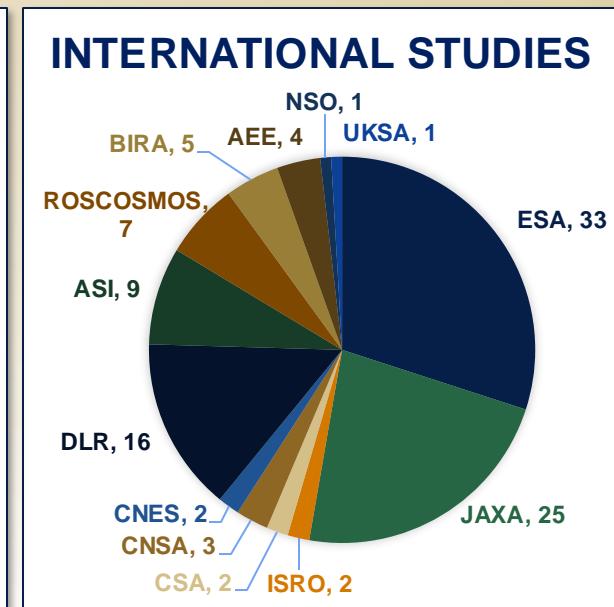
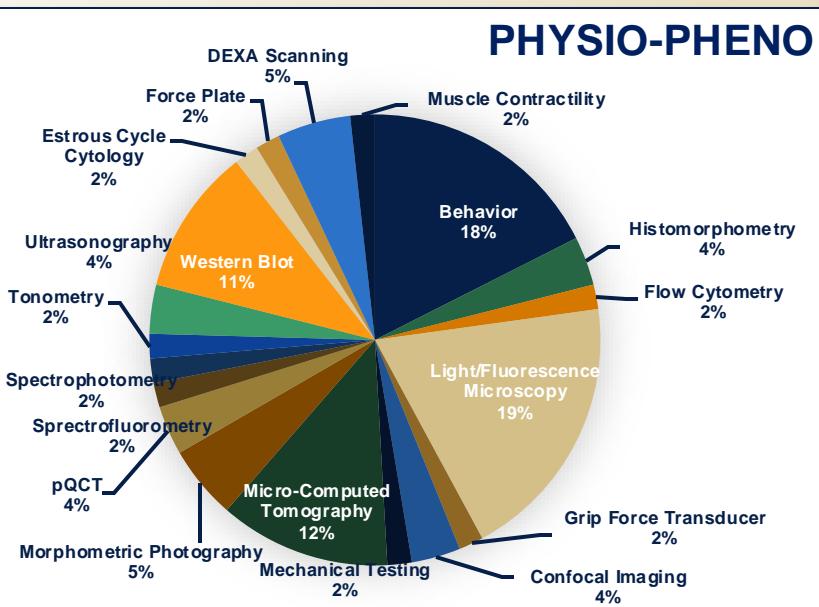
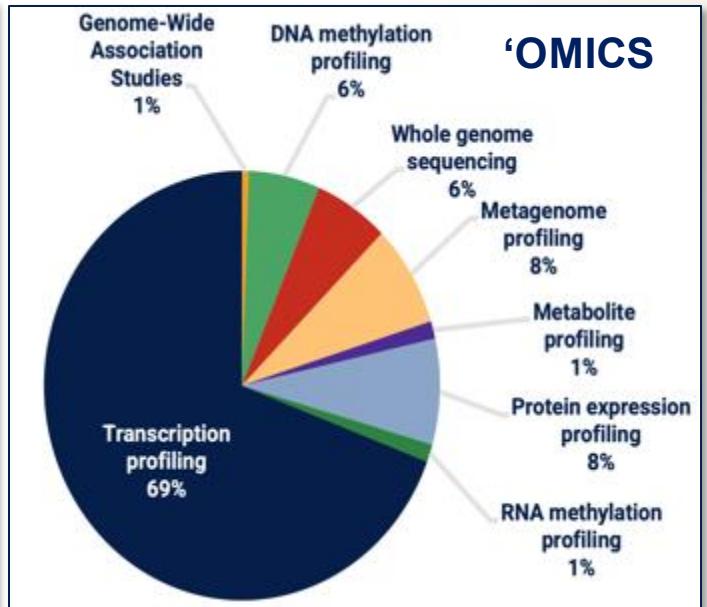
NASA Open Science
Data Repository (OSDR)
osdr.nasa.gov/bio

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



NASA OSDR – By The Numbers

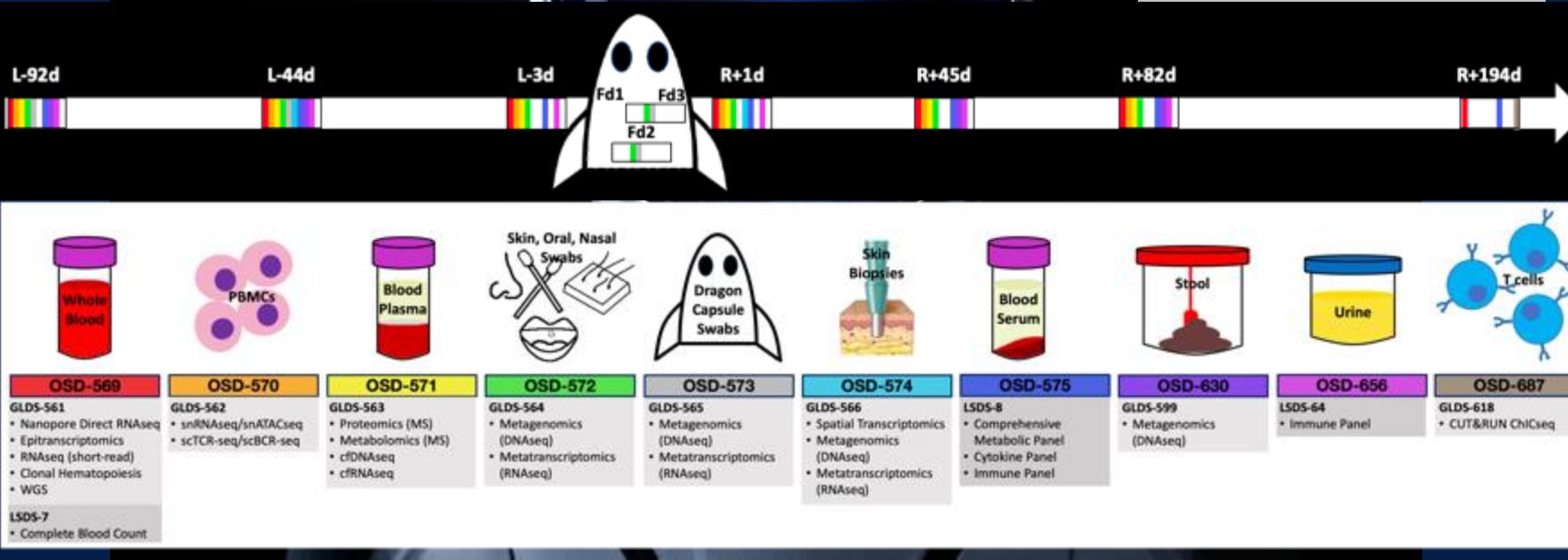
505 Studies
969 Datasets
45 Species
>80 Assays
>160TB Data



85 Enabled Publications linked to OSDR
109 Original Publications linked to OSDR
150+ Datasets used in enabled publications

INSPIRATION 4

- Nature Portfolio Collection (2024): Space Omics and Medical Atlas (SOMA) across orbits
- I4 Image Credit: Chris Mason's Lab



- Commercial Sub-Orbital Galactic-05 – Flew Nov. 2

- Astroskin: ECG, systolic BP, heart rate (HR), breathing rate (BR), skin temperature, activity levels (provided by 3-axis accelerometer), and oxygen saturation in blood (SpO2)
- Blood Glucose Monitoring Device

Soon!

- Credit: Dr. Shawna Pandya





Data Access Info Available On DAR Portal

The screenshot shows the DAR Portal interface. At the top, there's a 'Data Access Requests' section for 'DAR-63: Study: OSD-571 Version 1' with a status of 'Approved' and expiration on '1/30/2025'. A large blue arrow points from this section to the 'Files' section below. The 'Files' section shows a 'Study Files' section with a 'Download' button highlighted in red. Below it is a 'Search Files' section. Under 'OSD-571', there are 'Study Metadata Files', 'RNA-Seq' (with a checked checkbox), and two files: 'Raw sequence data' (843.07 MB, Wed Oct 18 2023) and 'GLDS-563_rna-seq_cfRNA_I4_C004_R+45_rep2_R2_raw.fastq.gz' (782.29 MB, Wed Oct 18 2023).

Access Request Process for Private Data

1	2	3	4	5	6	7
Pre-Submission	Investigator Submits	Signing Official Reviews	Data Access Committee Reviews	Access Data	Renew Project (Optional)	Close Out Project
Request Button clicked in OSD dataset	OSDR Data Request Package sent w/ Data Use Agreement	OSDR Follow-Up with your institution	Review Board with Specific Stakeholders (OSDR, PIs, etc)	Approved / Denied Secure Access	Approved / Denied	Access removed at end of IRB, upon participant request, or if conditions of Data Use Agreement are violated

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

Navigating OSDR

The NASA Open Science Data Repository (OSDR) enables access to space-related data from experiments and missions that investigate biological and health responses of terrestrial life to spaceflight. The goal of OSDR is to enable multi-modal and multi-hierarchical fundamental space life science data be reused toward basic science, applied science, and operational outcomes for space exploration and knowledge discovery. These data include 'omics, phenotypic, physiological, behavioral, hardware, environmental telemetry; raw, processed; tabular, text, code, bioimaging, and video.

Learn More About OSDR

Explore and Contribute



Open Science Data Repository

Search and upload spaceflight datasets

Explore the Data Repository



Submit Data to OSDR

Have space-relevant data to submit?

Contribute via the Submission Portal



Join OSDR AWG's

We look forward to collaborating with you!

Interested in participating in an AWG?

The screenshot shows the BDME dashboard with a green '+ Create New' button highlighted by a red box. Below it, there are tabs for All, RDSA, Studies, and Experiments, and filters for All, Submitted, In Curation, In Review, Req Signature, Finalized, Public, and Private.

- PIs can now submit data through BDME.
- Allows efficient input of sample and assay level metadata alongside data files.
- Tailored to specialized needs of spaceflight samples.
- Used for both omics (GeneLab) and non-omics (ALSDA) data submissions.

Sample Name	Organism	Strain	Material Type	Protocol REF	Parameter Value		Ionizing Radiation	Time	Unit	Dose	Unit
					- Storage Temperature Setting	Unit					
Sample1	Mus musculus	X-5	liver	Sample collection	-80	degree Celsius	Silicon	1	hour	1	gray
Sample2	Mus musculus	X-5	liver	Sample collection	-80	degree Celsius	Sham-irradiated	0	hour	0	gray

Assay level data with assay-specific fields

Sample Name	Protocol REF	Parameter Value - QA Instrument	Parameter Value - QA Assay	Parameter Value - QA Score	Unit	Extract Name	Protocol REF
Sample1	Nucleic Acid Extraction	BioAnalyzer	NanoChip	8	RIN	Extract 1	Library Construction
Sample2	Nucleic Acid Extraction	BioAnalyzer	NanoChip	9	RIN	Extract 2	Library Construction
Sample3	Nucleic Acid Extraction	BioAnalyzer	NanoChip	9.1	RIN	Extract 3	Library Construction

A tooltip for the unit 'gray' is shown, providing its definition and source from the OBO library.

The screenshot shows a study preview with a private link for sharing.

DESCRIPTION		
Study Description: Study to understand the effects of gamma-rays and HZE particles on mouse tissues. This experiment was conducted at BMF on 25 mouse tissues. Radiation types studied included - Fe, Si, and X-ray at different doses and exposure times. The duration and dose for experiments varied: 0, 1, 3, Gy for 4, 6, 12, 24 hours. Total RNA was extracted and sequencing using the NextSeq 6000.		
CONTACTS		
NAME: Sam Gobin ROLE: Submitter ORGANIZATION: NASA E-MAIL: sam.gobin@nasa.gov		
STUDY DESIGN FACTORS	FACTOR	ONTOLOGY: CONCEPT
	Ionizing Radiation	Ionizing Radiation
	Time	Time
	Dose	Absorbed Radiation Dose

PIs are presented with a study preview and a private link for sharing with reviewers and collaborators.

The NASA Open Science Data Repository (OSDR) enables access to space-related data from experiments and missions that investigate biological and health responses of terrestrial life to spaceflight. The goal of OSDR is to enable multi-modal and multi-hierarchical fundamental space life science data be reused toward basic science, applied science, and operational outcomes for space exploration and knowledge discovery. These data include 'omics, phenotypic, physiological, behavioral, hardware, environmental telemetry; raw, processed; tabular, text, code, bioimaging, and video.

Learn More About OSDR

Explore and Contribute



Open Science Data Repository

Search and upload spaceflight datasets

[Explore the Data Repository](#)



Submit Data to OSDR

Have space-relevant data to submit?

[Contribute via the Submission Portal](#)



Join OSDR AWG's

We look forward to collaborating with you!

