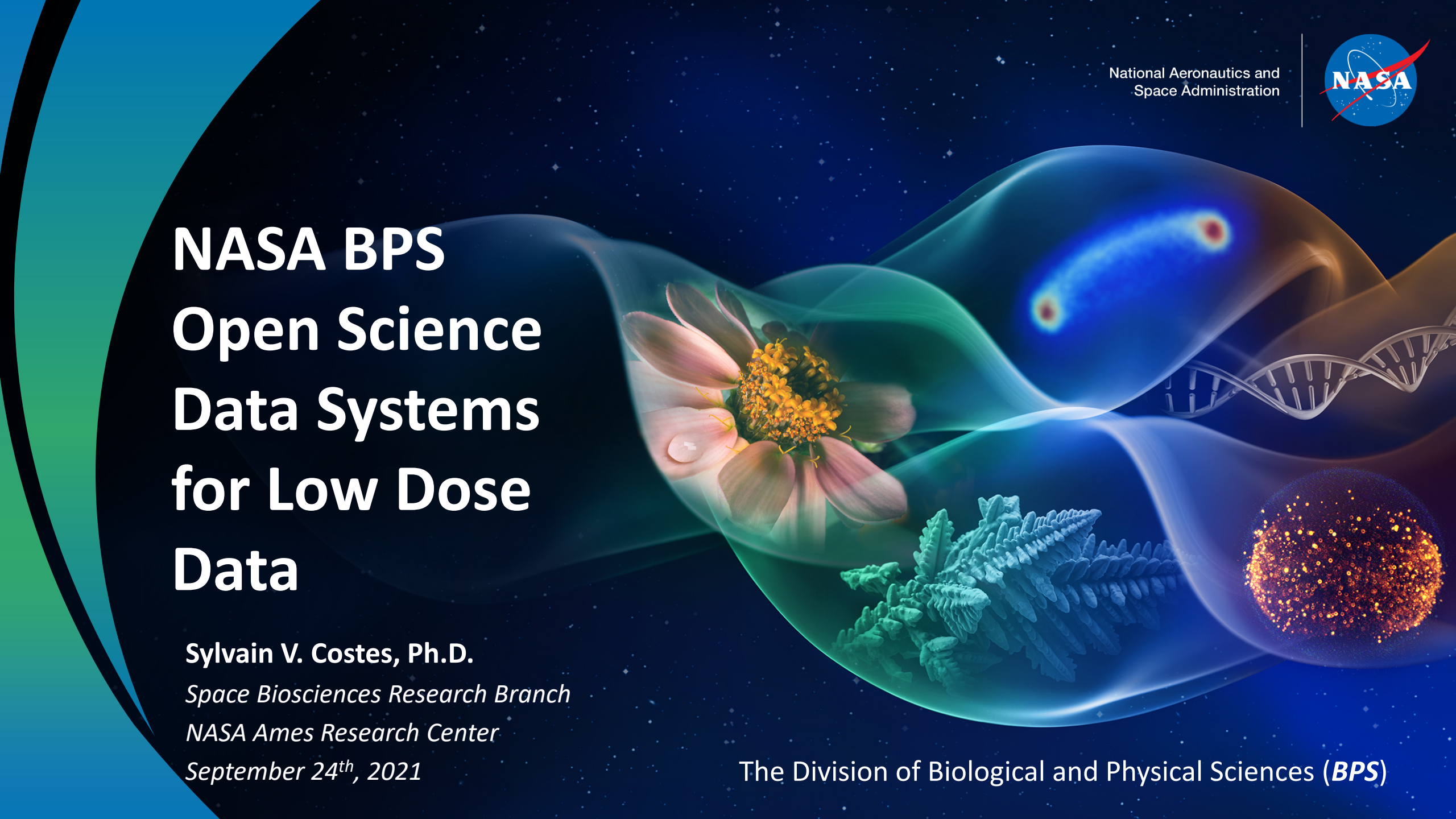
National Aeronautics and Space Administration



NASA BPS Open Science Data Systems for Low Dose Data

Sylvain V. Costes, Ph.D.
Space Biosciences Research Branch
NASA Ames Research Center
September 24th, 2021

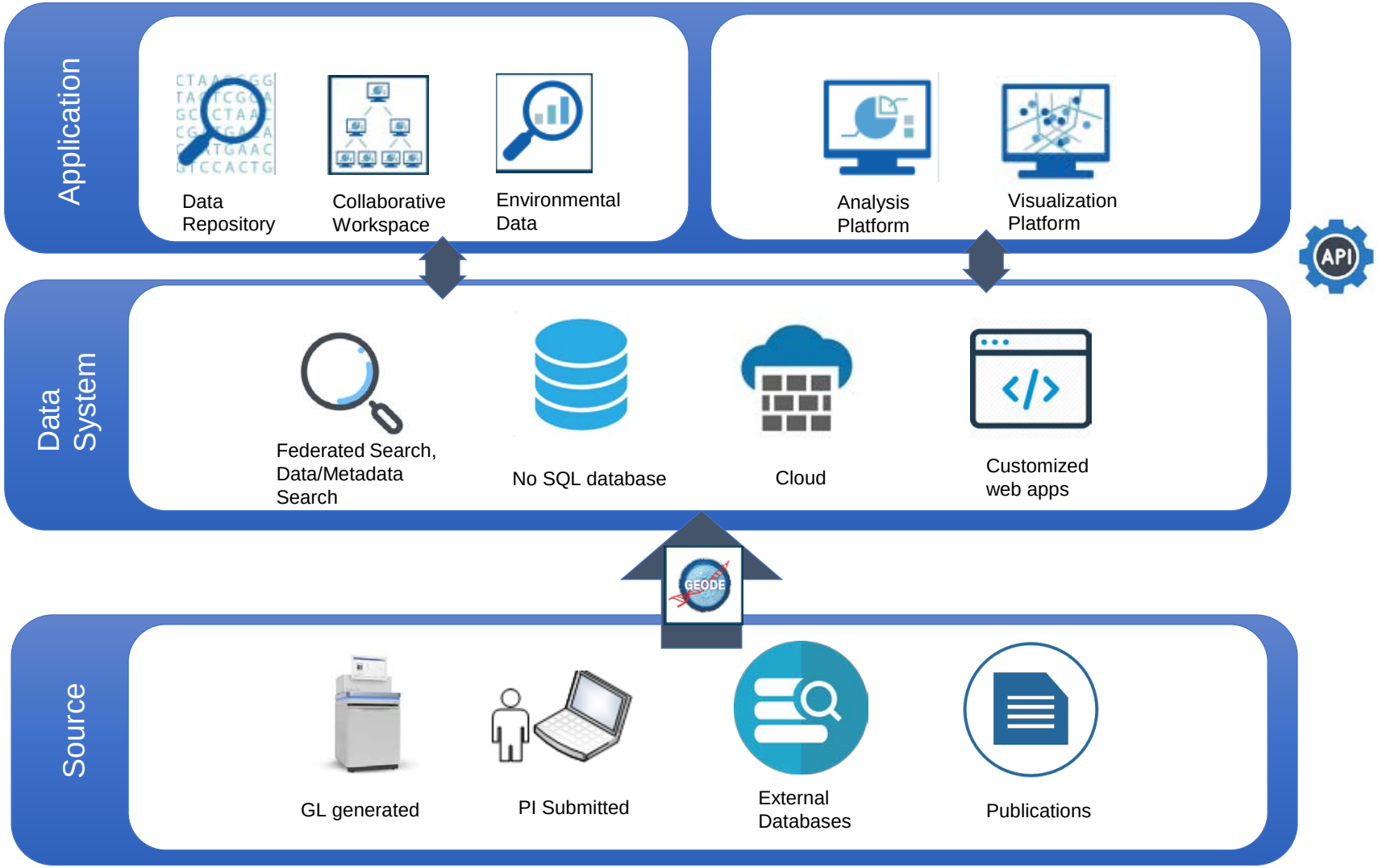
The Division of Biological and Physical Sciences (***BPS***)

