National Aeronautics and Space Administration

Introduction to NASA, GeneLab, and the Command Line

Biological and Physical Sciences

Amanda M. Saravia-Butler, Ph.D.
GeneLab Data Processing Lead
NASA Ames Research Center

Introduction to the National Aeronautics and Space Administration (NASA)

A Brief History of NACA

https://www.nasa.gov/ames/the-national-advisory-committee-for-aeronautics

A Brief History of NASA

1957

Soviet Union launches Sputnik 1

1958

'An Act to provide for research into the problems of flight within and outside the Earth's atmosphere, and for other purposes.'

1958

NASA established and absorbs itself into NACA

US launches it's first Earth satellite

1961-3

Mercury's single astronaut program: Can humans survive in space?

1965-6

Project Gemini: Two astronauts practice space operations

1968-72

Project Apollo: Explore the Moon

1981

Space Shuttle's first mission: STS-1

1984

Congress authorized NASA to build the Space Station

2011

Last Space Shuttle Mission: STS-135

NASA's Vision

We reach for new heights and reveal the unknown for the benefit of humankind

NASA's Mission

Drive advances in science, technology, aeronautics, and space exploration to enhance knowledge, education, innovation, economic vitality and stewardship of Earth

https://www.nasa.gov/content/the-crawlers

Cell Science Bioculture System

NASA astronaut Christina Koch

https://www.nasa.gov/cygnss/overview

How does NASA achieve its Mission?

- **NASA HQ** – Provides overall guidance and direction to the Agency
- **NASA Centers and Facilities** – Where the day-to-day NASA work is performed in laboratories, airfields, wind tunnels, etc.
- NASA conducts its work though four main organizations called **Mission Directorates**:
  - Aeronautics Research (ARMD)

  - Human Exploration and Operations (HEOMD)

  - Science (SMD)

  - Space Technology (STMD)

https://www.nasa.gov/offices/olia/overview/index.html

NASA Organization

Note: Administrator may delegate direct reports to Deputy Administrator at his/her discretion.
* NMO oversees the Jet Propulsion Laboratory contract.
** Programmatic reporting to the Science Mission Directorate Associate Administrator.
JPL will participate in Agency-level functions, such as APMC.
JPL is a Federally Funded Research and Development Center (FFRDC).

February 2021
https://www.nasa.gov/about/org_index.html

NASA Leadership

Administrator: Bill Nelson

Sen. Bill Nelson was sworn in as the 14th NASA administrator on May 3, 2021, tasked with carrying out the Biden-Harris administration's vision for the agency. Nelson served in the U.S. Senate for 18 years from Florida and as a payload specialist on space shuttle mission 61-C in 1986. Bill Nelson's full bio: https://www.nasa.gov/feature/nasa-administrator-bill-nelson/

Deputy Administrator: Pamela Melroy? (Nominee)

One of only two women to command a space shuttle, Melroy logged more than 38 days in space. All three of her missions were assembly missions to build the International Space Station. After serving more than two decades in the Air Force and as a NASA astronaut, Melroy took on a number of leadership roles, including at Lockheed Martin, the Federal Aviation Administration, the Defense Advanced Research Projects Agency, Nova Systems Pty, Australia, and as an advisor to the Australian Space Agency. She currently is an independent consultant and a member of the National Space Council's Users Advisory Group. Pam Melroy's full bio: https://www.nasa.gov/sites/default/files/atoms/files/melroy_pamela.pdf

NASA Centers and Facilities

NASA Centers and Facilities

NASA HQ

Ames Research Center

Armstrong Flight Research Center

Glenn Research Center

Goddard Space Flight Center

Jet Propulsion Lab

Johnson Space Center

Kennedy Space Center

Langley Research Center

Marshall Space Flight Center

Stennis Space Center

White Sands Test Facility

https://history.nasa.gov/

What has NASA done for us?

A Workforce Innovating To Go Farther

Space Launch System/Multi-Purpose Crew Vehicle
NASA is innovating to continue American leadership in an ambitious space program that takes us farther than ever before. The Agency is supporting technology innovation and commercial partnerships that fuel the American economy. With the development of the newest generation of space transportation, the Space Launch System (SLS) and the Orion Multi-Purpose Crew Vehicle, NASA is developing capabilities to access deep space destinations like the Moon, an asteroid, and Mars.

This capability is being built right here in the United States. Thousands of workers in over 30 states are currently building our Nation's next generation of space travel.

An exploration flight test of Orion will take place in 2014 with a follow-on integrated Orion/SLS uncrewed flight in 2017.

A crewed flight of the integrated deep space system will occur as early as 2021.

Game-Changing Innovation

International Space Station—A Laboratory for Life Improvement Today
The United States has an active astronaut program. In fact, American human space flight occurs daily on board the International Space Station (ISS).

