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

Acknowledgements

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."

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

Space Travel Introduces Risks for Living Systems

Afshinnekoo et al., *Cell*, 2020

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

(Scott et al., 2023 https://rdcu.be/c8jSO; Sanders et al., 2023 https://rdcu.be/c8jSS)

(Berliner et al., 2022 https://doi.org/10.1038/s44172-022-00012-9; Everroad et al., 2021 https://ntrs.nasa.gov/citations/20210023324)

NASA Biological Open Science Resources

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

NASA Open Science Data Repository Overview

Physiological/Phenotypic/Imaging/ Environmental Telemetry Data

Tabular, text, imaging, video, code

ALSDA

Molecular/Omics Data

GeneLab

Biospecimens

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)
- Maximally Open Access with Necessary Controls for Sensitive Data
- Data Maximally FAIR

What are ‘omics?

(Andy Vierstraete 1999)

What are ‘omics?

Why is studying omics important for spaceflight?

- 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

NASA Open Science Data Repository Overview

Physiological/Phenotypic/Imaging/ Environmental Telemetry Data

Tabular, text, imaging, video, code

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)
- 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

Civilian Human Spaceflight Data

INSPIRATION 4

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

Chris Mason

Eliah Overbey

Weil Cornell Medicine

IAS International Institute for Astronautical Sciences

- **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**

Shawna Pandya
IIAS Space Medicine

Access Request Process for Private Data

Data Access Info Available On DAR Portal

If Approved

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

Open Science Data Repository (nasa.gov/osdr)

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

Open Science Data Repository

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?

Submission Portal – BDME (https://osdr.nasa.gov/bio/submission-sso-login.html)

Sample level data

Sample Name	Organism	Strain	Material Type	Protocol REF	Parameter Value - Storage Temperature Setting	Unit	Ionizing Radiation	Time	Unit	Dose	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

Terms are connected to a controlled ontology

- 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.

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

Open Science Data Repository (nasa.gov/osdr)

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

Open Science Data Repository

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?

Open Science Data Repository (https://osdr.nasa.gov/bio/repo/)

NASA Open Science for Life in Space

Home | About | Data & Tools | Working Groups | Help

General Search Filters

Data Source

- [x] GeneLab
- [x] ALSDA
- [ ] NIH GEO
- [ ] EBI PRIDE
- [ ] ANL MG-RAST

Data Type

- [x] 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

Open Science Data Repository Search

Search Datasets

Sort By: Release Date

Items per page: 25 | 1 – 25 of 435

Study OSD-385

Persistence of Escherichia coli in the microbiomes of red Romaine lettuce (Lactuca sativa cv. 'Outredgeous')- does seed sanitization matter?

Organisms	Factors	Assay Types	Release Date	Description
Microbiota	Treatment Seed Sanitization Tissue	Amplicon Sequencing	19-Apr-2024	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*

Study OSD-580

Transcriptional profiling of heart tissue from mice flown on the RRRM-2 mission

Organisms	Factors	Assay Types	Release Date	Description
Mus musculus	Spaceflight Age Euthanasia Location	transcription profiling	03-Jan-2024	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*

Study OSD-576

Transcriptional profiling of tibialis anterior muscle from mice flown on the RR-23 mission

Organisms	Factors	Assay Types	Release Date	Description
Mus musculus	Spaceflight	transcription profiling	12-Dec-2023	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*

Study OSD-524

Ionizing radiation induces transgenerational effects of DNA methylation in zebrafish

Organisms	Factors	Assay Types	Release Date	Description
Danio rerio	Ionizing Radiation Generation	DNA methylation profiling	31-Aug-2023	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 ...

OSDR: Study Page

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

2.50 TB

OSD-137 Version 8

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

Study

Submitted Date: 04-Aug-2017
Initial Release Date: 28-Aug-2017

GeneLab ID: GLDS-137
ALSDA ID: LSDS-28
DOI: 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.