NASA BPS Open Science Enterprise Solution: genelab.nasa.gov

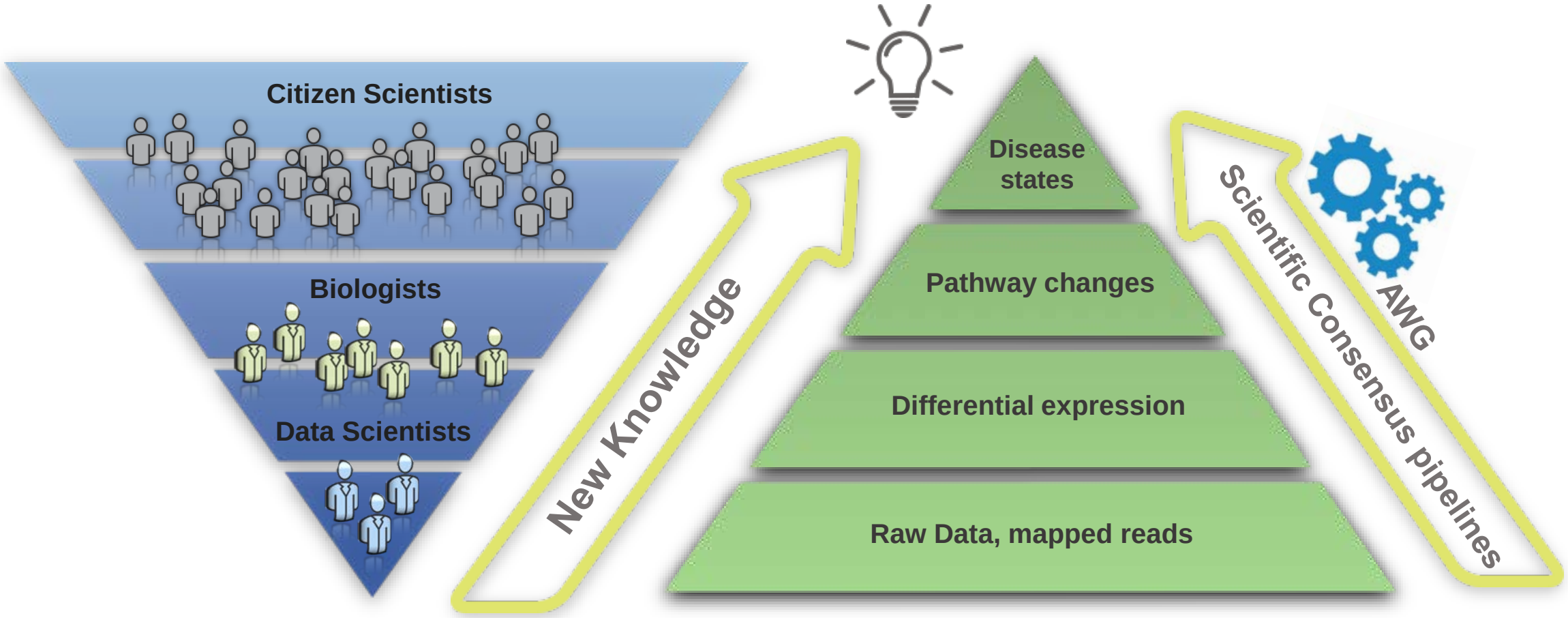
- Open access data
- **FAIR (Findable, Accessible, Interoperable, Reusable)**
- Controlled access tools
- API - internal and external
- User Friendly Interface
- Tutorials
- Self-service Submission Portal

- Federated search – GEO, PRIDE, MG-RAST, **ALSDA**
- Database & Cloud – Scalable, easy access, fast
- Web apps
  - Data Access & Management
  - Security
  - Operation
  - Governance and Integration
- Open Source software – *no maintenance cost for software*