NASA is committed to maintaining American excellence in science, technology, engineering, and mathematics (STEM) while sustaining a long-term human presence in space. The ISS provides important technology and innovation development, not only for future long-term human space missions, but to enhance life on Earth today.

Outstanding advances in microbial vaccine development, unprecedented cancer research involving treatment delivery, and improved water purification technologies are just some of the technological advances that research on board the ISS has provided to life here on Earth.

Science
NASA is committed to groundbreaking science missions today and in the future. Our missions focus on providing critical data about our home planet, unraveling the mysteries of our universe, understanding solar variability and the impacts of space weather, and exploring the diverse planetary bodies of our solar system.

Aeronautics
Aeronautics research is leading the Nation to less congested airways and a more environmentally friendly aviation industry.

Investment in American Technology
By pushing the boundaries of technology, NASA's programs keep American business and workers on the cutting edge. Investing in technology development today enables future missions of NASA while growing the economy by creating new industries, jobs, products, and services. To ensure access to the most innovative tech development, NASA partners with American companies throughout the Nation in all 50 states.

$12.5 Billion in FY 2011 Obligated to American Industry

Inspiring the Next Generation

Over the past year, NASA's K–12 education projects reached more than 1 million students through STEM programs and initiatives. The Summer of Innovation Program engaged over 45,000 students and over 5,500 educators through camps and activities in 46 states, the District of Columbia, and Puerto Rico.

$941 Million in FY 2011 Obligated to Educational Institutions Throughout the Country

Investing in a New Aerospace Economy

Commercial Partners
NASA is committed to developing a new aerospace economy. The Agency is currently partnering with American industry to achieve safe and reliable access to low-Earth orbit and the ISS.

In fact, private industry began test flights in 2010, and these will continue through 2012. This progress will culminate with human access to the ISS by 2017.

Building a strong commercial space industry enables the development of a capability that will assure the Nation's future in space. Doing so will increase national revenue while positioning American companies as competitors in the global marketplace in the space economy.

NASA Technologies Americans Use Daily

TV Satellite Dish
NASA developed ways to correct errors in the signals coming from spacecraft. This technology is used to reduce noise in TV signals coming from satellites.

Medical Imaging
NASA developed ways to process signals from spacecraft to produce clearer images. This technology also makes photolike images of the insides of our bodies possible.

Vision Screening System
This technology uses techniques developed for processing space images to examine the eyes of children and find out quickly if they have any vision problems. The child doesn't have to say a word!

Ear Thermometer
Instead of measuring temperature using a column of mercury (which expands as it heats up), this thermometer has a lens like a camera and detects infrared energy, which we feel as heat. The warmer something is (like your body), the more infrared energy it puts out. This technology was originally developed to detect the birth of stars.

Firefighter Equipment
Firefighters wear suits made of fire-resistant fabric developed for use in spacesuits.

Smoke Detector
These devices were first used in the Earth-orbiting space station called Skylab (launched back in 1973) to help detect toxic vapors. Now they are used in most homes and other buildings to warn people of fire.

Automobile Design Tools
A computer program developed by NASA to analyze a spacecraft or airplane design and predict how parts will perform is now used to help design automobiles.

Thermal Gloves and Boots
Thermal gloves and boots have heating elements that run on rechargeable batteries worn on the inside wrist of the glove or embedded in the sole of the ski boot. This technology was adapted from a spacesuit design for the Apollo astronauts.

Advanced Plastics
Spacecraft and other electronics need very special, low-cost materials as the base for printed circuits (like those inside your computer). Some of these "liquid crystal polymers" have turned out to be very good, low-cost materials for making containers for food and beverages.

National Aeronautics and Space Administration
Headquarters
Office of Legislative and Intergovernmental Affairs
300 E Street SW
Washington, DC 20546-0001
Phone: 202-358-1055
Fax: 202-358-4340
http://www.nasa.gov/offices/olia

What's in store for NASA?

1. Extend and sustain human activities across the solar system.

2. Expand scientific understanding of the Earth and the universe in which we live.

3. Create the innovative new space technologies for our exploration, science, and economic future.

4. Advance aeronautics research for societal benefit.

5. Enable program and institutional capabilities to conduct NASA's aeronautics and space activities.

6. Share NASA with the public, educators, and students to provide opportunities to participate in our mission, foster innovation and contribute to a strong national economy.