[Interested in participating in an AWG?](#)

General Search Filters

Data Source

- GeneLab
- ALSDA
- NIH GEO
- EBI PRIDE
- ANL MG-RAST

Data Type

- Study
- Experiment
- Subject
- Biospecimen
- Payload

Show more ▾

Study Search Filters

Project Type

- Ground
- Spaceflight
- High Altitude

Assay Type

- Amplicon Sequencing Assay
- Bisulfite Sequencing
- ChIP-Seq
- Behavior (Gait)
- Gel Electrophoresis

Show more ▾

Organism

NASA Open Science for Life in Space

Home About Data & Tools Working Groups Help

Open Science Data Repository Search

Search Datasets

Sort By: Release Date				
Items per page: 25 1 – 25 of 435				
 Persistence of Escherichia coli in the microbiomes of red Romaine lettuce (<i>Lactuca sativa</i> cv. 'Outredgeous')- does seed sanitization matter? Study OSD-385 Microbiota Treatment: Seed Sanitization Factors: Tissue Assay Types: Amplicon Sequencing Release Date: 19-Apr-2024 Description: Seed sanitization via chemical processes removes/reduces microbes from the external surfaces of the seed and thereby could have an impact on the plants, health or productivity. To determine the impact ... Highlights: cgene				
 Transcriptional profiling of heart tissue from mice flown on the RRRM-2 mission Study OSD-580 <i>Mus musculus</i> Spaceflight Age Euthanasia Location Factors Assay Types: transcription profiling Release Date: 03-Jan-2024 Description: In the Rodent Research Reference Mission (RRRM-2), forty female C57BL/6NTac mice were flown on the International Space Station. To assess differences in outcomes due to age, twenty 12-week-old and twe... Highlights: cgene				
 Transcriptional profiling of tibialis anterior muscle from mice flown on the RR-23 mission Study OSD-576 <i>Mus musculus</i> Spaceflight Factors Assay Types: transcription profiling Release Date: 12-Dec-2023 Description: The objective of the Rodent Research-23 mission (RR-23) was to better understand the effects of spaceflight on the eyes, specifically on the structure and function of the arteries, veins, and lymphati... Highlights: cgene				
 Ionizing radiation induces transgenerational effects of DNA methylation in zebrafish Study OSD-524 <i>Danio rerio</i> Ionizing Radiation Generation Factors Assay Types: DNA methylation profiling Release Date: 31-Aug-2023 Description: Ionizing radiation is known to cause DNA damage, yet the mechanisms underlying potential transgenerational effects of exposure have been scarcely studied. Previously, we observed effects in offspring ...				

- [Description](#)
- [Experiments](#)
- [Payloads](#)
- [Missions](#)
- [Protocols](#)
- [Samples](#)
- [Assays](#)
- [Publications](#)
- [Files](#)
- [Version History](#)
- [Visualization](#)

 OSD-137 Version 8

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

2.50 TB

[Study](#)

GeneLab ID: GLDS-137
ALSDA ID: LSDS-28
DOI: [10.26030/9k6w-4c28](https://doi.org/10.26030/9k6w-4c28)
License: CC0-1.0

[Cite this Study](#)

 **Description**

Description

The Rodent Research-3 (RR-3) mission was sponsored by the pharmaceutical company Eli Lilly and Co. and the Center for the Advancement of Science in Space to study the effectiveness of a potential countermeasure for the loss of muscle and bone mass that occurs during spaceflight. Twenty BALB/c, 12-weeks old female mice (ten controls and ten treated) were flown to the ISS and housed in the Rodent Habitat for 39-42 days. Twenty mice of similar age, sex and strain were used for ground controls housed in identical hardware and matching ISS environmental conditions. Basal controls were housed in standard vivarium cages. Spaceflight, ground controls and basal groups had blood collected, then were euthanized, had one hind limb removed, and finally whole carcasses were stored at -80 C until dissection. All mice in this data set received only the control/sham injection.

Submitted Date:
04-Aug-2017

Initial Release Date:
28-Aug-2017

OSD-137 Version 8

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

2.50 TB **Study**

GeneLab ID: GLDS-137
ALSDA ID: LSDS-28
DOI: [10.26030/9k6w-4c28](https://doi.org/10.26030/9k6w-4c28)
License: CC0-1.0

Cite this Study

Smith R, Cramer M, Globus R, Galazka J. "Rodent Research-3-CASIS: Mouse liver transcriptomic, proteomic, epigenomic and histology data", NASA Open Science Data Repository, Version 8, <http://doi.org/10.26030/9k6w-4c28>

[Download Citation \(BibTex\)](#) [Download Citation \(RIS\)](#)

Description

Experiments

Payloads

Missions

Protocols

Samples

Assays

Publications

Files

Version History

Visualization

Samples

Select Export Columns

Select columns of the samples panel below to export [Export CSV](#)

Source Name	Sample Name	Characteristics: Organism	Characteristics: Strain	Characteristics: Animal Source	Characteristics: sex	Characteristics: Age at Launch	Characteristics: Diet	Characteristics: Strain	Characteristics: Diet	Comment: Feeding Schedule	Characteristics: Material Type	Factor Value: Spaceflight	Protocol REF
CB1	Mmus_BAL-TAL_LVR_BSL_Rep1_B1	Mus musculus	BALB/c	Taconic Biosciences	female	12 week	Nutrient Upgraded Rodent Food Bar (NuRFB)	ad libitum	liver				
CB2	Mmus_BAL-TAL_LVR_BSL_Rep2_B2	Mus musculus	BALB/c	Taconic Biosciences	female	12 week	Nutrient Upgraded Rodent Food Bar (NuRFB)	ad libitum	liver				
CB3	Mmus_BAL-TAL_LVR_BSL_Rep3_B3	Mus musculus	BALB/c	Taconic Biosciences	female	12 week	Nutrient Upgraded Rodent Food Bar (NuRFB)	ad libitum	liver				

Search Bar

- Description
- Experiments
- Payloads
- Missions
- Protocols
- Samples
- Assays
- Publications
- Files**
- Version History
- Visualization

Files

Study Files

Download

Search Files

OSDR API

OSDR AWS s3

Selected: 3

OSD-137

Study Metadata Files

Histology

RNA-Seq

Whole Genome Bisulfite Sequencing

GeneLab Processed RNA-Seq Files

Differential Expression Analysis Data

GLDS-137_rna_seq_contrasts.csv 407.0 B Fri Feb 17 2023

GLDS-137_rna_seq_differential_expression.csv 25.5 MB Fri Feb 17 2023

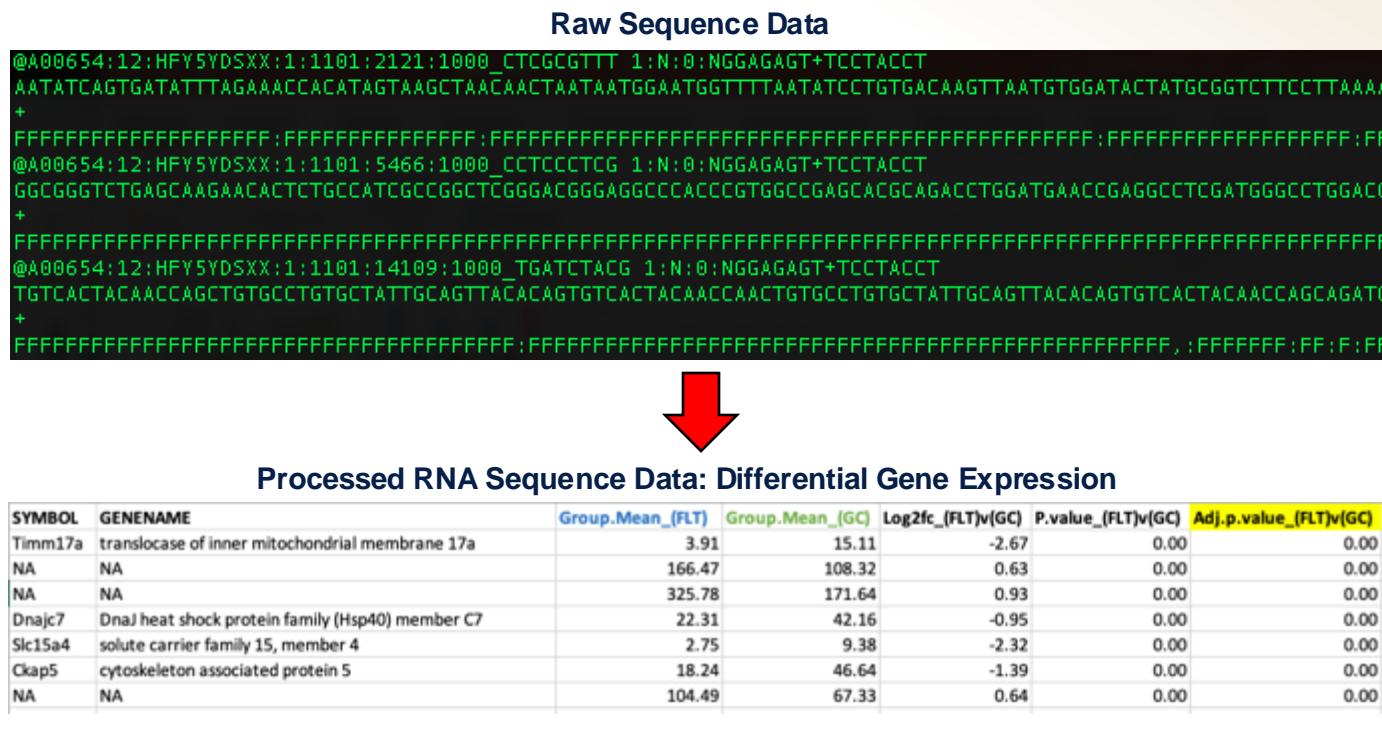
GLDS-137_rna_seq_SampleTable.csv 855.0 B Fri Feb 17 2023

Merged sequence data

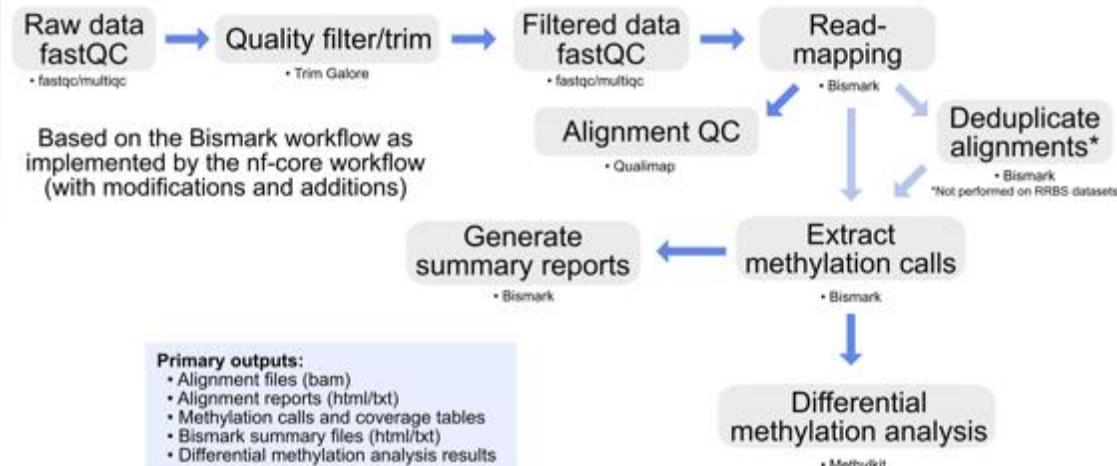
Trimmed sequence data

Aligned sequence data

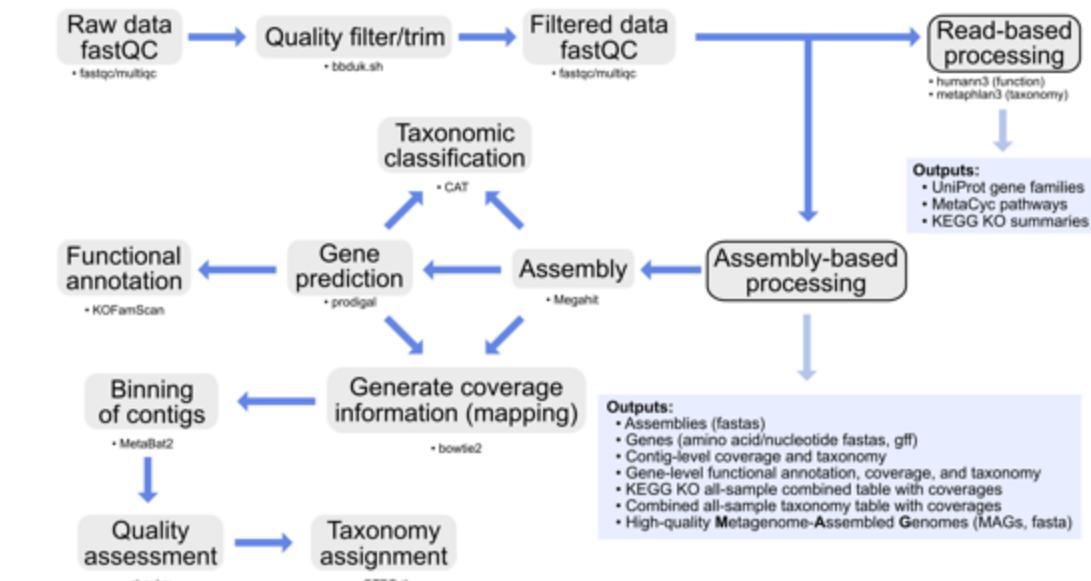
Build consensus data processing pipelines with the scientific community