- Multiple data sources
  - Standard metadata organization
  - Open file formats



GeneLab Omics Data Democratization



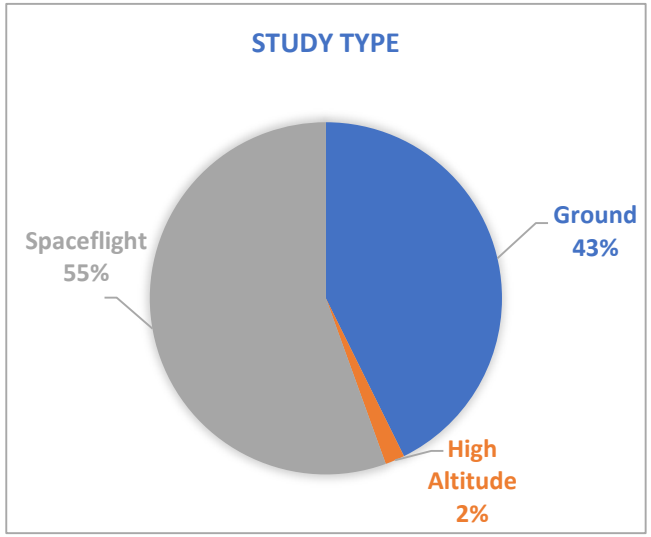
335
Studies

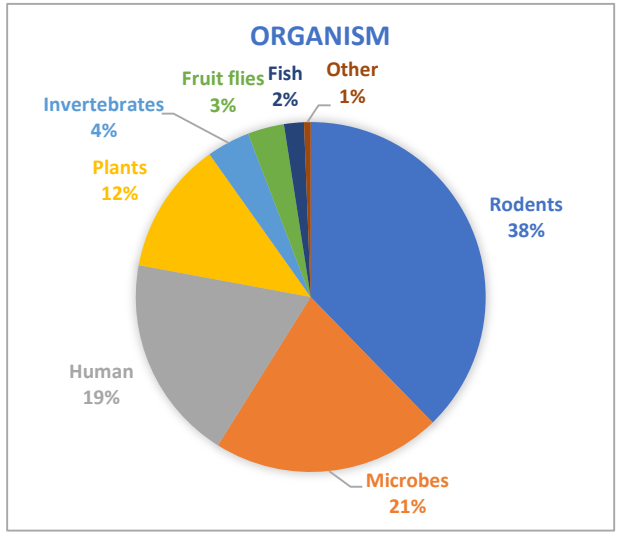
371
Datasets

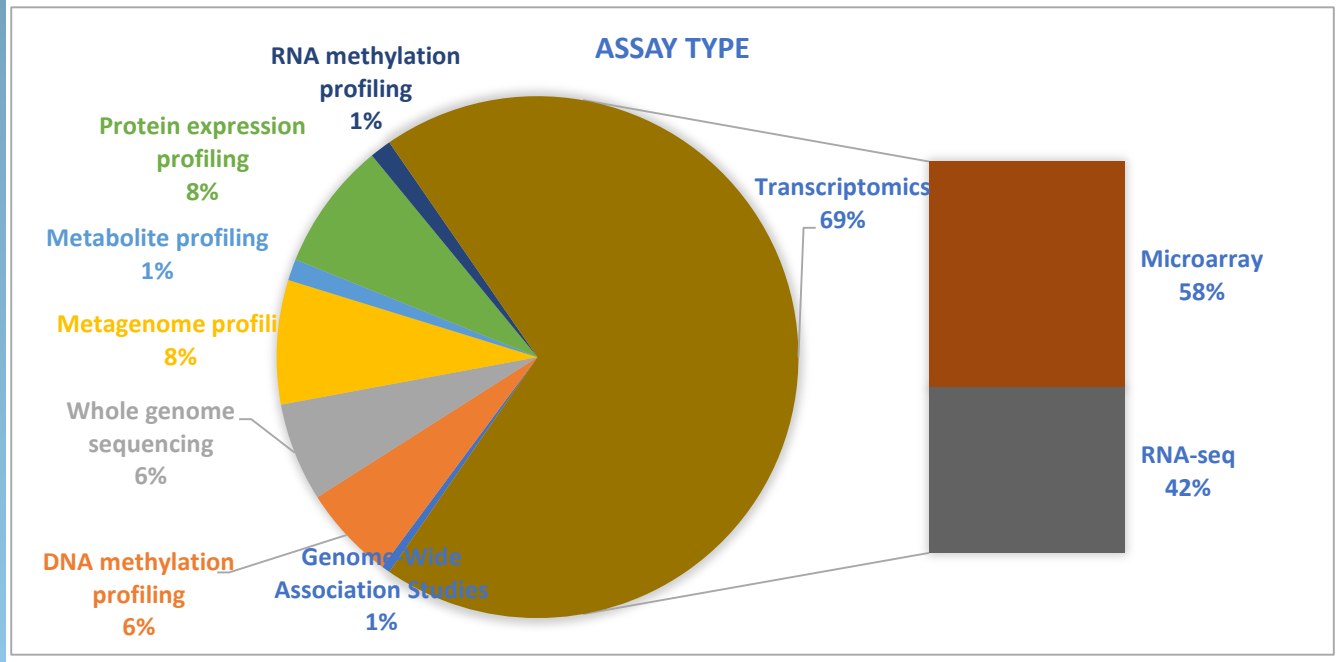
45
Species

>10
Assays

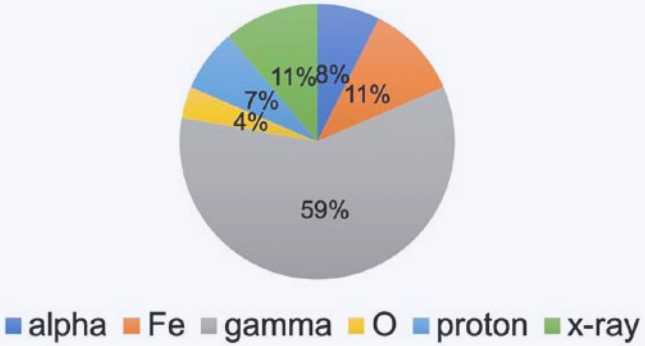
>135TB
Data

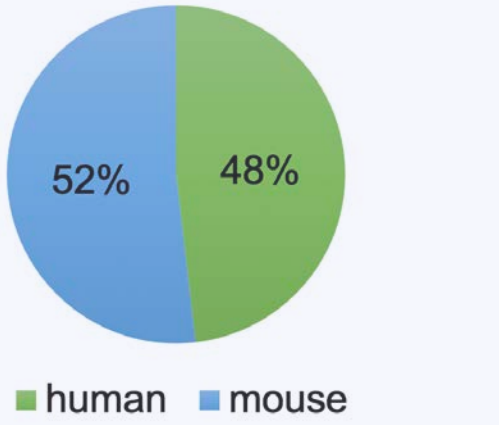




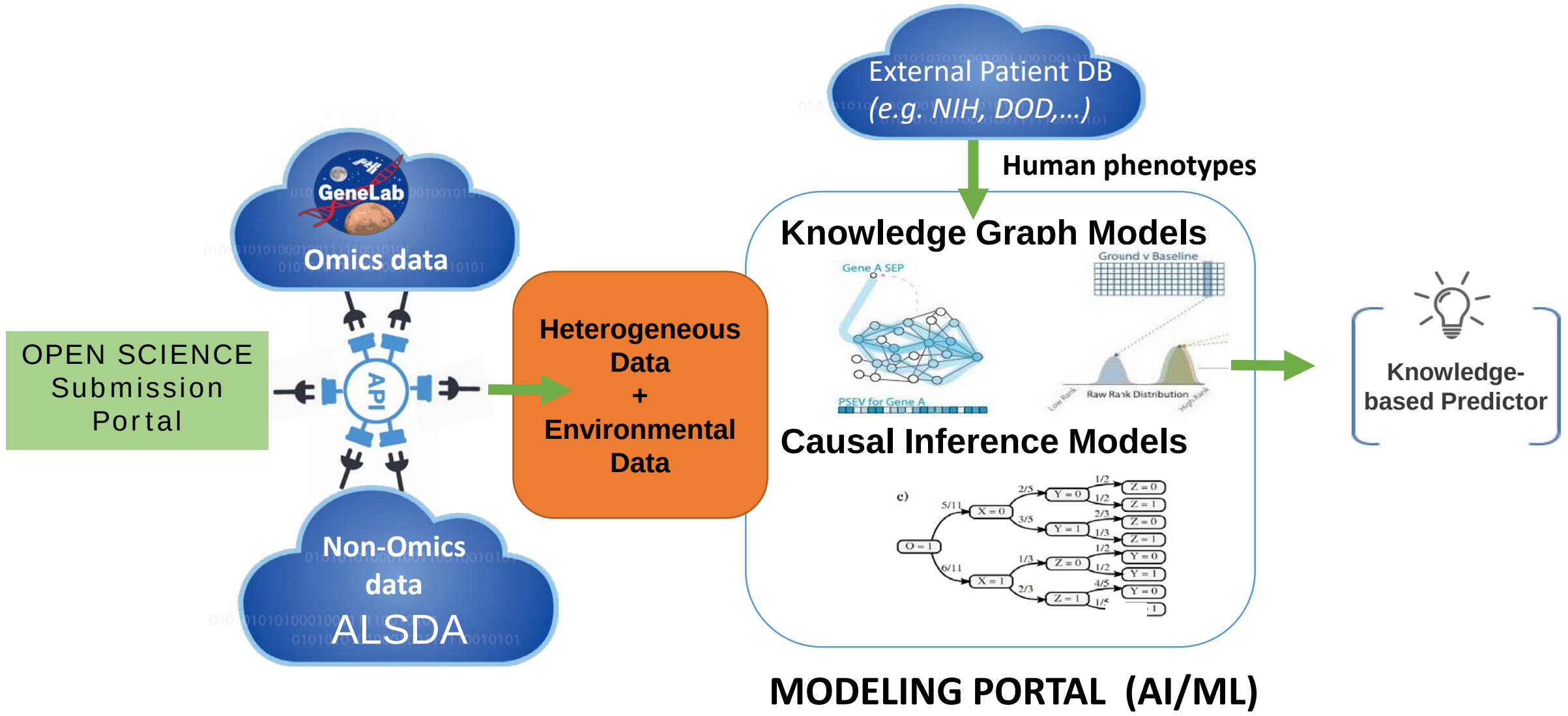


27 LOW DOSE RADIATION STUDIES IN GENELAB DATABASE





From multiple databases to a knowledge-based system



Open Science for Life in Space

genelab.nasa.gov

Data Portal

- First multi-omics space related database
- 335 publicly available studies
- High-order data for transcriptomics and metagenomics studies
- Spaceflight **environmental data** associated with sample metadata (radiation dosimetry, temperature, humidity)
- High-order Data **visualization Portal**

Tools

- User friendly submission portal to submit and publish data
- Analysis platform to tools and workflows to analyze your own data or data from the GeneLab repository.
- Tutorials and online resources to learn how to analyze RNA-Seq data
- Workspace to store, share, and organize data files

Sequencing

- State-of-the-art sequencing facility to process spaceflight samples
- Optimized SOPs with standard processing workflow
- Provide sequencing service for NASA funded PI without any university overhead
- Generate high quality data from shared tissues for open science access

Community

- Analysis Working Groups comprised of over 100+ scientists worldwide collaborating and analyze space omics data.
- Education Working Group focused on providing resources to educators and students to learn about space biology and bioinformatics