NASA SMD Science by the Numbers

NASA Science

Earth Science
The study of planet Earth

Planetary Science
the study of the origin and history of the solar system and the potential for extraterrestrial life

Astrophysics
The study of the origin, structure, evolution and destiny of the universe

Heliophysics
The study of the sun and its effects on the solar system

5+ Decades of NASA Life Sciences Research

Exposure to Spaceflight Changes our Physiological Systems

The Musculoskeletal System

Osteoporosis
Osteoarthritis
Muscle atrophy

The Cardiovascular System

CVD
Atherosclerosis
Anemia

The Lymphatics

Lymphatic pump disorders

The Central Nervous System

Memory loss
Retinal edema

The Reproductive System

Sperm motility
Infertility
Hormonal imbalances

The Digestive System

Liver fibrosis
NAFLD
IBS
Microbiome

Exposure to Spaceflight Changes Plant Physiology Too

Space Biology Program

https://science.nasa.gov/biological-physical/programs/space-biology

Spaceflight Hardware: Cell Culture Systems

Bioculture System

CellCult

Spaceflight Hardware: Microbiology

Biological Research in Canisters (BRIC)

Fig. 1 BRIC-60 and BRIC-60M with petri dish configuration.

Fig. 2 BRIC-60 containing

Fig. 3 BRIC-60 Canisters as flown

Spaceflight Hardware: Small Animals

Multi-use Variable gravity Platform (MVP)

Spaceflight Hardware: Plants

Advanced Plant Habitat

APH Flight Unit in Ground Support Equipment

Dwarf Wheat Growing in APH

Spaceflight Hardware: Plants

Veggie Habitat

Veggie Configured for Growth of Lettuce

Veggie LED light bank.

Veggie plant pillow with quick disconnect fitting for adding water.

Veggie with 6 plant pillows contained within adjustable bellows.

Astronaut Steve Swanson harvesting lettuce grown on ISS.

Astronauts Kjell Lindgren (left) and Scott Kelly (right) enjoying freshly harvested lettuce grown on ISS.

Astronaut Steve Swanson basking in the glow of Veggie.

Spaceflight Hardware: Rodent Research

Animal Access Unit (AAU)

RR Habitat

Space Biology Program

https://science.nasa.gov/biological-physical/programs/space-biology

The Future of Space Exploration

https://www.nasa.gov/about/whats_next.html | https://www.youtube.com/watch?v=mB1nAzrigRQ

Introduction to GeneLab

2011 NRC Decadal Survey and the Sequencing Paradigm Shift

*"…**genomics, transcriptomics, proteomics, and metabolomics** offer an immense opportunity to understand the effects of spaceflight on biological systems…"*

*"…Such techniques generate considerable amounts of **data that can be mined and analyzed** for information by multiple researchers…"*

What are ‘omics?

Beyond the Central Dogma

Gonzalez-Pardo and Alvarez, 2013

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

GeneLab: Open Science for Life in Space (https://genelab.nasa.gov)

GeneLab Open Science for Life in Space

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

Keywords

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

Data Repository
Search and upload spaceflight datasets

Analyze Data
Perform large-scale analysis of biological omics data

Environmental Data
Radiation data collected during experiments conducted in space

Collaborative Workspace
Share, organize and store files

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

Visualize Data
Interact with GeneLab processed data

GeneLab ecosystem: Maximizing knowledge by bringing experiments together

Identify Shared Processes/ Molecular Signatures

- Hypoxic Response ?
- Oxidative Stress
- Common Tissue (e.g. muscle, liver, heart, eyes, brain,…)

For Spaceflight

- High "n" number – statistically significant data
- Genetically identical animals
- Low resource requirements
- Short life cycle - multiple generations
- Measure response of a whole multicellular animal
- **Flies used as a model for humans for innate immunity, circadian rhythm, oxidative stress, neurobehavior, development, genetics, GWAS, "omics" studies etc**

GeneLab Data Systems

GeneLab: Submission Portal – GEODE (https://genelab.nasa.gov)

GeneLab Environment for Online Data Entry (GEODE)

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
Sample3	Mus musculus	X-5	liver	Sample collection	-80	degree Celsius	Iron-56 Ion Irradiation	24	hour	2	gray
Sample4	Mus musculus	X-5	liver	Sample collection	-80	degree Celsius	Silicon	12	hour	3	gray
Sample5	Mus musculus	X-5	liver	Sample collection	-80	degree Celsius	Sham-irradiated	0	hour	0	gray

Sample level data

Terms are connected to a controlled ontology

- Space Biology PIs can now submit data through GEODE.
- Allows efficient input of sample and assay level metadata alongside data files.
- Tailored to specialized needs of spaceflight samples.
- Now exploring extending this portal for use on ALSDA submissions.