OSDR: Cite Study & Export Specific Metadata Table Columns

OSDR: Download Study Files

Search Bar

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

Files

Study Files

Selected: 3

Download

Search Files

OSDR API **OSDR AWS s3**

- 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

GeneLab Data Processing

Build consensus data processing pipelines with the scientific community

Raw Sequence Data

```
@A00654:12:HFY5YDSXX:1:1101:2121:1000_CTCGCGTTT 1:N:0:NGGAGAGT+TCCTACCT
AATATCAGTGATATTTAGAAACCACATAGTAAGCTAACAACTAATAATGGAATGGTTTTAATATCCTGTGACAAGTTAATGTGGATACTATGCGGTCTTCCTTAAAA
+
FFFFFFFFFFFFFFFFFFF:FFFFFFFFFFFFFF:FFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFF:FFFFFFFFFFFFFFFFFF:FF
@A00654:12:HFY5YDSXX:1:1101:5466:1000_CCTCCCTCG 1:N:0:NGGAGAGT+TCCTACCT
GGCGGGTCTGAGCAAGAACACTCTGCCATCGCCGGCTCGGGACGGGAGGCCCACCCGTGGCCGAGCACGCAGACCTGGATGAACCGAGGCCTCGATGGGCCTGGACG
+
FFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFF
@A00654:12:HFY5YDSXX:1:1101:14109:1000_TGATCTACG 1:N:0:NGGAGAGT+TCCTACCT
TGTCACTACAACCAGCTGTGCCTGTGCTATTGCAGTTACACAGTGTCACTACAACCAACTGTGCCTGTGCTATTGCAGTTACACAGTGTCACTACAACCAGCAGATC
+
FFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFF:FFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFF,:FFFFFFF:FF:F:FF
```

Processed RNA Sequence Data: Differential Gene Expression

SYMBOL	GENENAME	Group.Mean_(FLT)	Group.Mean_(GC)	Log2fc_(FLT)v(GC)	P.value_(FLT)v(GC)	Adj.p.value_(FLT)v(GC)
Timm17a	translocase of inner mitochondrial membrane 17a	3.91	15.11	-2.67	0.00	0.00
NA	NA	166.47	108.32	0.63	0.00	0.00
NA	NA	325.78	171.64	0.93	0.00	0.00
Dnajc7	DnaJ heat shock protein family (Hsp40) member C7	22.31	42.16	-0.95	0.00	0.00
Slc15a4	solute carrier family 15, member 4	2.75	9.38	-2.32	0.00	0.00
Ckap5	cytoskeleton associated protein 5	18.24	46.64	-1.39	0.00	0.00
NA	NA	104.49	67.33	0.64	0.00	0.00

RNA Sequencing Data

Methylation Sequencing Data

Metagenomics Data

GeneLab Data Processing GitHub Repo

Search or jump to... / Pull requests Issues Marketplace Explore

nasa / GeneLab_Data_Processing Public

Edit Pins

README.md

Open Science for Life in Space

GeneLab_Data_Processing

About

The NASA GeneLab Data Processing team and Analysis Working Group members have created standard pipelines for processing omics data from spaceflight and space-relevant experiments. This repository contains the processing pipelines that have been standardized to date for the assay types indicated below. Each subdirectory in this repository holds current and previous pipeline versions for the respective assay type, including detailed descriptions and processing instructions as well as the exact processing commands used to generate processed data for datasets hosted in the GeneLab Data Repository.

Assay Types

Click on an assay type below for data processing information.

- Create GeneLab Reference Annotations
- Amplicon Sequencing
  - Illumina
  - 454 and Ion-Torrent
- Metagenomics
  - Removing human reads
  - Illumina
- (bulk) RNAseq
- single cell RNAseq

https://github.com/nasa/GeneLab_Data_Processing

OSDR: Data Visualization

Data Visualization Portal (https://visualization.genelab.nasa.gov/data/)

OSDR: Environmental Data Application (EDA)

Environmental Data Application (EDA)

Environmental Data Application (EDA)

View Biologically Relevant Mission Milestones

Mission Milestone Dates

Display Mission milestones

Launch: 2014-09-21 09:49:00

Reached orbit: 2014-09-23 17:19:00

Animal transfer: 2014-09-25 16:05:00

Sac1: 2014-10-12

Sac2: 2014-10-13

Return: 2014-10-25 19:36:00

Sac3: 2014-10-28 00:01:00

Compare Data From Multiple Missions

Temperature Ground

Copy Excel CSV PDF Print Search:

Mission	Min temperature Ground Degrees [C]	Max temperature Ground Degrees [C]	Mean temperature Ground Degrees [C]	Standard Deviation temperature Ground Degrees [C]	Median temperature Ground Degrees [C]
RR1	18.30	27.10	22.76	0.82	22.60
RR3	20.40	26.10	23.36	0.97	23.70
	Min	Max	Mean	Standard Deviation	Median

Temperature ISS

Copy Excel CSV PDF Print Search:

Mission	Min temperature ISS Degrees [C]	Max temperature ISS Degrees [C]	Mean temperature ISS Degrees [C]	Standard Deviation temperature ISS Degrees [C]	Median temperature ISS Degrees [C]
RR1	18.50	26.80	22.76	0.81	22.60
RR3	18.50	24.80	21.37	0.97	21.70
	Min	Max	Mean	Standard Deviation	Median