Low dose data are highly relevant for NASA space missions

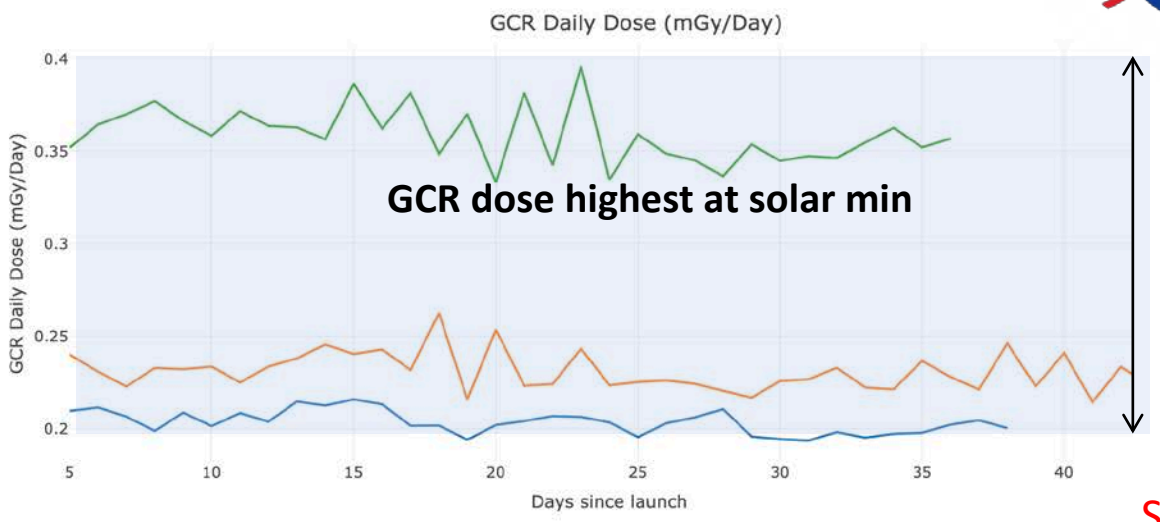
Application

Environmental Data

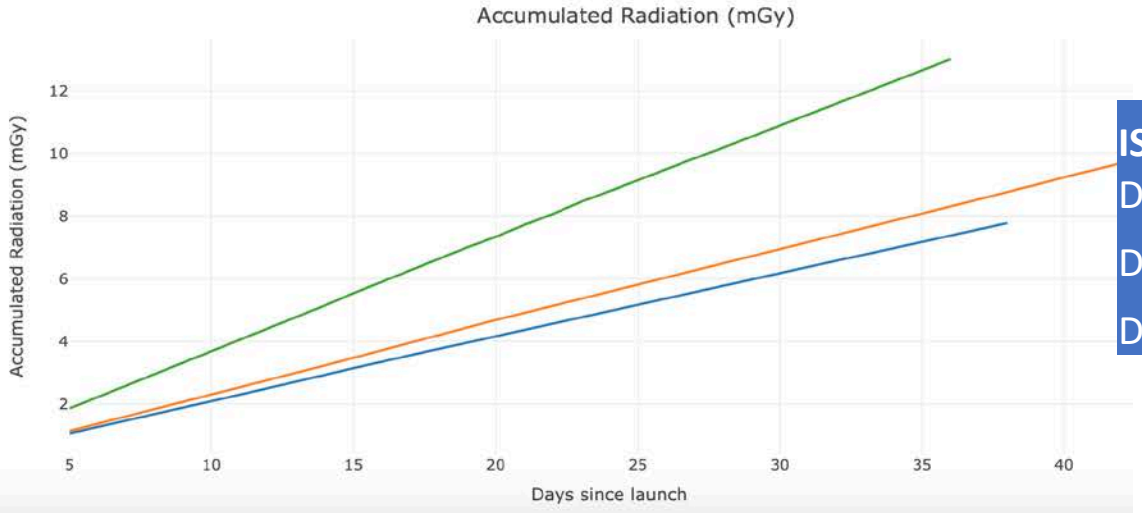
Mission	Date
RR1	2014-09-20 03:59:00
RR3	2016-04-08 04:00:00
RR5	2017-06-01 05:00:00
RR6	2017-12-06 15:06:00
RR7	2018-07-01 04:49:00
RR12	2019-04-17 19:19:00
RR19	0019-12-03 10:24:00

Select plots to display
- [x] Timeplot
- [ ] Boxplot
- [ ] Bar Chart
- [ ] Show landmarks

RR1
RR3
RR12



0.2-0.4 cGy/day



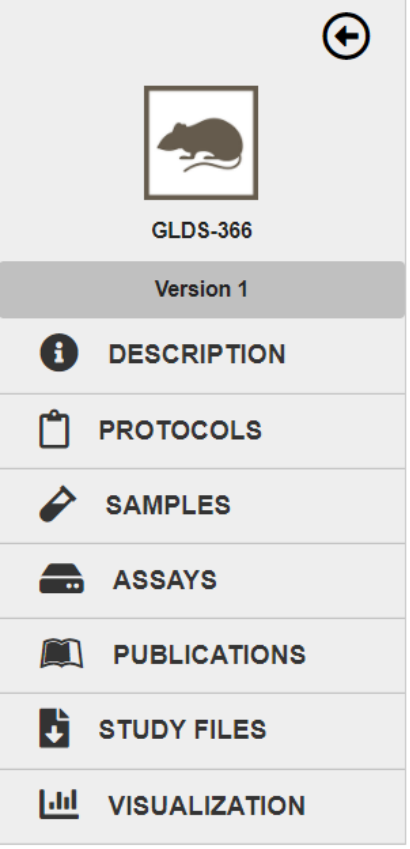
Space flight data:
Low dose rate: 0.2-0.4 cGy/day
& Low dose: <10 cGy

ISS	Mouse	Astronauts
Duration	30 days	180 days
Dose solar min	**1.2 cGy**	**7.2 cGy**
Dose solar max	**0.6 cGy**	**3.6 cGy**

HRP and DOE Low dose funded research – All dataset are assigned a DOI (FAIR database)

Application

Data Repository



Search Data

GLDS-366: Coalescence of DNA double strand breaks induced by galactic cosmic radiation is modulated by genetics in 15 inbred strains of mice

Version 1

Select a Version: 1

DOI: 10.26030/v8w4-rg83
Source Accession Number(s)
Total Data Volume: 64.2 MB

Submitted Date: 29-Sep-2020
Release Date: 19-Aug-2021

DESCRIPTION

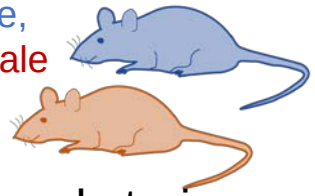
Study Description

This study analyzes the variability of responses to simulated deep space radiation among 15 commonly used mouse strains. Ex vivo primary skin fibroblast responses to high mass-high charge particles and X-rays were analyzed by quantifying DNA damage-sensing protein 53BP1 positive radiation-induced foci (RIF) as a surrogate biomarker of DNA double strand breaks (DSBs). Primary skin fibroblasts were isolated from 10 collaborative cross strains and five reference inbred mice (C57Bl/6, BALB/CByJ, B6C3, C3H and CBA/CaJ) and exposed to 350 MeV/n Ar and 600 MeV/n Fe particles as well as X-rays. Our results indicate that nearby DSBs coalesced into repair units characterized by large RIFs. Such model has the advantage of being much more efficient molecularly, but is poorly suited to deal with cosmic radiation, where energy is concentrated along the particle trajectory. Thus, we observed a large density of DSBs along each particle track and the percentage of unrepaired DSBs that increased with linear energy transfer of the particle over 48 hours post irradiation. Furthermore, persistent RIF levels ex vivo were well correlated with T and B lymphocyte survival in vivo in 10 collaborative cross strains 24 hours after 0.1 Gy whole-body dose of X-rays, suggesting that persistent RIFs might serve as an ex vivo biomarker for in vivo radiation toxicity. Finally, we performed genome-wide association study to identify the genomic associations with dose responses to ionizing radiation, marked as Foci per Gray (FPG), as well as with background DNA repair levels (BGD). This dataset includes GWAS data as well as FPG and BGD values for each sample and condition.