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

Assay level data with fields specific to that assay

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

GeneLab: Data Repository (https://genelab.nasa.gov)

GeneLab Open Science for Life in Space

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

Search Data

All | GeneLab | NIH GEO | EBI PRIDE | ANL MG-RAST

Search Filters (GeneLab Only)

Project Type | Factors | Organisms | Assay Type | Clear

Show Only:

Studies With Visualizations

Page 1 of 12 (Total Studies: 278) Next >

Studies Per Page: 25

GLDS-322

Comparative RNA-Seq transcriptome analyses reveal dynamic time dependent effects of 56Fe, 16O, and 28Si irradiation on the induction of murine hepatocellular carcinoma

Organisms	Factors	Assay Types	Release Date	Description
Mus musculus	Ionizing Radiation Time	transcription profiling	27-Aug-2020	One of the health risks posed to astronauts during deep space flights is exposure to high charge, high-energy (HZE) ions (Z>13) which can lead to induction of hepatocellular carcinoma (HCC). We perfor...

GLDS-320

Gamma radiation and HZE treatment of seedlings in Arabidopsis

Organisms	Factors	Assay Types	Release Date	Description
Arabidopsis thaliana	Genotype Time of sample collection after treatment Ionizing Radiation	transcription profiling	18-Sep-2014	Plants exhibit a robust transcriptional response to gamma radiation which includes the induction of transcripts required for homologous recombination and the suppression of transcripts that promote ce...

GeneLab: Data Repository

GLDS-137

Version 4

- DESCRIPTION
- PROTOCOLS
- SAMPLES
- ASSAYS
- PUBLICATIONS
- STUDY FILES
- VISUALIZATION

Search Data

GLDS-137: Rodent Research-3-CASIS: Mouse liver transcriptomic, proteomic, and epigenomic data Version 4

Select a Version: 4

DOI: 10.26030/9k6w-4c28
Source Accession Number
Total Data Volume: 2.5 TB

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

DESCRIPTION

Study 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, 18-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.

Contacts

NAME	ROLE	ORGANIZATION	E-MAIL
Ruth Globus	Project Scientist	NASA Ames Research Center	ruth.k.globus@nasa.gov
Jonathan Galazka	Project Scientist	NASA GeneLab	jonathan.m.galazka@nasa.gov
Rosamund Smith	Principal Investigator	Eli Lilly and Company	

GLDS-137

Version 4

DESCRIPTION

PROTOCOLS

Mission

MISSION NAME	START DATE	END DATE
SpaceX-8	08-Apr-2016	11-May-2016

Acknowledgments

Funding for sample processing and sequencing was provided to the GeneLab project by the NASA Space Biology Program Office, Space Life and Physical Sciences Research and Applications Division. Samples from the RR-3 experiment were provided to GeneLab through the Biospecimen Sharing Plan of the NASA Life Sciences Data Archive.

Data Citation

Download Citation (BibTex)

Download Citation (RIS)

Smith R, Cramer M, Globus R, Galazka J. "Rodent Research-3-CASIS: Mouse liver transcriptomic, proteomic, and epigenomic data", GeneLab, Version 4, http://doi.org/10.26030/9k6w-4c28

GLDS-137

Version 4

DESCRIPTION

PROTOCOLS

SAMPLES

ASSAYS

PUBLICATIONS

STUDY FILES

VISUALIZATION

Select Export Columns

Select columns of the samples panel below to export

Export CSV

- Source Name
- Sample Name
- Protocol REF
- Comment: Source description
- Characteristics: Organism
- Characteristics: Strain
- Comment: Animal Source
- Characteristics: sex
- Characteristics: Age at Launch
- Characteristics: Diet
- Comment: Feeding Schedule
- Parameter Value: Habitat
- Factor Value: Spaceflight
- Parameter Value: treatment
- Parameter Value: Duration
- Parameter Value: Euthanasia Chemical
- Parameter Value: Carcass Weight
- Parameter Value: Source Storage Temperature
- Comment: Initiation Date
- Comment: Completion Date
- Comment: Dissection date
- Protocol REF
- Comment: Source description
- Material Type
- Comment: Original Submitted Sample Name
- Comment: Sample Name Description
- Comment: LSDA Biospecimen Source ID
- Comment: LSDA Biospecimen Subject ID
- Parameter Value: Sample Preservation Method
- Parameter Value: Sample Storage Temperature

close select all unselect all

SAMPLES

Source Name	Sample Name	Protocol REF	Comment: Source description	Characteristics: Organism	Characteristics: Strain	Comment: Animal Source
CB1	Mmus_BAL-TAL_LVR_BSL_Rep1_B1	mouse habitation	previously dissected carcass	Mus musculus	BALB/c	Taconic Biosciences
CB2	Mmus_BAL-TAL_LVR_BSL_Rep2_B2	mouse habitation	previously dissected carcass	Mus musculus	BALB/c	Taconic Biosciences

GeneLab: Data Repository

STUDY FILES

VISUALIZATION

STUDY FILES

To view files, click on the folder of interest.