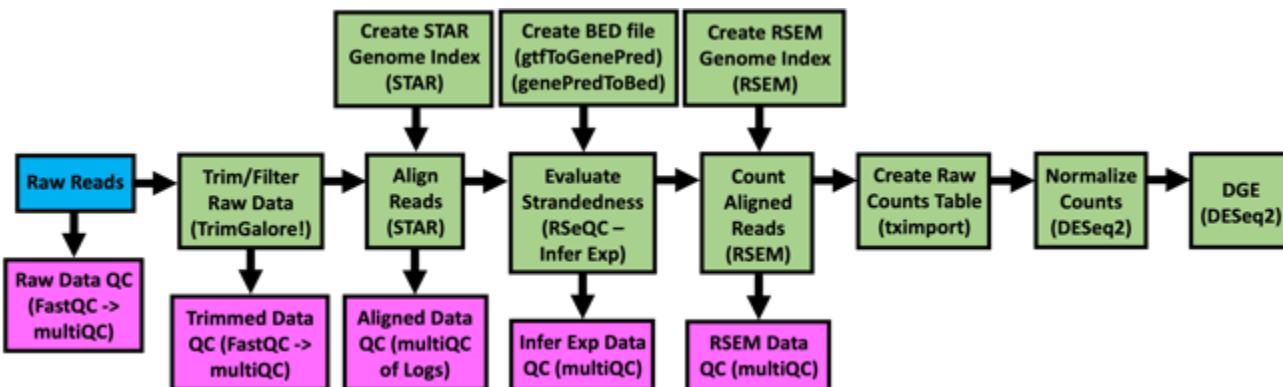
Methylation Sequencing Data



Metagenomics Data



RNA Sequencing Data



The screenshot shows the GitHub repository page for [nasa/GeneLab_Data_Processing](https://github.com/nasa/GeneLab_Data_Processing). The repository is public and contains a README.md file. The README features the GeneLab logo (a blue circle with a DNA helix and a planet Earth) and the text "Open Science for Life in Space". Below the logo, the repository name "GeneLab_Data_Processing" is displayed. A section titled "About" provides information about the repository's purpose, mentioning the NASA GeneLab Data Processing team and the Analysis Working Group. It explains that standard pipelines have been created for processing omics data from spaceflight and space-relevant experiments. The repository contains processing pipelines for various assay types, with subdirectories for each. A section titled "Assay Types" lists several categories, each with sub-options. The categories include "Create GeneLab Reference Annotations", "Amplicon Sequencing" (with options for Illumina and 454 and Ion-Torrent), "Metagenomics" (with options for Removing human reads and Illumina), "(bulk) RNAseq", and "single cell RNAseq".

https://github.com/nasa/GeneLab_Data_Processing



OSDR Home About ▾ Data & Tools ▾ Research & Resources ▾ Services ▾ Working Groups ▾ Engage with Us ▾ Help ▾

Data Repository
Submission Portal
Workspace
Data Visualization
RadLab
Environmental Data App
Environmental Data
API



ence Data
y

Data Repository (OSDR) enables
n experiments and missions that
health responses of terrestrial life to

Filter

Assay technology type

- DNA microarray 149
- RNA Sequencing (RNA-Seq) 154
- Nucleotide sequencing 75
- mass spectrometry 35
- microarray 8

Organism

- rodent 143
- bacteria 37
- fungal 79
- human 63
- plant 60

Tissue

- root 22
- liver 9
- leaf 8
- whole organism 6
- spleen 5

Factor

- spaceflight 101

Open Science for Life in Space

Home **About** **Data & Tools** **Research & Resources** **Working Groups** **Help** **BEST**

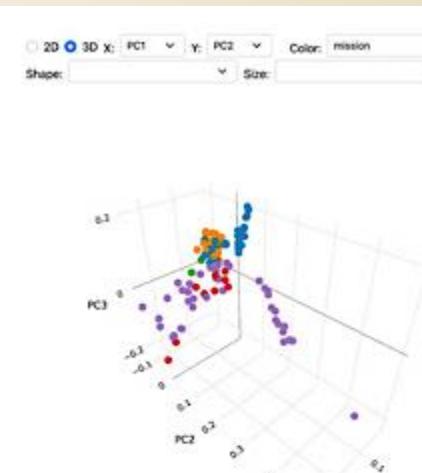
Number of studies
294

The figure consists of four pie charts arranged horizontally. The first chart, titled 'Assay technology type', shows a distribution where RNA Sequencing (RNA-Seq) is the largest segment (light blue), followed by DNA microarray (orange), Nucleotide sequencing (yellow), mass spectrometry (green), and microarray (purple). The second chart, titled 'Organism', shows a distribution where rodent is the largest segment (light green), followed by bacterial (blue), fungal (orange), human (red), and plant (purple). The third chart, titled 'Tissue', shows a distribution where root is the largest segment (light green), followed by liver (blue), leaf (orange), whole organism (purple), and spleen (red). The fourth chart, titled 'Factor', shows a distribution where spaceflight is the largest segment (orange), followed by infection (blue), ionizing radiation (green), and spaceflight (purple).

Show 10 entities

GLDS	Title	Assay	Organism	Tissue	Factor
GLDS-1	Expression data from drosophila melanogaster	DNA microarray	Drosophila melanogaster	whole organism	infection,ionizing radiation,spaceflight,
GLDS-4	Microarray Analysis of Space-flown Murine Thymus Tissue	DNA microarray	Mus musculus	thymus	spaceflight,
GLDS-18	Transcription profiling of rat to study the effect of hindlimb unloading on healing of medial collateral ligaments 3 weeks after injury	DNA microarray	Rattus norvegicus	Medial collateral ligament	hindlimb unloading,treatment,
GLDS-21	Effects of spaceflight on murine skeletal muscle gene expression	DNA microarray	Mus musculus	cell muscle,gastricremius	spaceflight,
GLDS-29	STS-135 Liver Transcriptomics	DNA microarray	Mus musculus	liver	spaceflight,
GLDS-26	Microbiomes of the Dust Particles Collected from the International Space Station during the Expedition 36 Flight	amplicon sequencing assay	cellular organisms	Cells	sample location,

Search

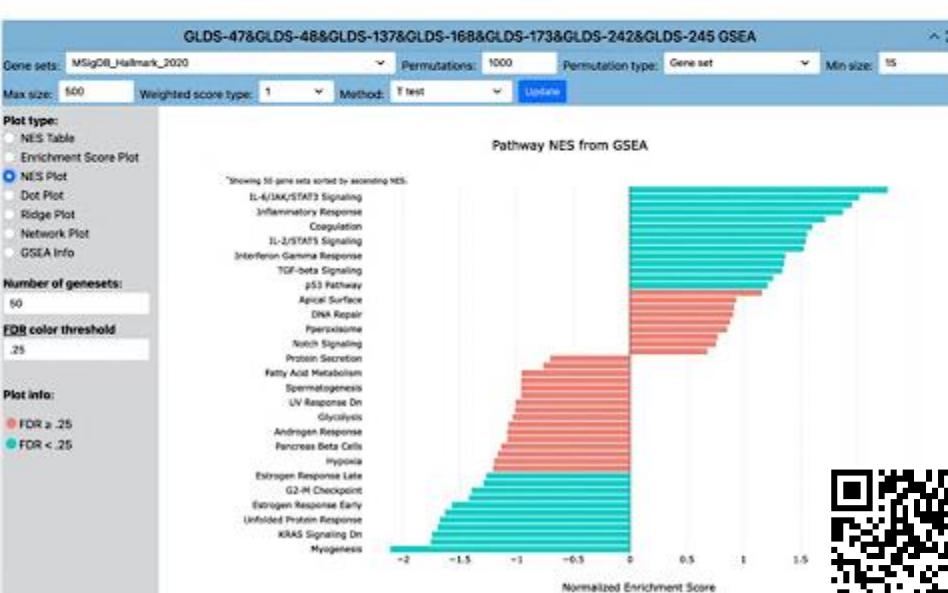
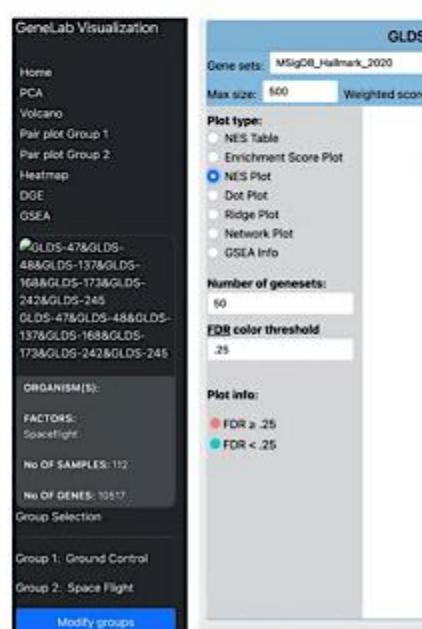


The expanded table will allow you to select the samples you want to use for the Differential Gene Expression Analysis. All samples are selected by default.

The following table lists some selected samples.

GLDS	Sample	mission	library selection
Search	Search	Search	Search
GLDS-47	<input type="checkbox"/> Mmuis_C57-6T_LVR_BSL_Rep1_B1	RR-1	mRNA enrichment
GLDS-47	<input type="checkbox"/> Mmuis_C57-6T_LVR_BSL_Rep2_B2	RR-1	mRNA enrichment
GLDS-47	<input type="checkbox"/> Mmuis_C57-6T_LVR_BSL_Rep3_B3	RR-1	mRNA enrichment
GLDS-47	<input checked="" type="checkbox"/> Mmuis_C57-6T_LVR_FLT_Rep1_F1	RR-1	mRNA enrichment
GLDS-47	<input checked="" type="checkbox"/> Mmuis_C57-6T_LVR_FLT_Rep2_F2	RR-1	mRNA enrichment

GLDS-47&GLDS-48&GLDS-137&GLDS-168&GLDS-173&GLDS-242&GLDS-245 DGE					
					Maximum p-value: Maximum adjusted p-value:
					Search:
Copy	CSV	Excel	PDF	Print	
ENSEMBL	Symbol	LOG2FC	PVAL	ADJP	
Search ENS	Search Symbol	Search LOG2FC	Search PVAL	Search ADJP	
ENSMUSG0000000001543	Ank1	-1.8700696226	5e-10	0.0000069515	
ENSMUSG00000000014309	Apol7c	0.956731259	1.42e-8	0.0000561777	
ENSMUSG00000000020108	Ddit4	-1.2769254891	3.85e-8	0.0001143909	
ENSMUSG00000000032080	Apos4	-1.2372393607	1.34e-7	0.0003180109	
ENSMUSG00000000030302	Atp2b2	-1.1445791174	1.526e-7	0.0003293267	
ENSMUSG00000000029188	Slc34a2	-1.51889323708	2.281e-7	0.0004512668	
ENSMUSG00000000030244	Gys2	0.5579028642	0.00000010983	0.0017383056	
ENSMUSG00000000021579	Lrrc14b	-0.9323304158	0.00000012277	0.0017388094	
ENSMUSG00000000048521	Cxcr6	0.60734688945	0.00000012451	0.0017388094	
ENSMUSG0000000006574	Slc4a1	-1.4885038062	0.00000015841	0.0020765442	
ENSEMBL	Symbol	LOG2FC	PVAL	ADJP	



ence Data
y

Data Repository (OSDR) enables
n experiments and missions that
health responses of terrestrial life to

OSDR Home About ▾ Data & Tools ▾ Research & Resources ▾ Services ▾ Working Groups ▾ Engage with Us ▾ Help ▾

Data Repository

Submission Portal

Workspace

Data Visualization

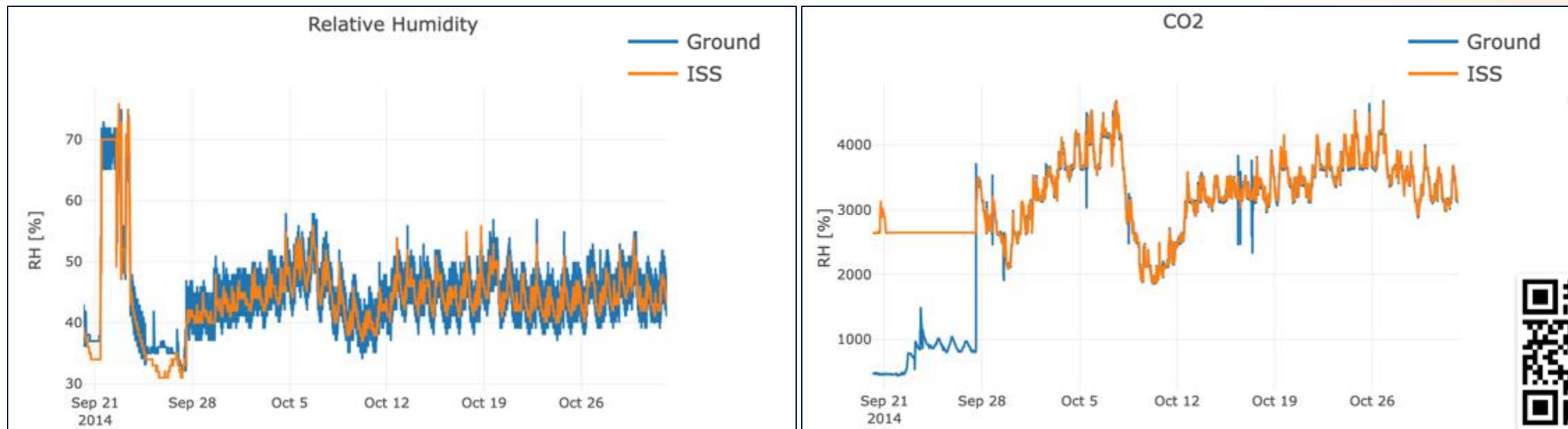
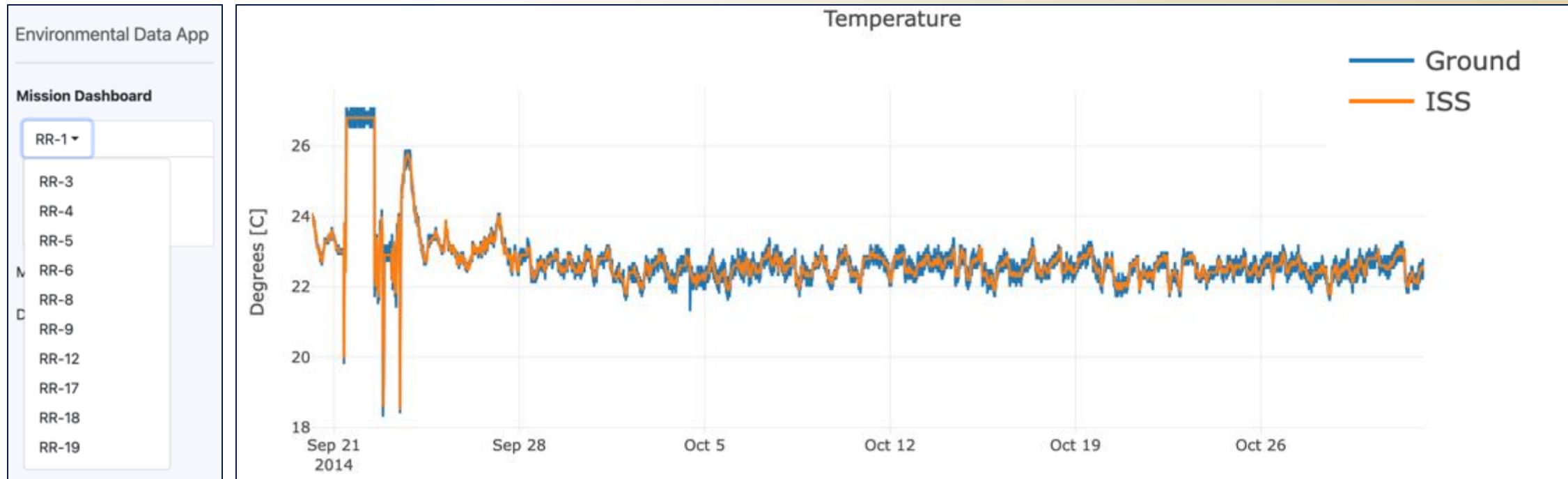
RadLab