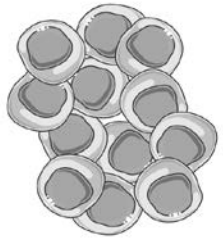
Mouse variability and genomic associations with radiation responses

Low dose animal model developed by Lawrence Berkeley National Laboratory Low dose program

3 male,
3 female
each



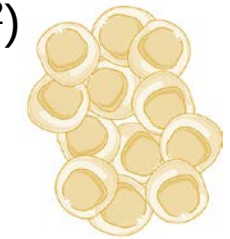
5 inbred strains
10 Collaborative Cross strains



Cultured fibroblasts

2 fluences
(1 and 3 particles/100 um^2)

- ^{56}Fe (600 MeV/n)
- ^{40}Ar (350 MeV/n)
- **Gamma rays, 0.1 Gy & 1 Gy**

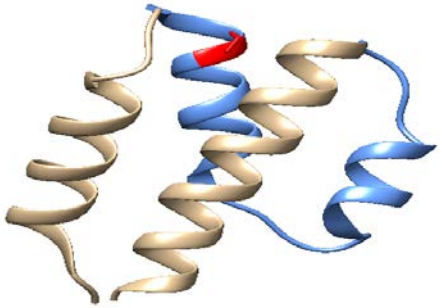


GWAS

4h & 24h post irradiation

- DNA DSB – 53BP1+ foci

Target genes



Predicted protein-level impact

Different strains for GWAS are easily visible in sample tables

Application

Data Repository

GLDS-366

Version 1

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

SAMPLES

Source Name	Sample Name	Characteristics: Organism	Factor Value: Strain	Characteristics: sex	Characteristics: Material Type	P
B6C3F1	B6C3F1_P235_F4_B6C3...	Mus musculus	B6C3 Mouse	female	primary cultured fibroblast cell	growt
BALBCF1	BALBCF1_P235_A8_BAL...	Mus musculus	BALB/cByJ	female	primary cultured fibroblast cell	growt
C3HF1	C3HF1_P235_B4_C3H_F...	Mus musculus	C3H/HeMsNsrf	female	primary cultured fibroblast cell	growt
C57BLF1	C57BLF1_P235_A6_C57...	Mus musculus	C57BL/6J	female	primary cultured fibroblast cell	growt
CBAF1	CBAF1_P235_E2_CBA_F...	Mus musculus	CBA/CaJ	female	primary cultured fibroblast cell	growt
B6C3F1	B6C3F1_P234_F4_B6C3...	Mus musculus	B6C3 Mouse	female	primary cultured fibroblast cell	growt

Assays are selectable

Application

Data Repository

GLDS-366

Version 1

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

Select Export Columns

ASSAYS/MEASUREMENTS

Assay Name: Genome-wide Association Study

Technology Platform: MegaMUGA

Technology Ty

Genome-wide Association Study

histology

Sample Name	Protocol REF	Parameter Value: QA Instrument	Parameter Value: Platform	Extract Name	Array Data File	Pr
B6C3	Genotyping	NanoDrop 2000 UV-vis spectrophotometer	MegaMouse Universal Genotyping Array (MegaMUGA platform)	B6C3	GLDS-366_SNP_Merged_GENO…	data t
BALBC	Genotyping	NanoDrop 2000 UV-vis spectrophotometer	MegaMouse Universal Genotyping Array (MegaMUGA platform)	BALBC	GLDS-366_SNP_Merged_GENO…	data t
C3H	Genotyping	NanoDrop 2000 UV-vis spectrophotometer	MegaMouse Universal Genotyping Array (MegaMUGA platform)	C3H	GLDS-366_SNP_Merged_GENO…	data t
C57	Genotyping	NanoDrop 2000 UV-vis spectrophotometer	MegaMouse Universal Genotyping Array (MegaMUGA platform)	C57	GLDS-366_SNP_Merged_GENO…	data t
CBA		NanoDrop 2000 UV-vis	MegaMouse Universal	CBA	GLDS-	

ALSDA non-omics histology data are linked to GeneLab via the assay selection

Application

Data Repository

GLDS-366

Version 1

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

Select Export Columns

ASSAYS/MEASUREMENTS

Assay Name: histology

Technology Platform:

Technology Ty ...py assay

Genome-wide Association Study

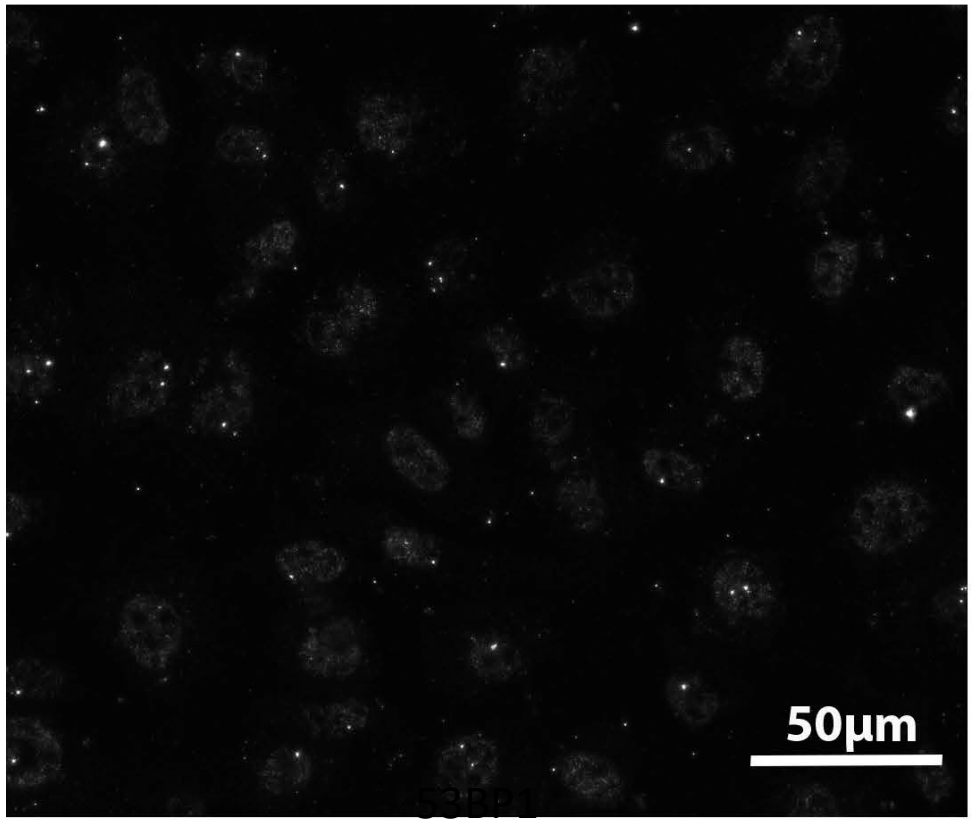
histology

Sample Name	Protocol REF	Parameter Value: primary antibody	Comment: Dilution factor	Comment: Incubation time	Parameter Value: secondary antibody	Di
B6C3F1_P235_F4_B6C3...	immunostaining and imaging	rabbit polyclonal anti-53BP1	1 to 400	1 hour	Alexa Fluor 488 goat anti-rabbit	1 to 4
B6C3F2_P235_E3_B6C3...	immunostaining and imaging	rabbit polyclonal anti-53BP1	1 to 400	1 hour	Alexa Fluor 488 goat anti-rabbit	1 to 4
B6C3F3_P235_F3_B6C3...	immunostaining and imaging	rabbit polyclonal anti-53BP1	1 to 400	1 hour	Alexa Fluor 488 goat anti-rabbit	1 to 4
B6C3M1_P235_C3_B6C3...	immunostaining and imaging	rabbit polyclonal anti-53BP1	1 to 400	1 hour	Alexa Fluor 488 goat anti-rabbit	1 to 4
B6C3M2_P235_D4_B6C3...	immunostaining and imaging	rabbit polyclonal anti-53BP1	1 to 400	1 hour	Alexa Fluor 488 goat anti-rabbit	1 to 4
B6C3M3_P235_D3_B6C3...	immunostaining and imaging	rabbit polyclonal anti-53BP1	1 to 400	1 hour	Alexa Fluor 488 goat anti-rabbit	1 to 4