RadLab Platform

RadLab Portal and Data API

RadLab Portal and Data API

Time Series Plots

Compare Two Detectors

RadLab Portal and Data API

Geospatial Plots

Maximizing Scientific Return Using Open Science Resources

Principal Investigation

Dissections

Tissue Storage

Tissue Request

Analysis and Publication

PI Results

Submitted & Curated Datasets

Data Reused

Soleus (SOL) Muscle Data:

- Proteomics (quantitative mass spectrometry)
- Transcriptomics (bulk RNA-seq)
- Epigenomics (DNA methylation)
- Calcium Reuptake

Tibialis Anterior (TA) Muscle Data:

- Transcriptomics (bulk RNA-seq)
- Epigenomics (DNA methylation)
- Calcium Reuptake

- Omics & phenotypic data: RR-9, RR-1
- Qlattice xAI (explainable AI) in AWG collaboration
- Li et al., 2023 npj Microgravity 🎉

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

GeneLab for Colleges and Universities (GL4U)

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 Content Design

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

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)

Unix Intro JN

R Intro JN

AmpSeq JN

RNAseq JN

GL4U Approach

GL4U Direct Approach

MSI/HBCU 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
- Training focused on MSIs and HBCUs
- 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

MSI/HBCU Educators

GL4U RNAseq Certification

GL4U Introduction Module
- ☑ Pre-Intro Module Survey
- ☑ Intro Lectures
- ☑ Intro JNs
- ☑ Post-Intro Module Survey

+

GL4U RNAseq Module
- ☑ Pre-RNAseq Module Survey
- ☑ RNAseq Lectures
- ☑ RNAseq JNs
- ☑ 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*

FALL 2024: GL4U ON-DEMAND

GL4U Student & Educator Bootcamps To-Date

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 with MSIs/HBCUs, 06/2022

- Virtual 9-day long bootcamp with 6 professors and 4 graduate students from **4 HBCUs/MSIs**:
  - 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

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

Complete NASA's open science curriculum!

Open Science 101:

A community-developed introduction to **core open science skills**

- Know how to write a NASA open science and data management plan
- Learn about tools and best practices
- Increase the impact & visibility of your science
- Earn your digital NASA open science badge

All 5 modules available through a self-paced online course and through in-person and virtual instructor-led workshops.

Take OS101!

https://go.nasa.gov/40pPQMx

CC BY

TOPS-T *ScienceCore* AI/ML Training – Synchronous

- *ScienceCore*: discipline-specific training materials showcasing open science workflows
- *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_rnaseq
https://registry.opendata.aws/bps_microscopy

canvas.instructure.com/courses/9450036

AI/ML 2024 › Syllabus

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

Account
Dashboard
Courses
Calendar
Inbox
History
Commons
Help

Home
Announcements
Assignments
Discussions
Grades
People
Pages
Files
Syllabus
Outcomes
Rubrics
Quizzes
Modules
BigBlueButton
Collaborations
Attendance
New Analytics
Settings

AI/ML AWG Summer 2024 Training Program

Jump to Today | Edit

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

This class was designed the 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

- *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

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

TOPS-T *ScienceCore* AI/ML Training – Asynchronous

Pre-recorded lectures

Quizzes

Interactive Python notebooks

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

colab

GitHub

Get Involved!

Open Science Analysis Working Groups (AWGs)

Join an AWG!

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

Example of Open Science AWG Collaboration and Publication

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	Mouse	♀	C57BL/6	500.0	~1.5yr sim	PE+1	10											
	MGsim				N/A	14	PE+1	10											
	simGCRsim + MGsim				500.0	~1.5yr sim	PE+1	10											
BNL-2	simGCRsim	Mouse	♀	C57BL/6	500.0	~1.5yr sim	PE+1	10											
	MGsim				N/A	14	PE+1	10											
	simGCRsim + MGsim				500.0	~1.5yr sim	PE+1	10											
BNL-3	simGCRsim	Mouse	♀	C57BL/6	500.0	~1.5yr sim	PE+1	10											
	MGsim				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/6	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											
					13.2	75	L+75	5											
				C57BL/6	4.4	25	L+25	5											
					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?

From Multiple Databases to a Knowledge-Based System

AI ready FAIR data for modeling and risk assessment

Credit: Nelson et al., 2021, https://doi.org/10.3390/life11010042; Genewein et al., 2020, arXiv:2010.12237v1 [cs.AI])

Big Picture for Open Science for Life in Space

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

Acknowledgements

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