Environmental Data App

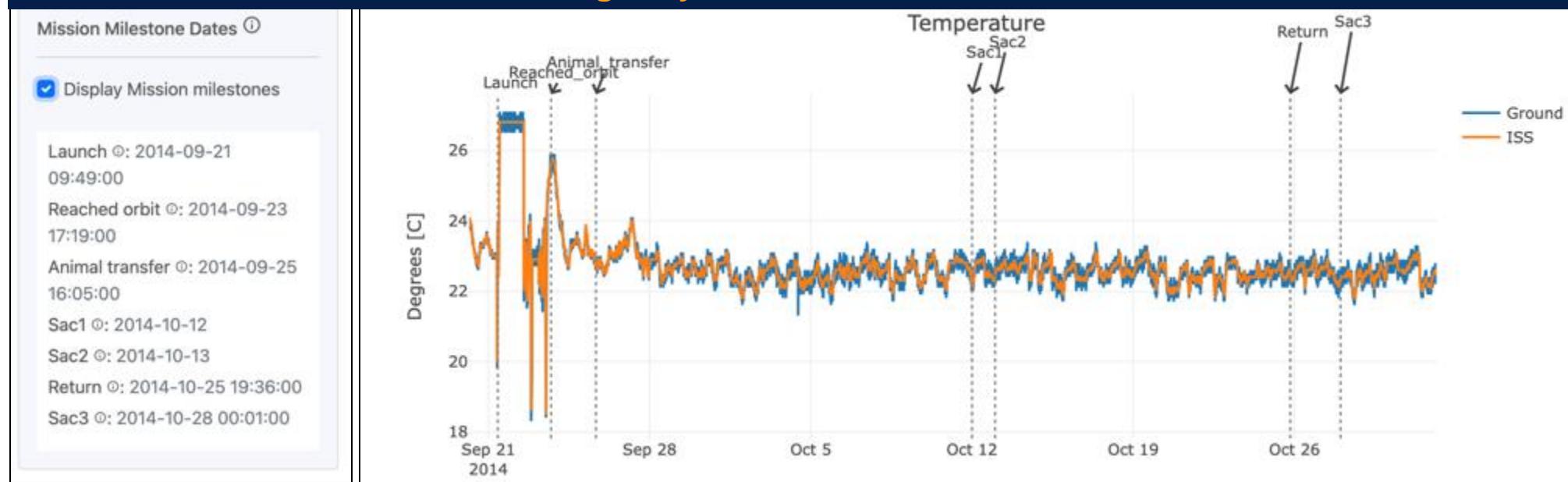
Environmental Data

API

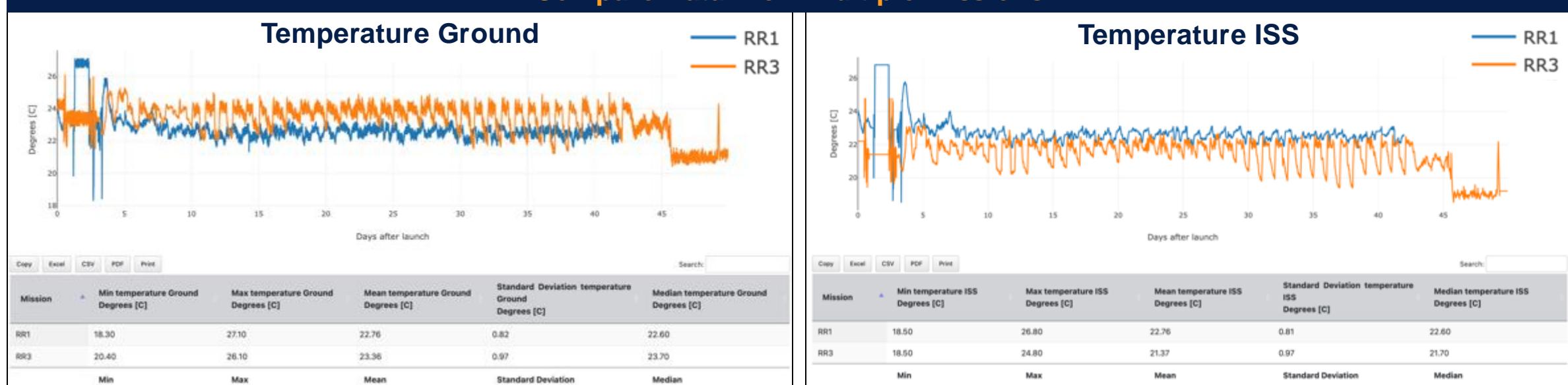




View Biologically Relevant Mission Milestones



Compare Data From Multiple Missions



OSDR Home About ▾ Data & Tools ▾ Research & Resources ▾ Services ▾ Working Groups ▾ Engage with Us ▾ Help ▾

Data Repository

Submission Portal

Workspace

Data Visualization

RadLab

Environmental Data App

Environmental Data

API

ence Data
y

Data Repository (OSDR) enables
n experiments and missions that
health responses of terrestrial life to



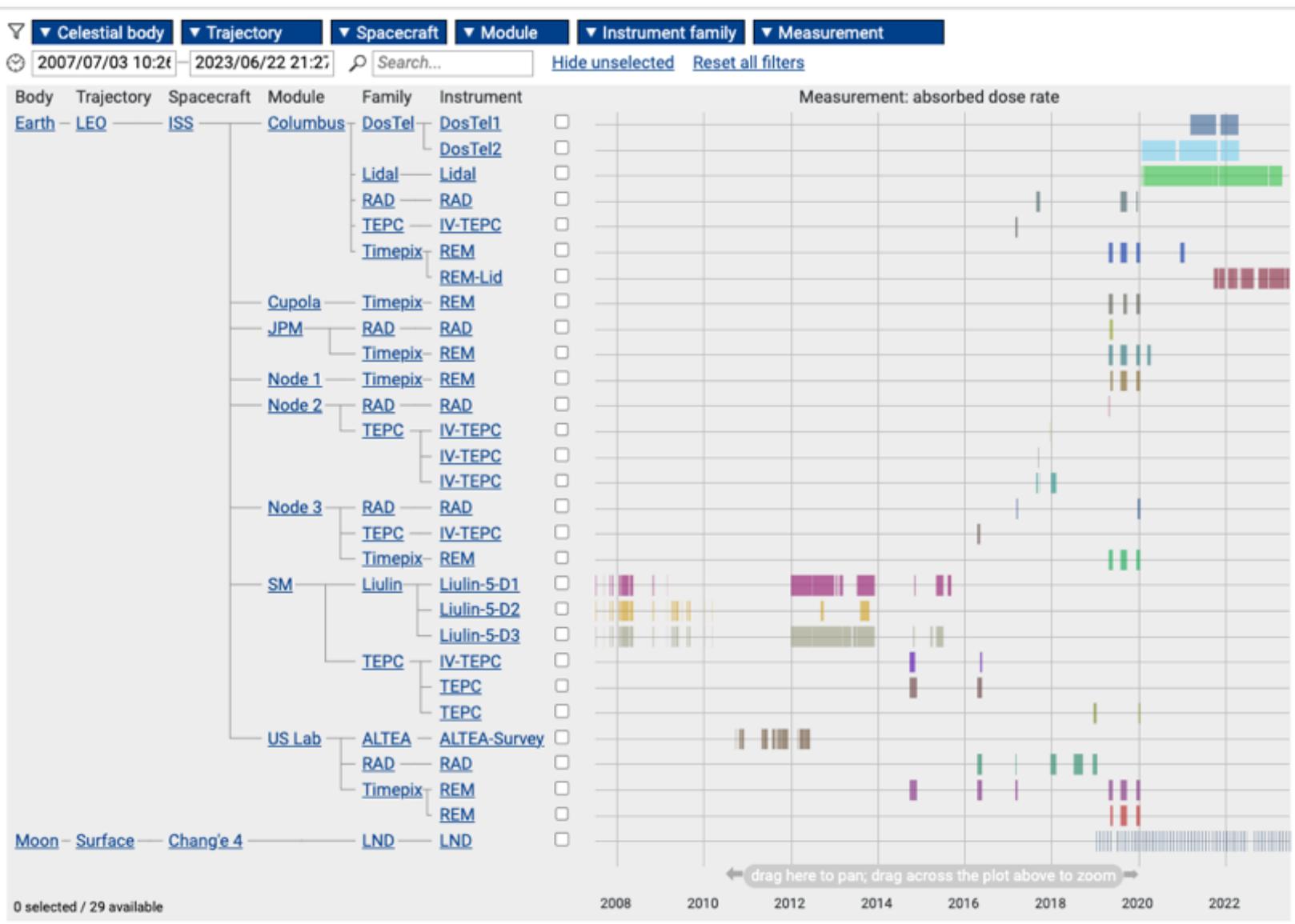


RadLab
Overview
LEO
ISS
BLEO
Time series plots
Data comparison
Geospatial plots
Knowledgebase
Data API
Settings

Overview of available instrument readings

Explore [the data available from the ISS](#) or from [instruments in BLEO](#), or directly explore the overview table below to filter for instruments (using the dropdowns above the table) and time spans of interest (by panning and zooming the interactive plot).

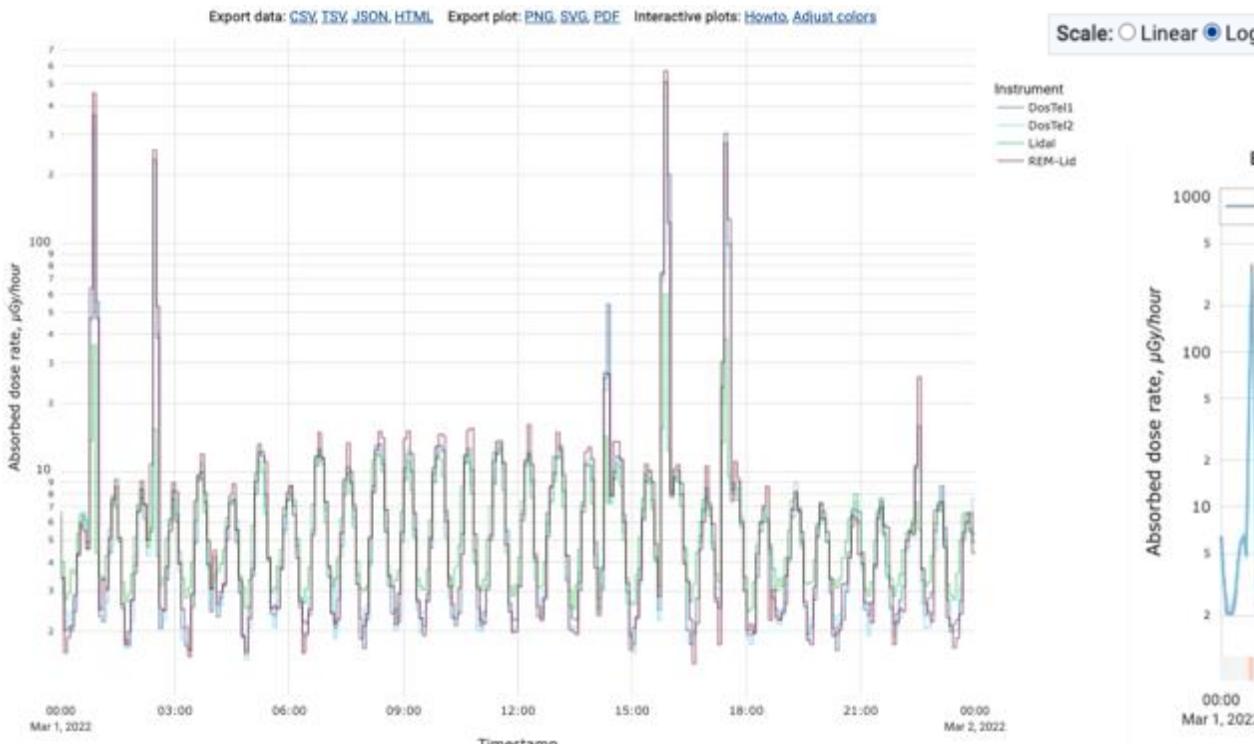
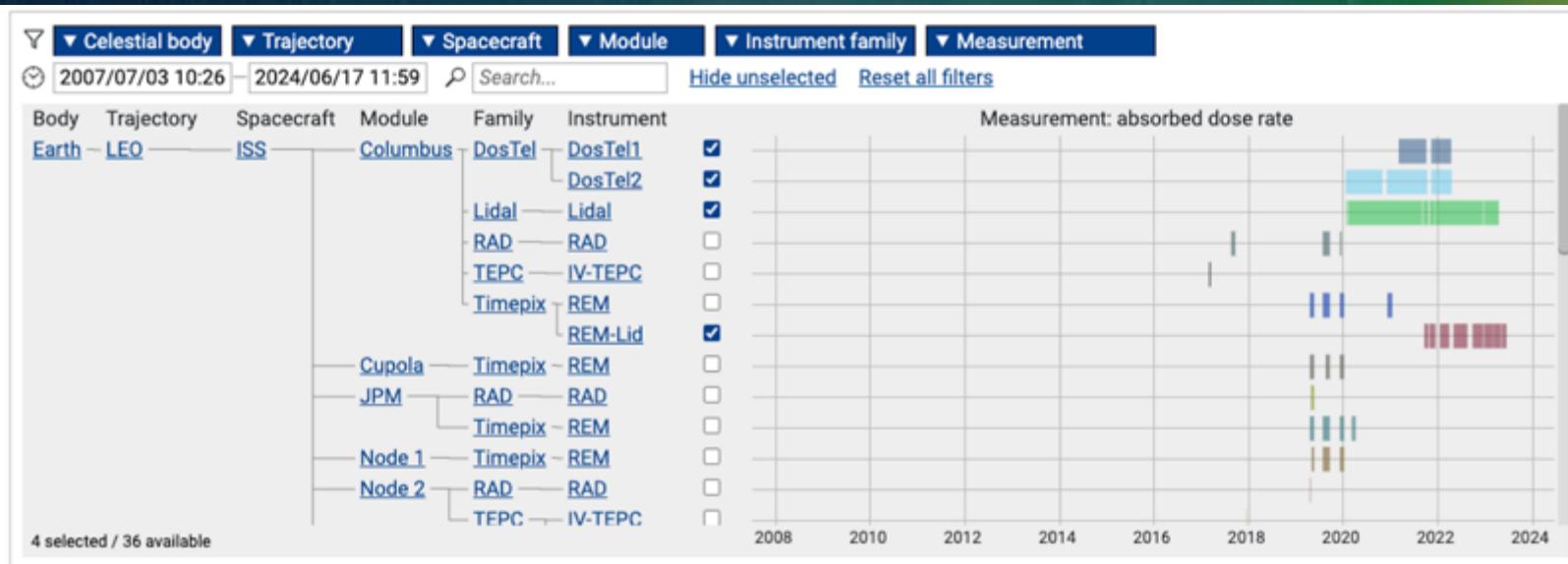
Tick the checkboxes next to the instruments of interest to proceed with more in-depth visualizations.