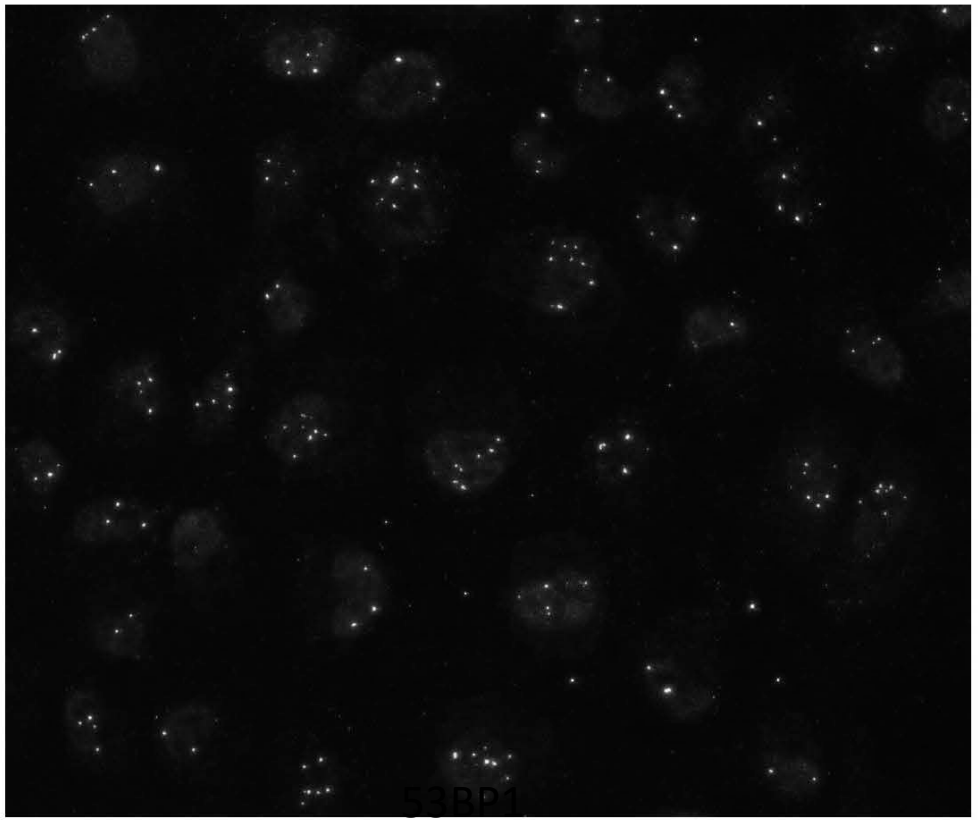
Images can be accessed via ALSDA (work in progress)

Application

Data Repository



0 cGy, 4 hours post-IR, 360 KV Xray



10 cGy, 4 hours post-IR, 360 KV Xray

All peer-reviewed publications associated to dataset are listed

Application

Data Repository

GLDS-366

Version 1

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

PUBLICATIONS

Considering Cell Proliferation to Optimize Detection of Radiation-Induced 53BP1-Positive Foci in 15 Mouse Strains Ex Vivo
Authors: Sebastien Penninckx, Eloise Pariset, Ana Uriarte Acuna, Stephane Lucas, Sylvain V. Costes
PubMed ID: 33181852
DOI: 10.1667/RADE-20-00165.1

Dose, LET and Strain Dependence of Radiation-Induced 53BP1 Foci in 15 Mouse Strains Ex Vivo Introducing Novel DNA Damage Metrics
Authors: Sebastien Penninckx, Egle Cekanaviciute, Charlotte Degorre, Elodie Guiet, Louise Viger, Stephane Lucas, Sylvain V. Costes
PubMed ID: 31081741
DOI: 10.1667/RR15338.1

53BP1 Repair Kinetics for Prediction of In Vivo Radiation Susceptibility in 15 Mouse Strains
Authors: Eloise Pariset, Sebastien Penninckx, Charlotte Degorre Kerbaul, Elodie Guiet, Alejandra Lopez Macha, Egle Cekanaviciute, Antoine M Snijders, Jian-Hua Mao, Francois Paris, Sylvain V Costes
PubMed ID: 32991727
DOI: 10.1667/RADE-20-00122.1

STUDY FILES

To view files, click on the folder of interest.

0 files selected

Download Selected Files

- All Files
  - Genome Wide Association Study Data Files
  - Histology Data Files
  - SNP array
  - Study Metadata Files

FILES	FILE SIZE	RESOURCE CATEGORY	RESOURCE DESCRIPTION
GLDS-366_GWAS_processed_associations.csv	53.82 MB	Processed data	Compressed collection of processed data files and quality report associated with this study. Formats are platform specific.

Data file manager allows access to all levels of data (from raw to processed)

Application

Data Repository

GLDS-366

Version 1

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

STUDY FILES

To view files, click on the folder of interest.

0 files selected

Download Selected Files

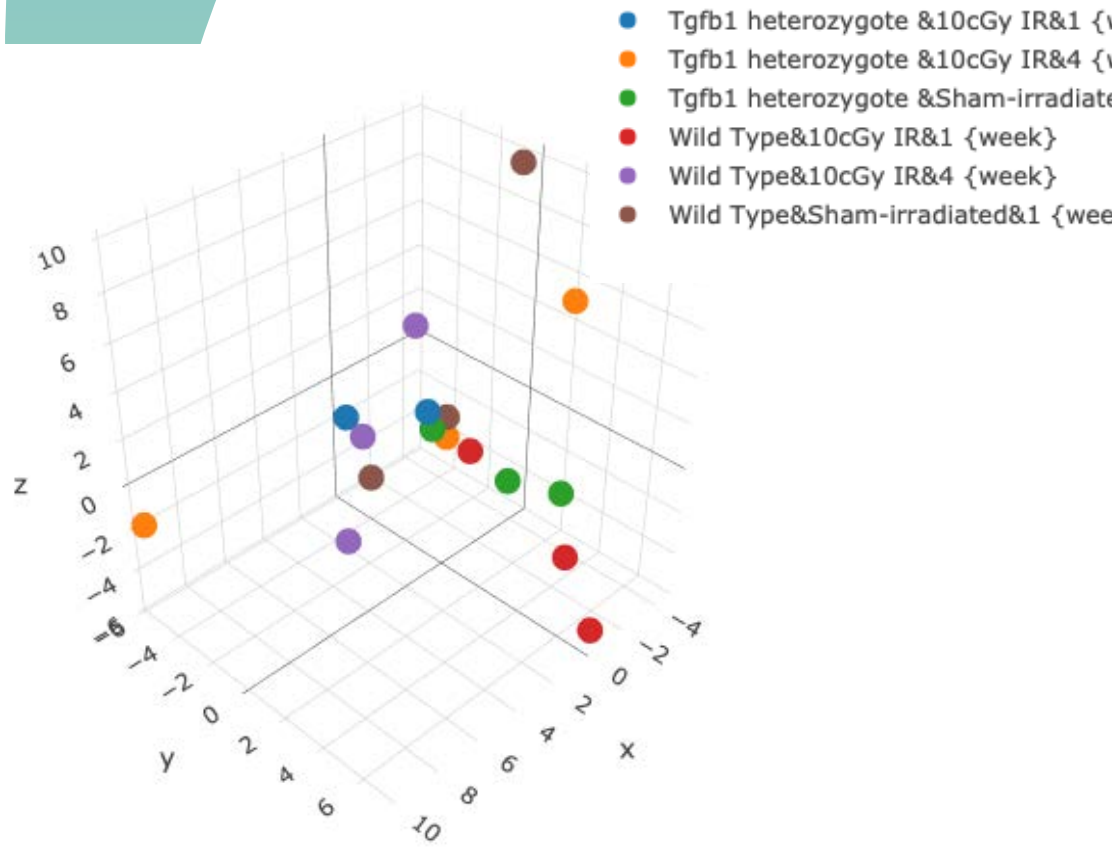
- All Files
  - Genome Wide Association Study Data Files
    - Processed data
  - Histology Data Files
    - Raw Data Files
    - Processed Data Files
    - Raw Data Files/Supplemental Materials
  - SNP array
    - Raw Data Files
  - Study Metadata Files