4 files selected

Download Selected Files

- All Files
  - RNA-Seq Data
  - Proteomics Data Files
  - Whole Genome Bisulfite Sequencing Data
  - GeneLab Processed RNA-Seq Files
    - Aligned sequence data
    - Trimmed sequence data/FastQC reports
    - Merged sequence data/FastQC reports
    - Aligned sequence data/Alignment logs
    - Raw counts data
    - Trimmed sequence data/MultiQC report
    - Differential expression analysis data
    - Normalized counts data
    - Merged sequence data/MultiQC report
    - Trimmed sequence data
    - Merged sequence data
    - Trimmed sequence data/Trimming reports
  - Study Metadata Files

	FILES	FILE SIZE	RESOURCE CATEGORY	RESOURCE DESCRIPTION
☑	GLDS-137_rna_seq_Mmus_BAL-TAL_LVR_FLT_Rep5_F5_SJ.out.tab	7.26 MB	Aligned sequence data	Data processed by the GeneLab Team, providing users high order data using standardized methods.
☑	GLDS-137_rna_seq_Mmus_BAL-TAL_LVR_BSL_Rep4_B4_R1_trimmed_fastqc.zip	433.13 KB	Trimmed sequence data/FastQC reports	Data processed by the GeneLab Team, providing users high order data using standardized methods.
☐	GLDS-137_rna_seq_Mmus_BAL-TAL_LVR_GC_Rep4_G5_R2_raw_fastqc.zip	456.51 KB	Merged sequence data/FastQC reports	Data processed by the GeneLab Team, providing users high order data using standardized methods.
☑	GLDS-137_rna_seq_Mmus_BAL-TAL_LVR_FLT_Rep2_F2_R2_raw_fastqc.html	636.29 KB	Merged sequence data/FastQC reports	Data processed by the GeneLab Team, providing users high order data using standardized methods.
☑	GLDS-137_rna_seq_Mmus_BAL-TAL_LVR_GC_Rep3_G3_R1_trimmed_fastqc.zip	434.84 KB	Trimmed sequence data/FastQC reports	Data processed by the GeneLab Team, providing users high order data using standardized methods.
☐	GLDS-137_rna_seq_Mmus_BAL-TAL_LVR_GC_Rep1_G1_R1_raw_fastqc.zip	449.76 KB	Merged sequence data/FastQC reports	Data processed by the GeneLab Team, providing users high order data using standardized methods.

GeneLab: Data Repository

302
Studies

45
Species

>10
Assays

>130TB
Data

Experiment environment

Assay type

Organism

GeneLab: Environmental Data (https://genelab.nasa.gov)

GeneLab Open Science for Life in Space

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

Keywords

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

Data Repository
Search and upload spaceflight datasets

Analyze Data
Perform large-scale analysis of biological omics data

Environmental Data
Radiation data collected during experiment conducted in space

Collaborative Workspace
Share, organize and store files

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

Visualize Data
Interact with GeneLab processed data

GeneLab: Environmental Data

Users can view environmental data relevant to a dataset.

This includes temperature, relative humidity, CO2, radiation exposure.

Users can compare conditions of different datasets.

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

GeneLab Open Science for Life in Space

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

Keywords

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

Data Repository
Search and upload spaceflight datasets

Analyze Data
Perform large-scale analysis of biological omics data

Environmental Data
Radiation data collected during experiments conducted in space

Collaborative Workspace
Share, organize and store files

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

Visualize Data
Interact with GeneLab processed data

GeneLab: Data Visualization Portal

GeneLab Data Democratization

GeneLab User Growth

Top 10 Countries – Sessions

1. United States
2. India
3. United Kingdom
4. Germany
5. Canada
6. France
7. South Korea
8. Japan
9. China
10. Australia

Usage: May 2017- Jan 24, 2021

3494 Avg Monthly Sessions

1718 Avg Monthly Users

Downloads: Jan 2019 - Dec 2020

**represents downloads per GLDS Study, not per file*

405 Avg Monthly Downloads

10146 Total Study Downloads

GeneLab Power Users

GeneLab Analysis Working Groups (AWGs) *consist of* ***130+ scientists*** *from multiple space agencies, international institutions, and industry. Scientists meet monthly with each group to analyze data in the GeneLab repository. Majority of members are non-NASA PI's – many have applied for NASA funding following AWG interactions.*
https://genelab.nasa.gov/awg/join

AWG Members represent:
- 48 US Universities
- 4 NASA Centers
- 4 Other Government-funded Organizations
- 3 Institutes or Private Industry
- 3 International Universities

Educational Working Group (EWG) *consists of educators and scientists spanning across high school and college, focusing on developing content to learn about spaceflight OMICS, promoting the use of GeneLab's database via the repository, and the visualization and analytical portals.*