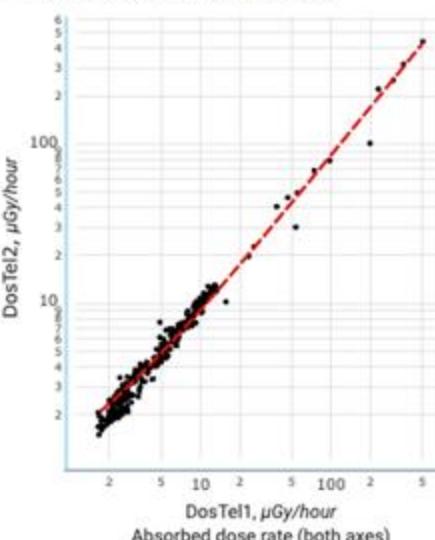
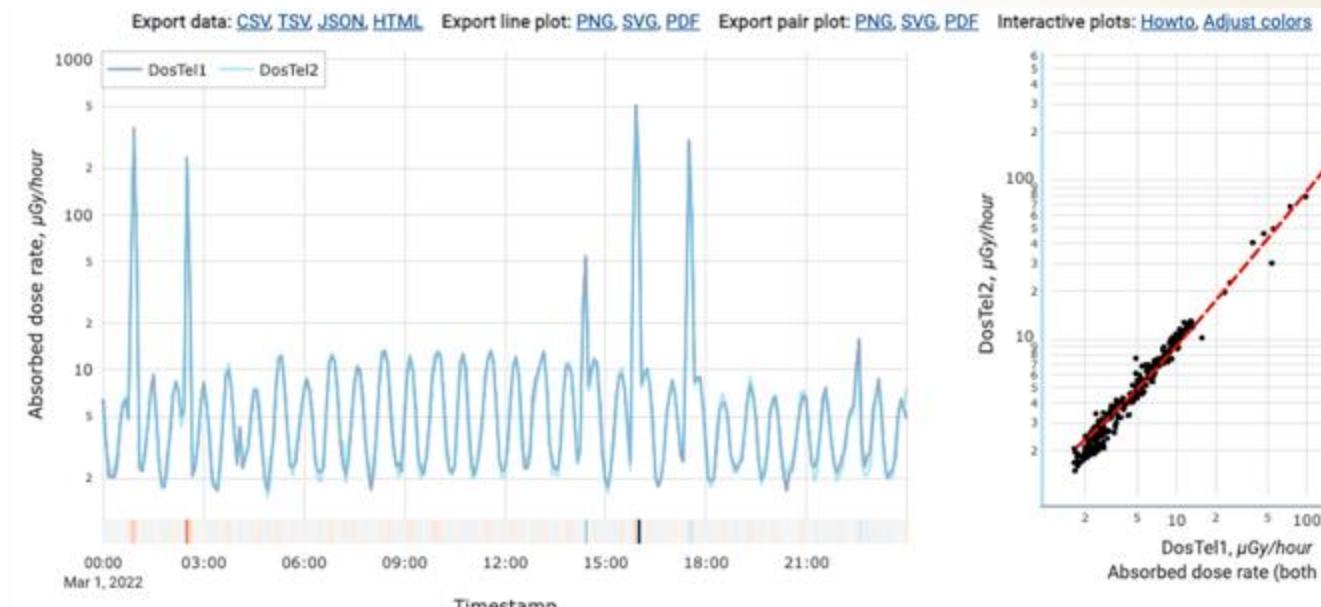
RadLab Portal and Data API

Time Series Plots

RadLab	
Overview	
LEO	
ISS	
BLEO	
Time series plots	
Data comparison	
Geospatial plots	
Knowledgebase	
Data API	
Settings	

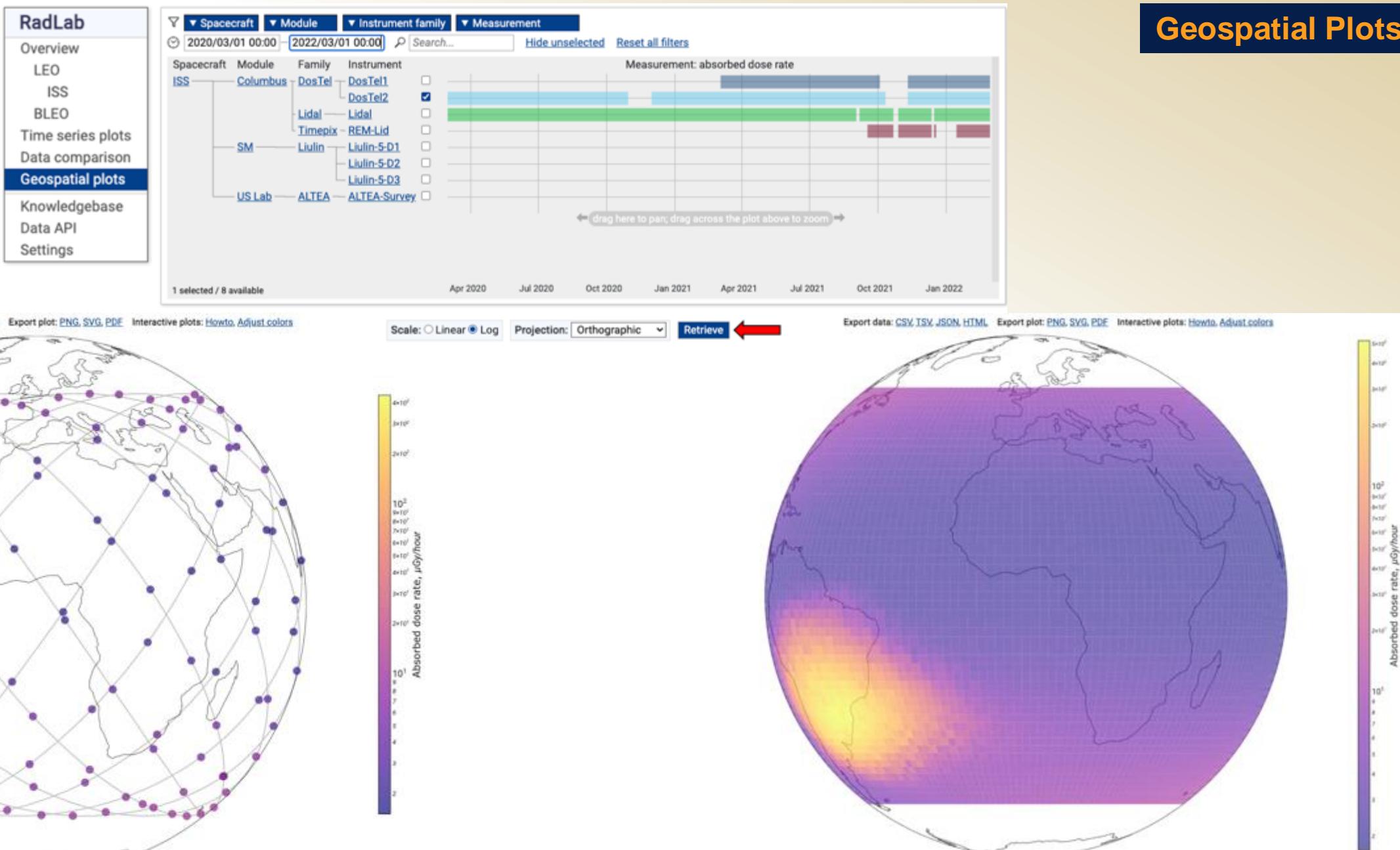


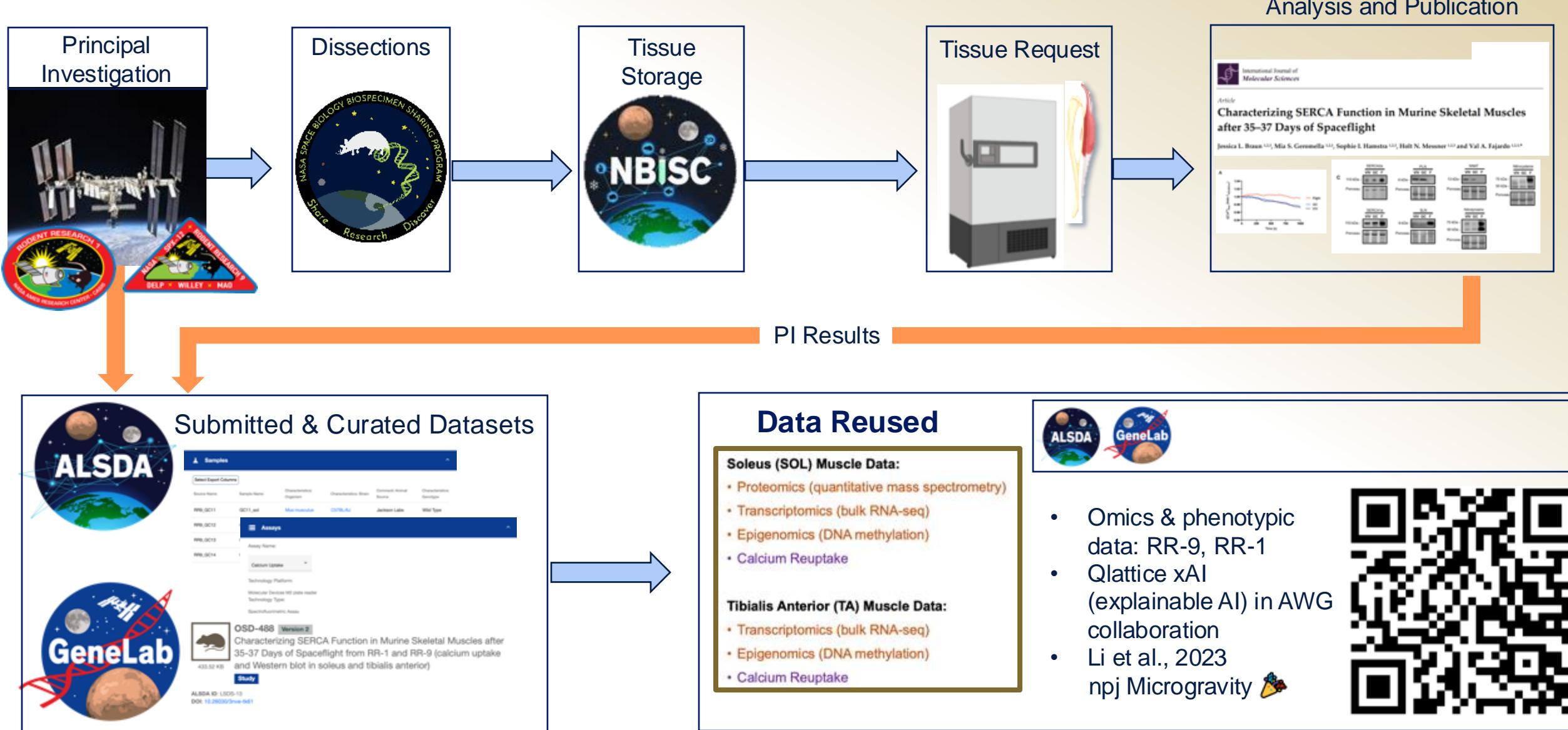
Compare Two Detectors



RadLab	
Overview	
LEO	
ISS	
BLEO	
Time series plots	
Data comparison	
Geospatial plots	
Knowledgebase	
Data API	
Settings	

Geospatial Plots





Training Programs Using OSDR Data

GeneLab for High Schools (GL4HS)



Overview:

- **GeneLab for High Schools (GL4HS)** is a four-week space biology virtual training program for rising high school juniors and seniors in the US.
- For the **GL4HS Teacher Program**, educators learn alongside students for weeks 1-2 then spend weeks 3-4 developing personalized curriculum material using teacher template documents.
- The program has three components:
 - **LEARN** general cell biology, bioinformatics and omics, and applications for space biology both on Earth and in spaceflight conditions.
 - **NETWORK** with professionals including NASA researchers, and other guests from academia, industry, and private and non-profit sectors.
 - **RESEARCH** and present an analysis of a transcriptomic dataset from the Open Science Data Repository.