FILES	FILE SIZE	RESOURCE CATEGORY	RESOURCE DESCRIPTION
GLDS-366_GWAS_processed_associations.csv	53.82 MB	Processed data	Compressed collection of processed data files and quality report associated with this study. Formats are platform specific.
GLDS-366_Histology_raw_pheno_V3.csv	1.02 MB	Raw Data Files	raw or processed data files and quality report associated with this study. Formats are platform specific.
GLDS-366_Histology_processed_pheno_V2.csv	251.08 KB	Processed Data Files	raw or processed data files and quality report associated with this study. Formats are platform specific.
GLDS-366_Histology_Phenotypes_description.txt	6.51 KB	Raw Data Files/Supplemental Materials	raw or processed data files and quality report associated with this study. Formats are platform specific.
GLDS-366_SNP_Merged_GENO.txt	8.8 MB	Raw Data Files	An oligonucleotide DNA microarray used to detect polymorphisms in DNA samples.

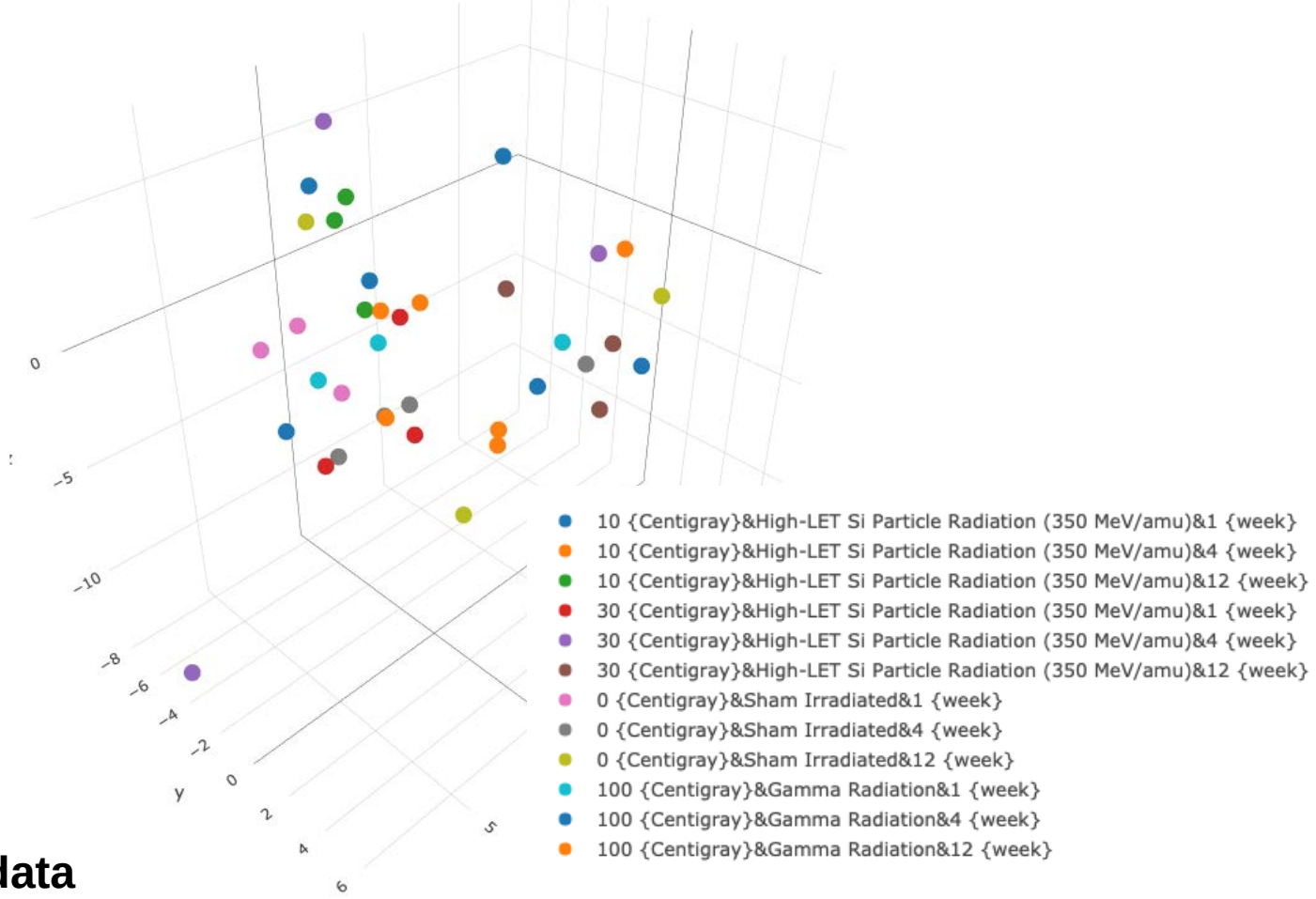
Low dose radiation transcriptomic data can be visualized

Visualization Platform

GLDS-153: Non-targeted effects of low dose ionizing radiation act via TGF-beta to promote mammary carcinogenesis



GLDS-80: Response of mammary tissue to high-LET HZE particle radiation or low-LET gamma-rays