GeneLab for High Schools (GL4HS)
GeneLab for College/University (GL4U – Pilot with SJSU)

SJSU SAN JOSÉ STATE UNIVERSITY

GeneLab included in Bioinformatics curriculum of degree granting university

GL4HS (Tutorial + Tools)

GeneLab Interns
Undergrads, Grads, Postdoc

Annual AWG Workshop

Analysis Working Groups (AWGs) are Driving Data Reuse

Publications Using GeneLab Data

Publication Metrics

50
Original Publication linked to GeneLab

30
Derived Publication linked to GeneLab

60+
Presentations linked to GeneLab

92
Datasets used in derived publications

ROI grows faster than publications linked to original

Original – Publications with original dataset submitted to GeneLab
Derived – Publications using data submitted to GeneLab repository

Most Recent Publications

A special collection on **the biology of spaceflight**, published on November 25th, 2020 in *Cell* and other Cell Press journals, done in collaboration between NASA and other space agencies around the world, uncovers the impact of known hazards of spaceflight, such as radiation and microgravity, and discusses the standards for multi-omics from space and the preparations needed for Mars and other missions in the next two decades.

- 29 scientific papers, 200+ authors
- **9 papers utilize data in GeneLab or GeneLab resources**
- 2 AWG papers
- *Covered by 200 News Articles worldwide with 43 articles mentioning GeneLab*

Cell Bundle - Highlights

Highlights

- Multi-omics analysis and techniques with NASA's GeneLab platform.
- Mitochondrial dysregulation driving spaceflight health risks.
- NASA Twin Study data validates mitochondrial dysfunction during space missions.

Largest amount of astronaut data in one paper!

Highlights

- Comparative transcriptomics in *C. elegans* exposed to hypergravity and spaceflight
- Bioinformatics identifies novel putative regulators of altered gravitational load
- Candidate molecules infer a close gravity > *daf-16*/FOXO > neuronal link

Highlights

- Data integration between GeneLab and ALSDA furthers the ability to utilize technologies such as artificial intelligence (AI) and machine learning (ML) methodologies to develop new hypotheses.
- Database standards and accessibility such as FAIR principles and APIs enables open-access science promoting collaborative efforts to interpret spaceflight effects by integrating omics and physiological data to the systems level.

Highlights

- Experimentation is necessary to understand how organisms respond to space
- RNAseq datasets are impacted by preservation protocols used on the ISS
- Impacts can be alleviated with improved carcass preservation protocols

- ISSOP is an international consortium of scientists
- Optimizing the conditions for scientists and the general public to derive valid hypotheses from these precious space omics data by reducing confounding factors and increasing interoperability at the global level.

https://www.cell.com/c/the-biology-of-spaceflight

GeneLab Teams

Introduction to the Command Line

What is a Terminal?

- A terminal window is a command-line interface used to interact with your computer (without graphics)
- The terminal window displays a blank text window, or “command line”, where you can type commands to communicate with your computer
- A terminal uses a specific language interpreter, or “shell”, to interpret commands typed into the command line
- Unix is a computer operating system written in the C programming language
  - By default, terminals on Linux and Mac computers run a Unix shell
  - Unix shells come in a variety of flavors including zsh, csh, tcsh, and bash, each having slightly different features and functionality
  - We will be using the Unix shell, bash, during this bootcamp

Command Syntax

- General command line syntax: command arguments
- Depending on the command used, arguments can be optional or required
  - Note: Arguments are also referred to as “options”, “parameters”, or “flags”
- All available arguments for a command can be found by looking at the command’s manual or help page, which can be opened in the terminal.
  - To view a command’s manual, we use the `man` function in Unix followed by the command of interest
    - Example: man <command_of_interest>
  - To view a command’s help page, we use the `--help` or `-h` flag for the command of interest
    - Example: <command_of_interest> --help
- Below is the manual of a common Unix command, mkdir, which was displayed in a terminal by running the command: man mkdir

```
asaravia — less • man mkdir — 192×27

MKDIR(1)                BSD General Commands Manual                MKDIR(1)

NAME
     mkdir -- make directories

SYNOPSIS
     mkdir [-pv] [-m mode] directory name ...

DESCRIPTION
     The mkdir utility creates the directories named as operands, in the order specified, using mode rwxrwxrwx (0777) as modified by the current umask(2).

     The options are as follows:

     -m mode
             Set the file permission bits of the final created directory to the specified mode.  The mode argument can be in any of the formats specified to the chmod(1)
             command.  If a symbolic mode is specified, the operation characters ``+'' and ``-'' are interpreted relative to an initial mode of ``a=rwx''.

     -p      Create intermediate directories as required.  If this option is not specified, the full path prefix of each operand must already exist.  On the other hand,
             with this option specified, no error will be reported if a directory given as an operand already exists.  Intermediate directories are created with permission
             bits of rwxrwxrwx (0777) as modified by the current umask, plus write and search permission for the owner.

     -v      Be verbose when creating directories, listing them as they are created.

     The user must have write permission in the parent directory.

:
```