Objectives:

- To inspire the next generation of scientists to become involved in **space-based research**
- To introduce and teach high school students about '**omics data**' and how it relates to biological research, with a specific focus on **space-based research**
- To introduce and teach high school students **basic bioinformatics skills** enabling them to excel in both their college curriculum and academic research laboratories
- To provide students with **networking opportunities** with their peers, NASA research scientists, NASA Space Biology management and University professors
- To present students with an opportunity to conduct **hypothesis driven research** of a publishable standard in collaboration with a NASA mentor at a NASA research facility
- To provide tools for teachers to **include GeneLab and bioinformatics analysis in their classrooms**



Background:

- NASA's GeneLab project empowers researchers with open access to space-relevant multi-omics data through the [Open Science Data Repository \(OSDR\)](#)
- GeneLab for Colleges and Universities (GL4U) offers **bioinformatics training for space biology** to:
 - Increase accessibility and interpretability of multi-omics Space Biology data
 - Train a diverse next generation of Space Biology researchers
 - Enhance awareness and understanding of Space Biology data



Objectives:

- **Educate and train** the next generation of scientists to process, analyze, and interpret space-relevant 'omics data using publicly available data and bioinformatics tools
- **Maximize** the number of scientists who understand and utilize NASA's open-source 'omics data and tools
- **Provide educators with the knowledge and resources** required to train and inspire their students using GeneLab bioinformatic analyses as an entree into Space Biology

GL4U materials are organized into introduction and omics-specific modules

Introduction Module

- Overview Lecture (NASA, SMD, BPS, OSDR, GeneLab)
- Jupyter Lab Tutorial
- Unix Jupyter Notebook (hands-on training)
- R Jupyter Notebook (hands-on training)

The Intro Module serves as a pre-requisite for any omics-specific module

Unix Intro JN

5. Running commands

Using the foundational rules described above, we will begin running some commands.

```
In [2]: date
```

date is a command that prints out the date and time. This particular command doesn't require any arguments:

```
Sat Jul 8 22:48:55 PDT 2823
```

When we run `date` with no arguments, it uses some default settings, like assuming we want to know the time in our computer's currently set time zone. But we can provide optional arguments to `date`:

```
In [2]: date -u
```

`-u` is the argument to tell the `date` program to report UTC time instead of the local time – which will be the same if the computer we're using happens to be set to UTC time:

```
Sun Jul 9 05:41:28 UTC 2823
```

Note that if we try to run the command above without the dash, we get an error (ignore the message that prints out highlighted in red, we wouldn't normally see that outside of a notebook):

```
In [4]: date u
```

date: invalid date 'u'

Note

Notice that the error above comes from the program `date`. So the program we wanted to use is actually responding to us, but it doesn't seem to know what to do with the letter `u` we gave it. And this is because it wasn't prefixed with a dash, like `-u`.

Let's see what happens if we try to enter this without the "space" separating `(date)` and the optional argument `-u`, the computer won't know how to break apart the command and we get a different error (again, ignoring the red output):

```
In [5]: date-u
```

date-u: command not found

R Intro JN

1d. Data frame manipulations

Much of the data we work with in bioinformatics is in the data frame or matrix format. For example, gene expression data is usually held in matrix format, with samples as columns and genes as rows, where each entry (or cell) in the matrix contains the expression of a particular gene in a particular sample.

When analyzing numerical data in table format, it can be useful to be able to perform mathematical functions on all cells in a data frame, such as adding a value to all cells or taking the log of all cells. Fortunately, R makes that easy for us to do.

Below are some examples of common mathematical manipulations we often perform on data frames in bioinformatics.

Add a value to all cells

In R, you can add, subtract, multiply, or divide the number in every cell of a data frame by a specific value very easily. Run the command in the next cell to add `1` to every value in your `myDF` data frame.

```
In [24]: myDF = 1
```

```
A data.frame: 10 x 3
  column1 column2 column3
  <dbl> <dbl> <dbl>
row1  2     3     4
row2  5     6     7
row3  8     9     10
row4 11    12    13
row5 14    15    16
row6 17    18    19
row7 20    21    22
row8 23    24    25
row9 26    27    28
row10 1     1     1
```

The next cell to subtract 2 from all values in your `myDF` data frame.

```
In [25]: myDF = 2
```

Omics-specific Modules (RNAseq, Amplicon Seq, etc.)

- Omics Data Lectures
 - Experimental design
 - Sample preparation and quality control
 - Data processing tools and visualizations
 - Results analysis and interpretation
- Hands-on data processing and analysis of an OSDR dataset via Jupyter Notebooks (JNs)

AmpSeq JN

Note

It's worth noting again that these are not interpretable as "real" numbers of anything due to the nature of sequencing data, but they can still be useful as relative metrics of comparison within a study.

As a reminder:

- the left, `"Chao1"`, is an estimate of total richness (total number of unique "things")
- the right, `Bshannon`, is a metric of diversity – which incorporates "richness" and "evenness" (the relative proportions of all our unique things to each other)

Looking at the plots above:

1. Do you notice any immediate differences between the right and ground-control groups?

Some thoughts

We can also modify the parameters of the `plotRichness()` phyloseq function to group samples based on our treatment groups, which can be helpful sometimes:

```
In [26]: plotRichness(myPhyseq, x = "treatment", color = "treatment", measures = c("Chao1", "Shannon"),
  scale_color_by_treatment = unique(sample_info$sample_label),
  theme_bw() + theme(panel.title = element_blank(), text = element_textsize = 18,
  axis.title.x = element_textangle = 90, vjust = 0.5, hjust = 1)
```

RNAseq JN

4c. Volcano Plot

Finally, let's make a volcano plot to identify a few interesting genes. A volcano plot is a scatterplot which shows the relationships of the adjusted p-value to the log₂ fold change. Genes with large fold changes that are also statistically significant by adjusted p-value are labeled.

First, we'll use the default settings from the `EnhancedVolcano()` function: `log2FC.cutoff > 3`, and its adjusted p-value cutoff is < 10e-6.

Note: You can read more about the `EnhancedVolcano()` function and see some examples by clicking here

```
In [27]: # Usages give showing genes differentially expressed in GLT vs SC
EnhancedVolcano(GLT_vs_SC_output, table,
  lab = GLT_vs_SC_output$Sample,
  x = "log2FC", y = "negLog10P",
  p = "Adj.p.value", log2FC.cutoff = 3,
  title = "GLT versus SC",
  legendLabels = "lab", log2FC = 3,
  log2FC_min = -3, log2FC_max = 3,
  pCutoff = 1e-06,
  pointSize = 3.0,
  labelSize = 4.0,
  esColor = "#3182bd")
```

GLT versus SC

EnhancedVolcano

• NS • log2FC • Adj.p.value

GL4U Approach

GL4U Direct Approach

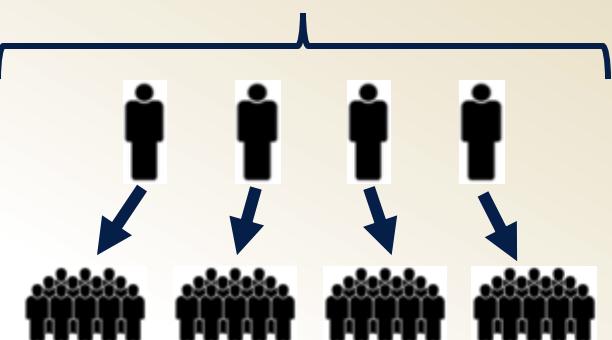


Students

- Space biology-relevant training in bioinformatics
- Uses **direct** (training students) and **indirect** (training educators) approaches
- The GL4U Introduction module and one omics-specific module is taught to students or educators during a 1-2 week-long bootcamp
- Compute resources are provided to all participants
 - For indirect bootcamps: educators receive materials and training to enable them to run the bootcamp at their home institutions
- All GL4U modules are made publicly available on GitHub (<https://github.com/nasa/GeneLab-Training/tree/main/GL4U>) and virtual bootcamps are recorded to enable independent learning

GL4U Indirect Approach

Educators



GL4U RNAseq Certification

GL4U Introduction Module
<input checked="" type="checkbox"/> Pre-Intro Module Survey
<input checked="" type="checkbox"/> Intro Lectures
<input checked="" type="checkbox"/> Intro JNs
<input checked="" type="checkbox"/> Post-Intro Module Survey



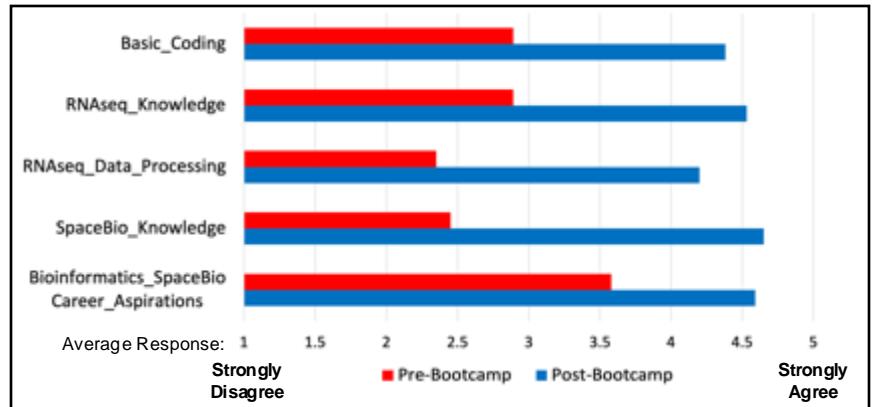
GL4U RNAseq Module
<input checked="" type="checkbox"/> Pre-RNAseq Module Survey
<input checked="" type="checkbox"/> RNAseq Lectures
<input checked="" type="checkbox"/> RNAseq JNs
<input checked="" type="checkbox"/> Post-RNAseq Module Survey

- Pre- and post- bootcamp surveys used to assess participant knowledge before and after training and to collect feedback
- GL4U certification offered for completing the GL4U Introduction module and one omics-specific module
 - *GL4U RNAseq certification shown as an example*

WINTER 2024: GL4U ON-DEMAND

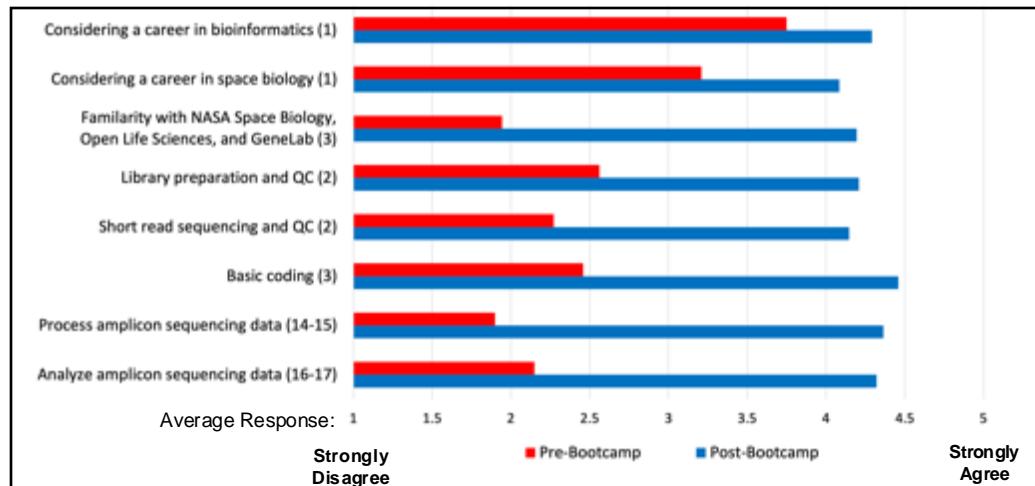
GL4U: RNAseq Student Bootcamp with SJSU, 06/2021

- Virtual 1-week long bootcamp with 17 SJSU students
- Compute resources:* SJSU HPC
- Bootcamp covered GL4U Introduction and RNA Sequencing (RNAseq) modules



GL4U: AmpSeq Student Bootcamp at CSULA, 07/2023

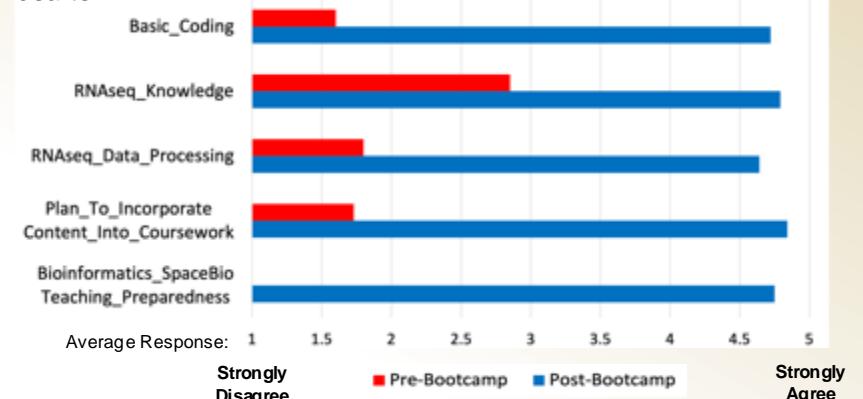
- In-person 4-day bootcamp with 24 CSULA students
- Compute resources:* NSF ACCESS
- Bootcamp covered GL4U Introduction and Amplicon Sequencing (AmpSeq) modules



GL4U: RNAseq Educator Bootcamp, 06/2022

- Virtual 9-day long bootcamp with 6 professors and 4 graduate students from **4 institutions**:
- CalState Northridge
- New Mexico State University
- Alabama A&M
- CalPoly Pomona
- Compute resources:* NASA NCCS SMCE
- Bootcamp covered GL4U Introduction and RNA Sequencing (RNAseq) modules

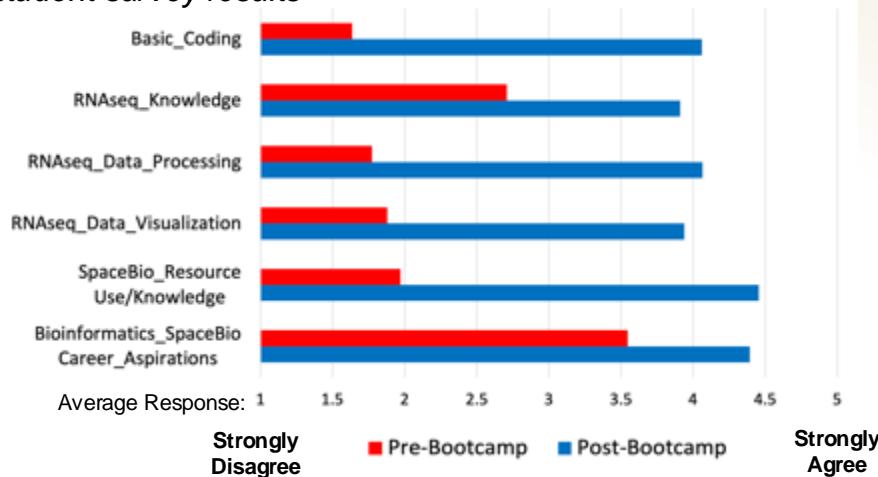
Educator survey results



CPP-GL4U RNAseq Independent Study Course, Fall 2022

- Dr. Lin taught GL4U RNAseq content for 11 junior & senior CalPoly Pomona (CPP) undergraduate students
- Compute resources:* NASA NCCS SMCE

CPP student survey results



- *Training in Artificial Intelligence and Machine Learning for Space Biological Sciences Using NASA Cloud-Based Data*
- AI/ML training with a focus on space biology
- Real-world use cases with NASA cloud-based scientific datasets

Cloud based NASA biology data:https://registry.opendata.aws/bps_rnaseqhttps://registry.opendata.aws/bps_microscopy

The screenshot shows a Canvas course page for 'AI/ML 2024 > Syllabus'. The title bar reads 'Training in Artificial Intelligence and Machine Learning for Space Biological Sciences Using NASA Cloud-Based Data'. The left sidebar includes links for Home, Account, Announcements, Assignments, Discussions, Grades, People, Courses, Calendar, Inbox, Syllabus, Outcomes, Rubrics, Quizzes, Modules, BigBlueButton, Collaborations, Attendance, New Analytics, and Settings. The main content area features the title 'AI/ML AWG Summer 2024 Training Program' and a welcome message: 'Welcome to "Using AI/ML for space biology research"! This class was designed to teach the basics of data analysis and machine learning in the context of space biology research. The course modules and presumptive schedule are as follows:'. It lists four modules: Module 1: Introduction to this course (with points about space biology and Google Colab), Module 2: Working with data (with points about tabular, image data, and visualizing), Module 3: Building basic models (with points about clustering, regression, and classification), and Module 4: Interpreting results (with points about bioinformatic tools, Explainable AI, and Open science). A 'Course Summary' table is at the bottom, showing two rows: one for 'Tue Jun 25, 2024' with 'Module 1 lecture (interactive)' due by 7am, and another for 'Thu Jun 27, 2024' with 'Module 1 hands-on notebook lecture' due by 7am. To the right of the main content, there is a green box containing a bulleted list of course details: '2024 Course: 6/25 – 7/25 Tues & Thurs 7-830am PT', 'Topics: Space biology, Building models, Working with data, Interpreting results, Advanced models', 'Content: 8 Lectures, 9 Jupyter Notebooks, 4 Quizzes', 'Participants: 141 High School Students, 7 TAs'. Logos for 'colab' and 'GitHub' are also present.

Training in Artificial Intelligence and Machine Learning for Space Biological Sciences Using NASA Cloud-Based Data

AI/ML AWG Summer 2024 Training Program

Welcome to "Using AI/ML for space biology research"!

This class was designed to teach the basics of data analysis and machine learning in the context of space biology research. The course modules and presumptive schedule are as follows:

Module 1: Introduction to this course

- Introduction to space biology and Google Colab
- Introduction to machine learning for space biology
- Introduction to space biology data

Module 2: Working with data

- Working with tabular data
- Working with image data
- Visualizing data

Module 3: Building basic models

- Introduction to clustering
- Introduction to regression
- Introduction to classification

Module 4: Interpreting results

- Using bioinformatic tools
- Explainable AI
- Open science

Course Summary:

Date	Details	Due
Tue Jun 25, 2024	Module 1 lecture (interactive)	due by 7am
Thu Jun 27, 2024	Module 1 hands-on notebook lecture	due by 7am

2024 Course:

- 6/25 – 7/25
- Tues & Thurs 7-830am PT

Topics:

- Space biology
- Building models
- Working with data
- Interpreting results
- Advanced models

Content:

- 8 Lectures
- 9 Jupyter Notebooks
- 4 Quizzes

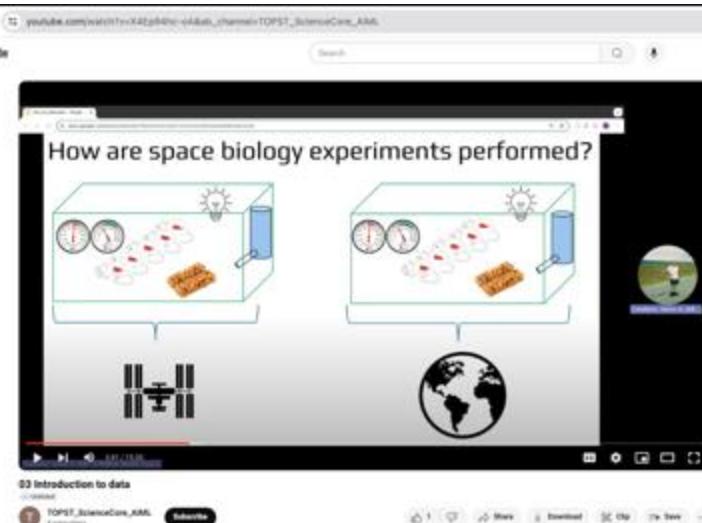
Participants:

- 141 High School Students
- 7 TAs

colab

GitHub

Pre-recorded lectures



Quizzes

Module 3 Quiz
Started: Jun 19 at 8:34am
Quiz Instructions

Question 1 1 pts

Regression is an unsupervised machine learning algorithm

True
 False

Question 2 1 pts

Which of the following is used to draw a regression line?

bias
 principal components
 variance
 ordinary least squares

Interactive Python notebooks

```

1 # create an array y that represents whether the sample is ground control or spaceflight
2 y = np.array([i['Metadata']['255']['Factor Value[spaceflight]'] for i in data])
3 target_names = np.array(['Ground Control', 'Space Flight'])
4
5 # run PCA to reduce dimensions from 23,419 to 2!
6 pca = PCA(n_components=2)
7 X_r = pca.fit_transform(X)
8
9 # Percentage of variance explained for each components.
10 print(pca.explained_variance_ratio_)
11 # explained variance ratio-(first two components): %
12 # % str(pca.explained_variance_ratio_)
13 )
14
15 # plot the pca plot
16 n=plt.figure()
17 colors = ["navy", "turquoise"]
18 bc = 2
19 for color, i, target_name in zip(colors, [0, 1], target_names):
20     n=plt.scatter(X_r[iy == i, 0], X_r[iy == i, 1], color=color, alpha=0.8, lw=1, label=target_name)
21 n=plt.legend(loc='best', shadow=False, scatterpoints=1)
22 n=plt.title('PCA of OSD-255 dataset')
23
24 # show the last 3 digits of the sample name
25 samples = list(df.columns)[1]
26 for i, ttt in enumerate(samples):
27     n=plt.annotate(ttt[-3:], (X_r[i][0], X_r[i][1]))
28 n=plt.show()
29
30 explained_variance_ratio_(first two components): [0.72997254 0.08189642]

```

PCA of OSD-255 dataset

To enroll in this course, visit:



<https://canvas.instructure.com/enroll/8JYKD7>

Training in Artificial Intelligence and Machine Learning for Space Biological Sciences Using NASA Cloud-Based Data

- 2 week curriculum
- 2.5 hours / day
- Topics:
 - Space biology
 - Working with data
 - Building models
 - Interpreting results
 - Capstone projects
- Available Summer 2024 for independent learning





Get Involved!



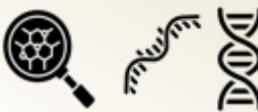
ANIMAL

180 members



MULTI-OMICS

306 members



MICROBIAL

163 members



PLANTS

132 members



AI/ML

230 members



ALSDA

(Physiological/BioMedical)

189 members



Feedback on Ingress/Egress Standards

Ingestion, Curation,
Processing Standards



Standards to Enable
Reuse/Data Mining

Collaborate on Data Mining/Publications



Cell Press Package 2020:
The Biology of Spaceflight
<https://www.cell.com/c/the-biology-of-spaceflight>



Nature Portfolio Collection 2024:
Space Omics and Medical Atlas across orbits (SOMA)
<https://www.nature.com/collections/ebdbcahdgc>



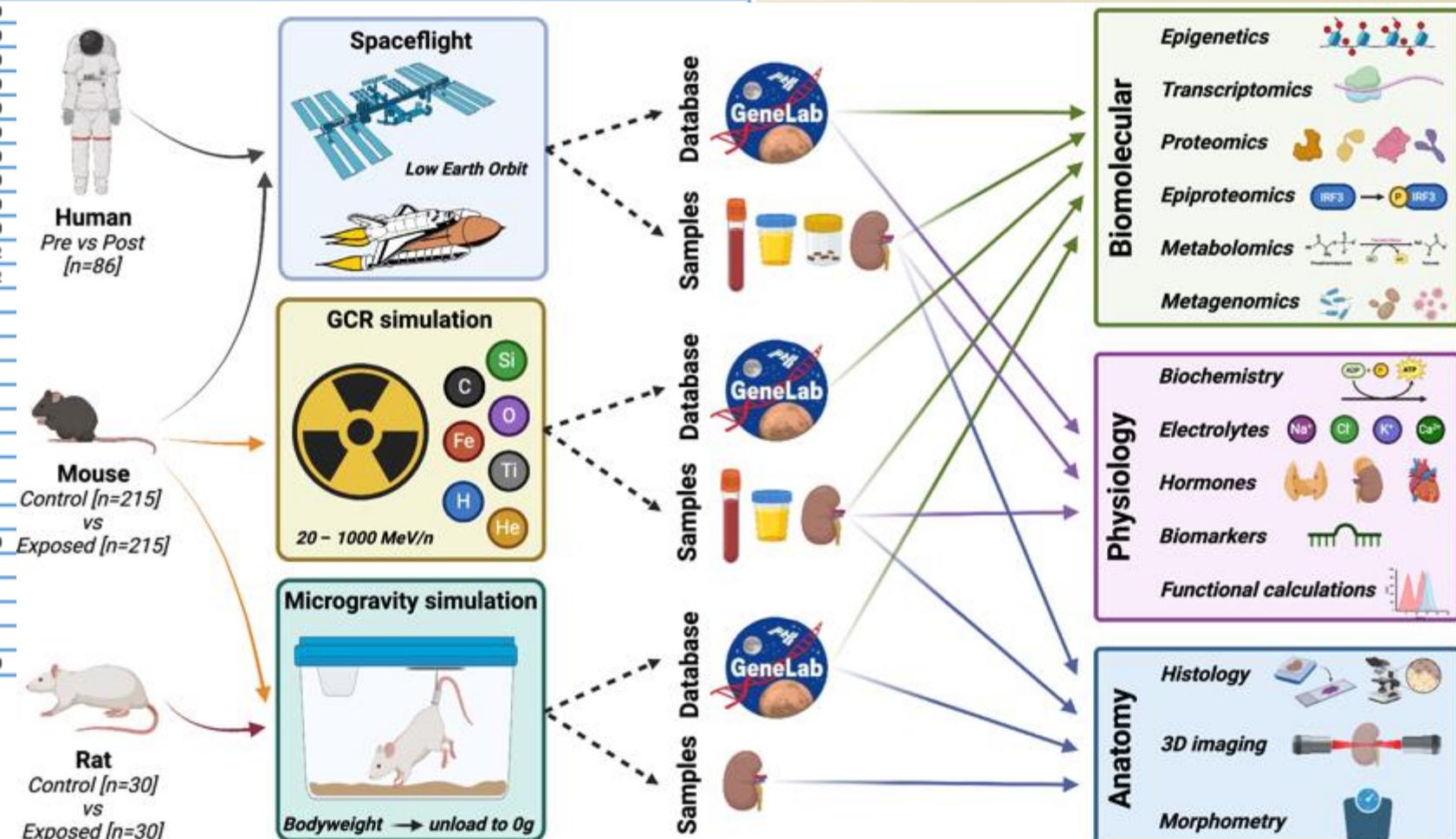
List of 74 publications enabled by OSDR
<https://osdr.nasa.gov/bio/data/publications.html>

[Join an AWG!](#)