Running Unix Commands

- The SYNOPSIS section of the mkdir manual shows how the command should be run:

```
SYNOPSIS
     mkdir [-pv] [-m mode] directory name ...
```

- For the mkdir command, `mkdir` is written first, followed by the flags indicated in the brackets, and ending with the directory name.
  - Note: Anything in brackets [] is optional and any part of the command not in brackets is required.
  - For the mkdir command, which parts are optional and which are required?
- Do you notice how the different parts of the command are separated? **With spaces!**
- On the command line, spaces are special
  - Spaces are used to know how to properly separate distinct parts of a command
  - For this reason, it's not ideal for file or directory (aka folder) names to contain spaces. Thus, it's better to use dashes (-) or underscores (_) instead of spaces when naming files and directories.
    - Example: "example_test1.txt" is preferred over "example test1.txt"

Running Unix Commands

- Before we run the mkdir command to make a new directory, we first need to decide where to make the new directory.
- On computers, files are organized into directories (aka folders).
- When you open a terminal, it opens to a particular directory on your system, referred to as your “working (aka current) directory”.
- You can see what your working directory is by executing the `pwd` command, which stands for “print working directory”
- In the example below, you’ll see the output of the `pwd` command:

```
(base) asaravia@ARLAL0121011036 ~ % pwd
/Users/asaravia
```

- What is the current (aka working) directory?
- You can list what’s in your working directory with the `ls` command as shown on the left below (note: `ls` stands for “list”) :

```
(base) asaravia@ARLAL0121011036 ~ % ls
Accellion       Documents       Music           igv
Applications    Downloads       Pictures        temp_files
COV-IRT         GLDS-121        Public
COV-IRT_tmp     Library         anaconda3
Desktop         Movies          gsea_home
```

- We can also navigate to the /Users/asaravia directory using the Finder window as shown on the right above.
  - Do you notice any similarities between the files shown in /Users/asaravia on the command line versus the Finder window?

Running Unix Commands

- Similar to how you can navigate to a different directory (aka folder) within your current directory by double clicking on the folder icon in the Finder window, you can use the `cd` command to change directories in your terminal (note: `cd` stands for “change directory”).
- Below, we will navigate to the Desktop using the command line (left) and in Finder (right):

```
(base) asaravia@ARLAL0121011036 ~ % cd Desktop
(base) asaravia@ARLAL0121011036 Desktop % pwd
/Users/asaravia/Desktop
```

- What is the working directory now?
- How are directories distinguished on the command line?

- Let’s see what’s on the Desktop using the `ls` command in the terminal below:

```
(base) asaravia@ARLAL0121011036 Desktop % ls
GLDS-104                        Miscellaneous
Google_Drive_Overload           Sat Feb 20 22-19-12 2021
IACUC_RR-10_correspondence      annotation_DBI_packages
Job_related_CS_challanges       druva
Logyx_NASA                      openEDITS12.fmp12
Microarray_Datasets             stuff_to_transfer
```

- Compare the content on the Desktop in the terminal with what is shown in Finder

Running Unix Commands

- Now that we've set the Desktop as our working directory, let's use the `mkdir` command to create a new folder on the Desktop.
- Recall the SYNOPSIS section of the mkdir manual, which shows how the command should be run:

```
SYNOPSIS
     mkdir [-pv] [-m mode] directory name ...
```

- In the example below, we'll make a directory titled "myDirectory" on the Desktop using the command line then view what's on the Desktop using both the command line and Finder:

```
(base) asaravia@ARLAL0121011036 Desktop % mkdir myDirectory
```

```
(base) asaravia@ARLAL0121011036 Desktop % ls
GLDS-104                        Sat Feb 20 22-19-12 2021
Google_Drive_Overload           annotation_DBI_packages
IACUC_RR-10_correspondence      druva
Job_related_CS_challanges       myDirectory
Logyx_NASA                      openEDI1512.tmp12
Microarray_Datasets             stuff_to_transfer
Miscellaneous
```

Note: The `mkdir` command is the equivalent of creating a new folder in Finder.

Creating Variables

- There will be times when we need to set our working directory to a folder that is deep within a directory structure. Rather than navigating one directory at a time, like you would in Finder, you can navigate directly to your folder of interest with a single `cd` command in the terminal by writing the explicit path to that directory as shown:

```
(base) asaravia@ARLAL0121011036 ~ % pwd
/Users/asaravia
(base) asaravia@ARLAL0121011036 ~ % cd /Users/asaravia/Desktop/myDirectory
(base) asaravia@ARLAL0121011036 myDirectory % pwd
/Users/asaravia/Desktop/myDirectory
```

- If there is a location that we need to visit often, rather than writing out the explicit path every time, we can create a variable containing the explicit path. In Unix, you can do this using an equal sign to assign the path to a variable name as shown:

```
(base) asaravia@ARLAL0121011036 ~ % pwd
/Users/asaravia
(base) asaravia@ARLAL0121011036 ~ % myDir=/Users/asaravia/Desktop/myDirectory
(base) asaravia@ARLAL0121011036 ~ % echo $myDir
/Users/asaravia/Desktop/myDirectory
(base) asaravia@ARLAL0121011036 ~ % cd $myDir
(base) asaravia@ARLAL0121011036 myDirectory % pwd
/Users/asaravia/Desktop/myDirectory
```

- What command was introduced in the example above?
- What was added to the variable name when using the variable in Unix?

Creating Variable Names

- When creating variable names there are a few import things to keep in mind:
  - Do not name your variable after a common function or a variable built-in to the coding language being used
  - Avoid using spaces and any characters other than letters, numbers, dashes, and underscores – Why?
  - Choose informative, memorable, and short variable names

- Example: You want to create a variable to store gene expression data from Drosophila melanogaster (aka fruit flies) that were grown at 10°C, 20°C, or 30°C. Which of the names below would be good variable names to use?

  1. GeneExpressionData_Drosophilamelanogaster_DifferentTemperatures
  2. Dmel_GE_dT
  3. FruitFlies_GE_dT
  4. FruitFlies_GE_10°C_ 20°C_ 30°C
  5. echo
  6. myData
  7. DmelGEdT
  8. Dmelanogaster GE Data

Jupyter Notebook

- Throughout this bootcamp, we will be running several of the commands we just discussed. However, since working directly on the command line takes time to get used to, we will be using Jupyter Notebooks to run our commands.
- Jupyter Notebook (JN) was created as part of Project Jupyter (https://jupyter.org/) and is a web application that allows you to create documents that have both narrative text and live code; meaning you can run commands through a web browser (instead of a terminal).
- Additional advantages of JNs:
  - ➢ Facilitates sharing code and results between groups
  - ➢ Ease of publishing your code
  - ➢ Allows for reproducibility

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 relationship of the adjusted p-value significant by adjusted p-value are labeled.

First, we'll use the default settings from the `EnhancedVolcano()` function: log2 Fold Change cutoff > |2|, and the adjusted p-value cutoff is < 10e-6.

You can read about `EnhancedVolcano()` and see some examples by clicking here

```
# Volcano plot showing genes differentially expressed in FLT vs GC
EnhancedVolcano(DGE_output_table,
    lab = DGE_output_table$SYMBOL,
    x = 'Log2fc_(FLT)v(GC)',
    y = 'Adj.p.value_(FLT)v(GC)',
    title = 'FLT versus GC',
    pCutoff = 10e-6,
    FCcutoff = 2,
    pointSize = 3.0,
    labSize = 6.0,
    colAlpha=0.5)

## Save your volcano plot
ggsave(file.path(DGE_plots,'GLDS-104_volcano_DGE.png'), width = 6, height = 8.5, dpi = 300)
```

Compute Cluster

- When we execute commands in our JNs, they will be run on the SJSU compute cluster.
- A compute cluster is a set of servers, which are like computers, that are all connected, and each server is used for a specific purpose. Below is an example of a relatively simple cluster containing 4 servers (aka nodes) – 1 head node and 3 compute nodes:

- **Head node**: The server that allows you to connect to a cluster (aka the login server). You should not run any commands that require a lot of computational resources on the head node. Why do you think that is?
- **Compute nodes**: Servers that are used to run commands, programs, and scripts (aka jobs). Compute nodes usually contain more computational resources than the head node.

Compute Cluster

- There are often multiple people logged into a cluster at the same time, and they all want to run jobs. This could be a problem for the servers in the cluster. Why?
  - Hint: Think about what happens when you try to run a bunch of applications on your personal computer. Now imagine multiple people each running several applications on a server…
- To avoid multiple people running jobs all at the same time and risking freezing the servers, compute clusters are equipped with a job scheduler.
- Rather than executing jobs directly, users submit their jobs to the job scheduler, which will check each compute node to see if there are enough resources available to run the job. If there are, the job will begin running on one of the compute nodes; if not, the job will get queued and will not start until enough resources are available.
- During this bootcamp, we will be submitting all computationally expensive jobs to the SJSU cluster job scheduler, SLURM.

Questions?