BPC Example of Open Science AWG Collaboration and Publication

46

Mission	Experimental Condition	Species	Sex	Background	GCR (mG)	Mission Duration (d)	Sample Collection	n	Control	Data Source	Tissue	DNA	RNA	Protein	PTMs	Metab	Microbes	Physiol	Anatomy
Inspiration4	Spaceflight	Human	Mix	USA	Unknown	3	R+0	4	Paired	Original	P, U		✓	✓	✓				
JAXA	Spaceflight	Human	♂	Japan	Unknown	120	L+120	6	Paired	Original	P		✓						
NASA	Spaceflight	Human	Mix	USA	Unknown	180	L+120	66	Paired	Database	P, U								✓
Roscosmos	Spaceflight	Human	♂	Russia	Unknown	199	R+1	10											
BNL-1	simGCRsim				500.0	~1.5yr sim	PE+1	10											
	MGsim	Mouse	♀	C57BL/6	N/A	14	PE+1	10											
	simGCRsim + MGsim				500.0	~1.5yr sim	PE+1	10											
BNL-2	simGCRsim				500.0	~1.5yr sim	PE+1	10											
	MGsim	Mouse	♀	C57BL/6	N/A	14	PE+1	10											
	simGCRsim + MGsim				500.0	~1.5yr sim	PE+1	10											
BNL-3	simGCRsim				500.0	~1.5yr sim	PE+1	10											
	MGsim	Mouse	♀	C57BL/6	N/A	14	PE+1	10											
	simGCRsim + MGsim				500.0	~1.5yr sim	PE+1	10											
NSRL-22A	GCRsim	Mouse	♀	C57BL/6	750.0	~2.5yr sim	PE+180	12											
	GCRsim	Mouse	♂	C57BL/6CR	750.0	~2.5yr sim	PE+180	12											
STS-135	Spaceflight	Mouse	♀	C57BL/6CR	4.7	13	R+0	7											
RR-1	Spaceflight	Mouse	♀	C57BL/6J	5.0	38	L+38	6											
RR-3	Spaceflight	Mouse	♀	BALB/c	5.9	41	L+41	6											
RR-6	Spaceflight	Mouse	♀	C57BL/6NTac	8.2	29	R+4	9											
					16.1	57	L+75	9											
RR-7	Spaceflight	Mouse	♀	C3H/Hej	4.4	25	L+25	5											
				C57BL/6	13.2	75	L+75	5											
RR-9	Spaceflight	Mouse	♂	C57BL/6J	5.5	33	R+1	5											
RR-10	Spaceflight	Mouse	♀	B6129SF2/J	4.2	28	L+28	10											
RR-19	Spaceflight	Mouse	♀	C57BL/6N	5.2	35	R+0	8											
RR-23	Spaceflight	Mouse	♂	C57BL/6J	5.8	37	R+1	9											
MHU-1	Spaceflight	Mouse	♂	C57BL/6J	9.0	35	R+2	6											
MHU-3	Spaceflight	Mouse	♂	C57BL/6NCrl	9.0	31	R+0	6											
CNSA	MGsim	Rat	♂	Sprague-Dawley	N/A	28	R+0	30											



- Renal/Kidney Research**
 - 20 spaceflight missions & ground simulations
 - 50+ datasets
 - 5 continents of collaborators

Cosmic Kidney Disease: The Effects of Spaceflight and Galactic Cosmic Radiation on Renal Structure and Function.

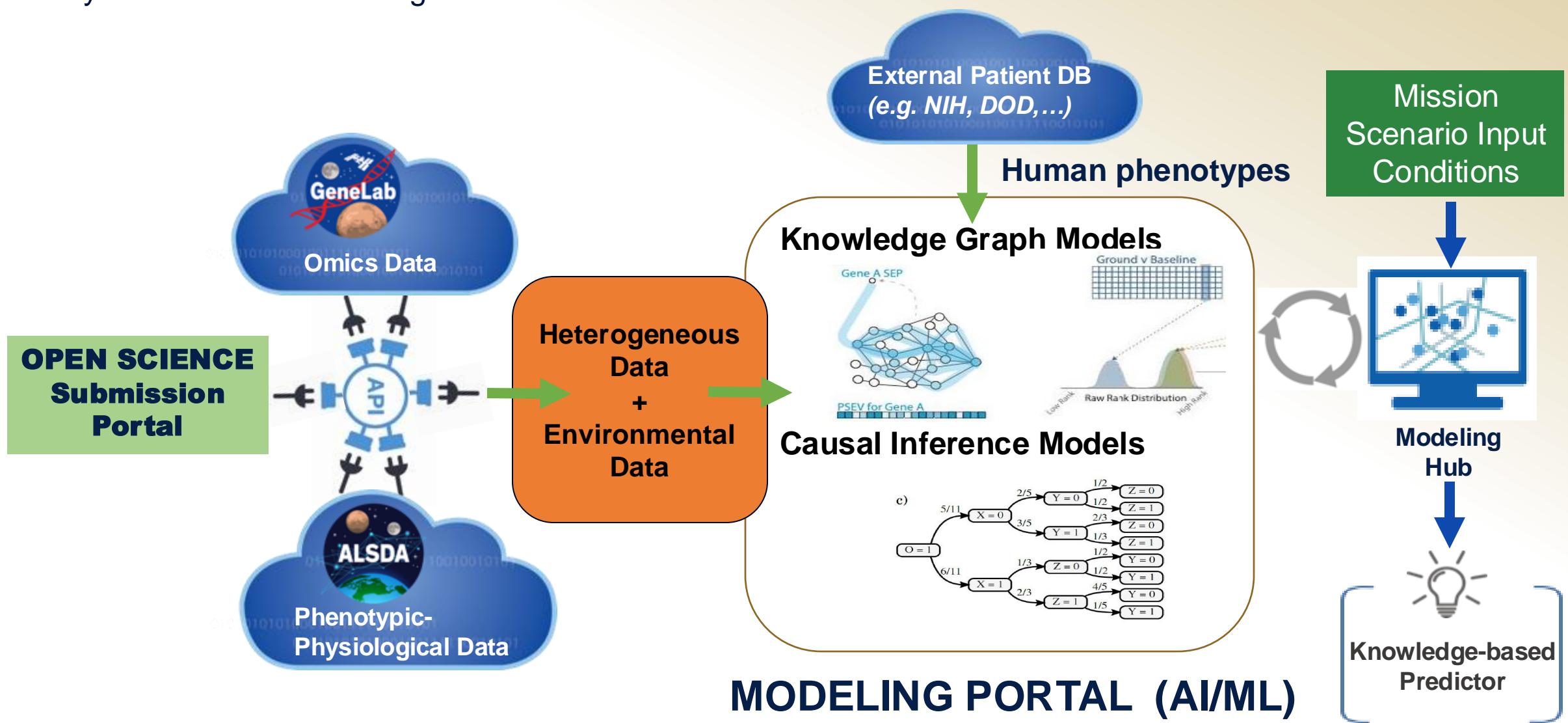
Siew, K., et al., Nat Communications 15, 4568 (June 2024). <https://doi.org/10.1038/s41467-024-49212-1>

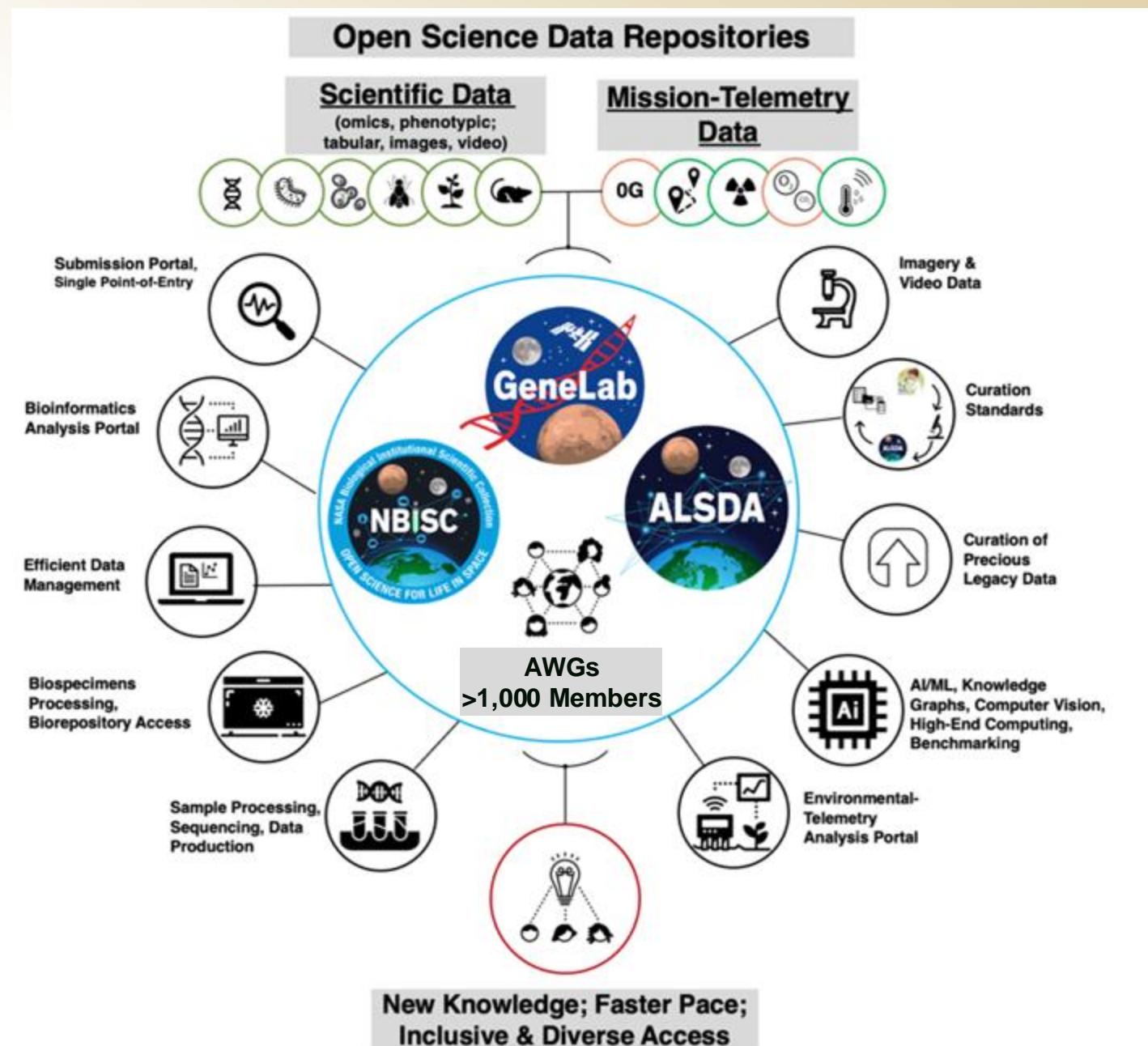
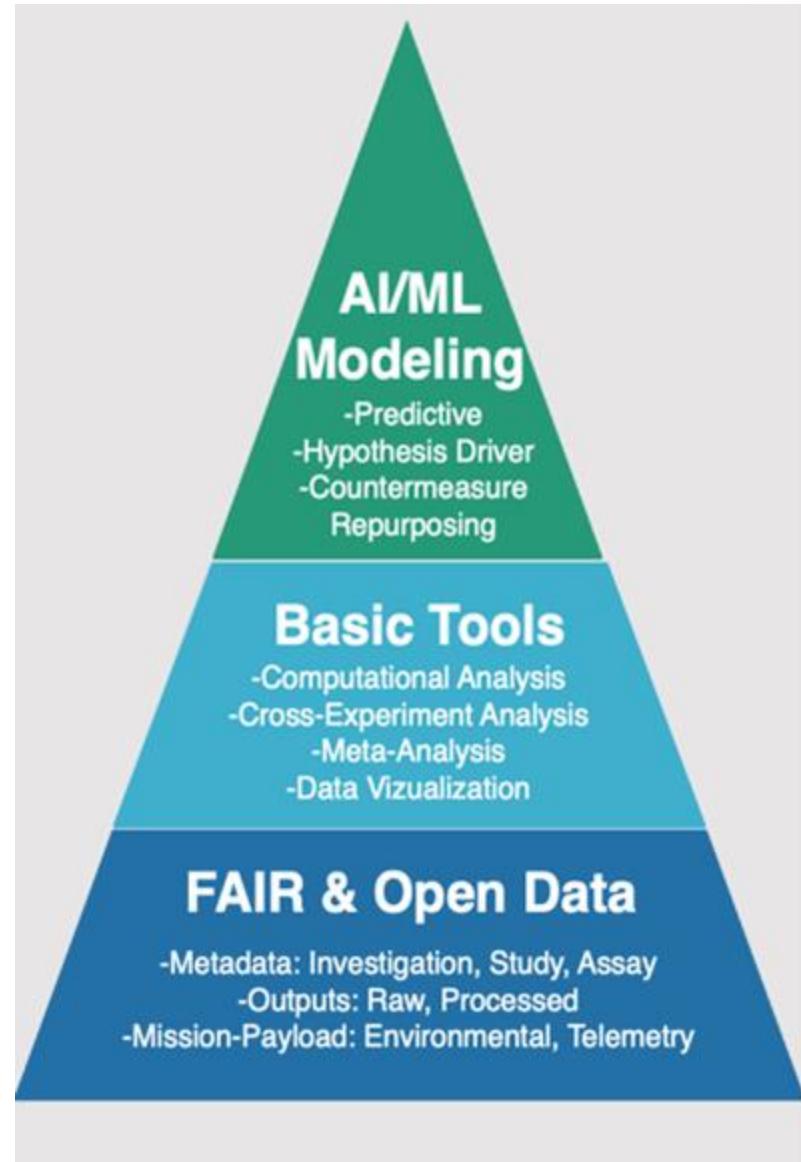




What's Next?

AI ready FAIR data for modeling and risk assessment





Opportunities & Resources

Training and Other Opportunities at NASA



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>

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.

Learn more and apply at : <https://npp.orau.org/>

NASA Pathways Program: Provides current students and recent grads with a dynamic career development program at a NASA center. Participants are given an opportunity for permanent employment with NASA.

Learn more and apply at : <https://www.nasa.gov/careers/pathways-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.

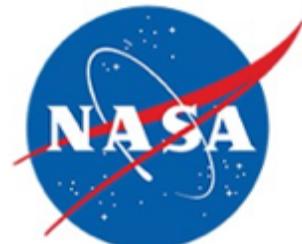
Learn more and apply at : <https://science.nasa.gov/biological-physical/programs/star>

Open Science for Life in Space Teams



Open Science Analysis Working Group Members

Support



NASA Space Biology Program
NASA Science Mission Directorate
NASA Human Research Program
NASA Biological and Physical Sciences



END GL4U: Introduction Lecture 2 of 4