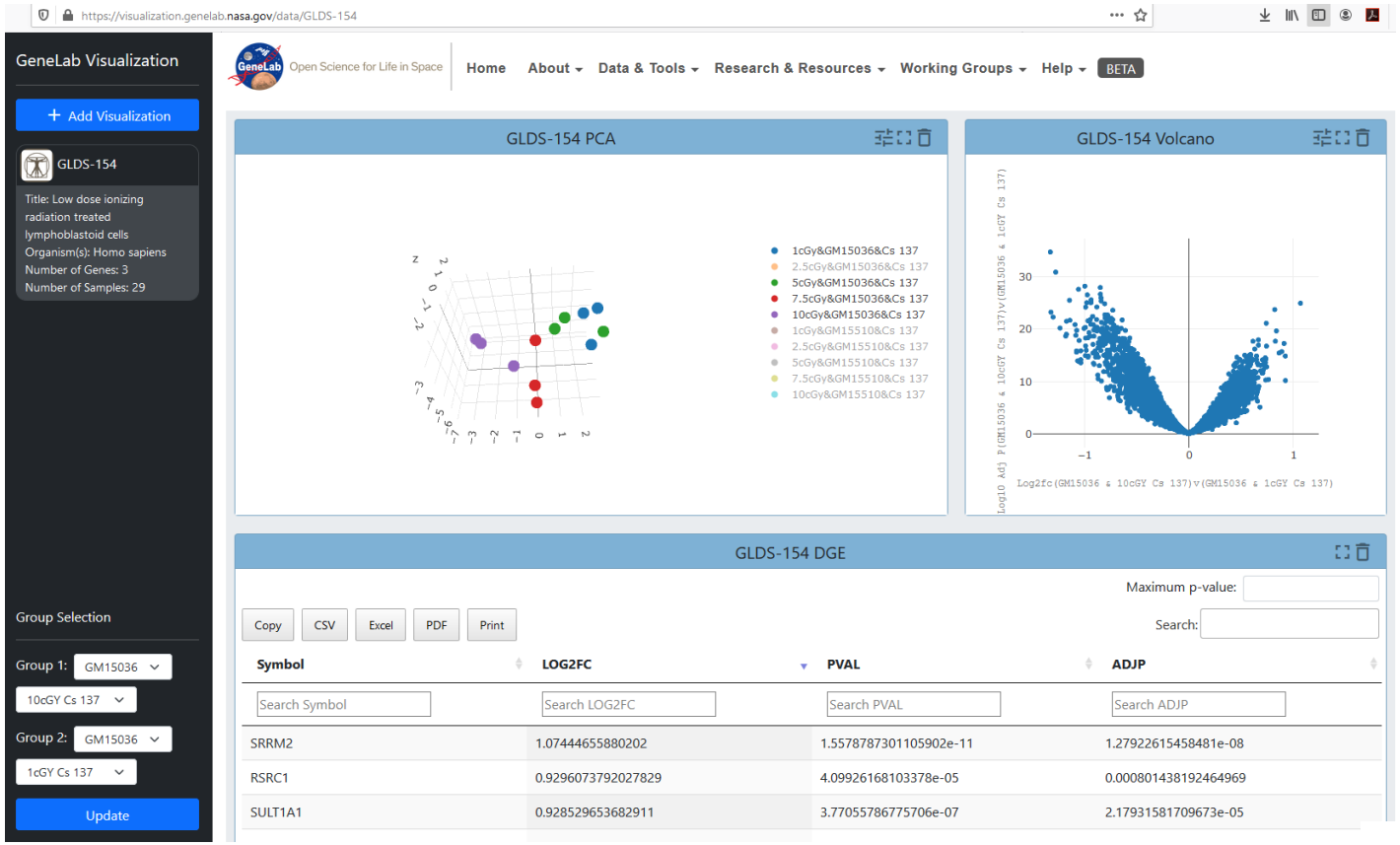


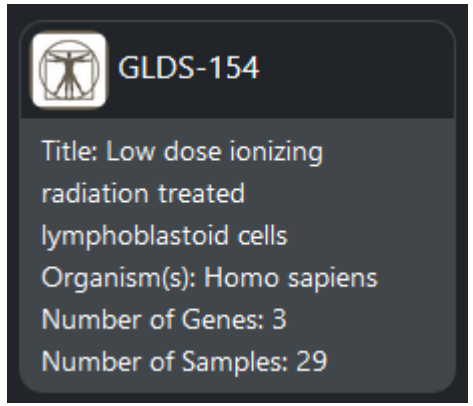
Example: Barcellos-Hoff Lab low dose data

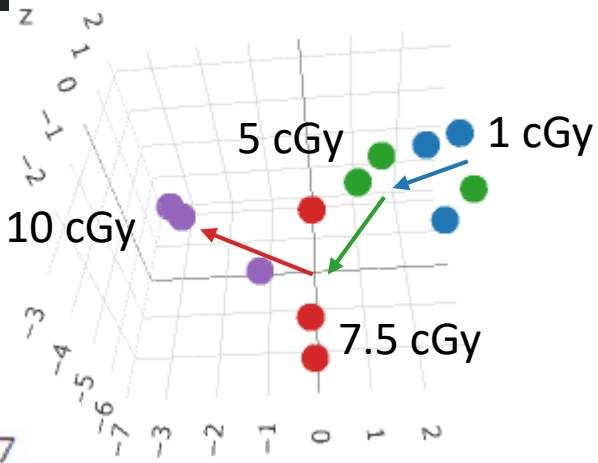
Low dose radiation transcriptomic data can be visualized

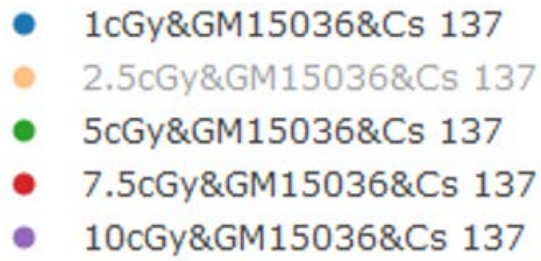
Application

Visualization Platform





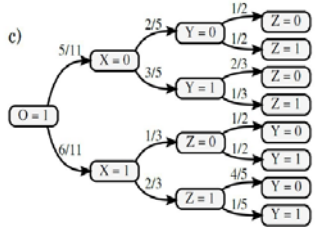




Wu P, Coleman M, Wyrobek AJ. "Low dose ionizing radiation treated lymphoblastoid cells", GeneLab, Version 3, http://doi.org/10.26030/hs0p-6w85

FDL-2021
SPACE MEDIC

CRISP Causal Inference Model



MODELING PORTAL (AI/ML)

Lauren Sanders, Ph.D.
AI/ML Scientist
GeneLab

GeneLab radiation data: enabling AI/ML meta-analysis

6 individual radiation exposure gene expression datasets:

GLDS-71: Immediate Transcriptional Changes in Response to High Dose Radiation Exposure Version 4

0.3 Gy – Cs-137

GLDS-152: Transcription profiling of human peripheral blood to development gene expression signatures for practical radiation biodosimetry Version 2

0.5 Gy – gamma ray

GLDS-156: Identifying radiation exposure biomarkers from mouse blood transcriptome Version 1

2 Gy – Cs-137

GLDS-157: Gene expression in human peripheral blood 48 hours after exposure to ionizing radiation Version 2

0.5 Gy – gamma ray

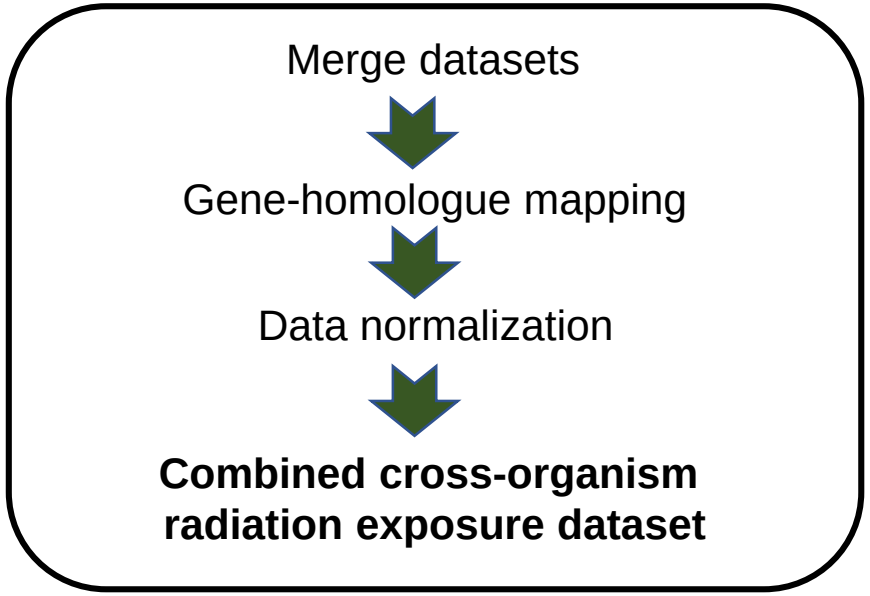
GSE124612: Transcriptomic responses in mouse blood during the first week after in vivo gamma irradiation

1.5 Gy – gamma ray

GSE62623: Gene expression in mouse blood following low dose-rate or acute x-ray exposure

1.1 Gy – X ray

Data Pipeline:



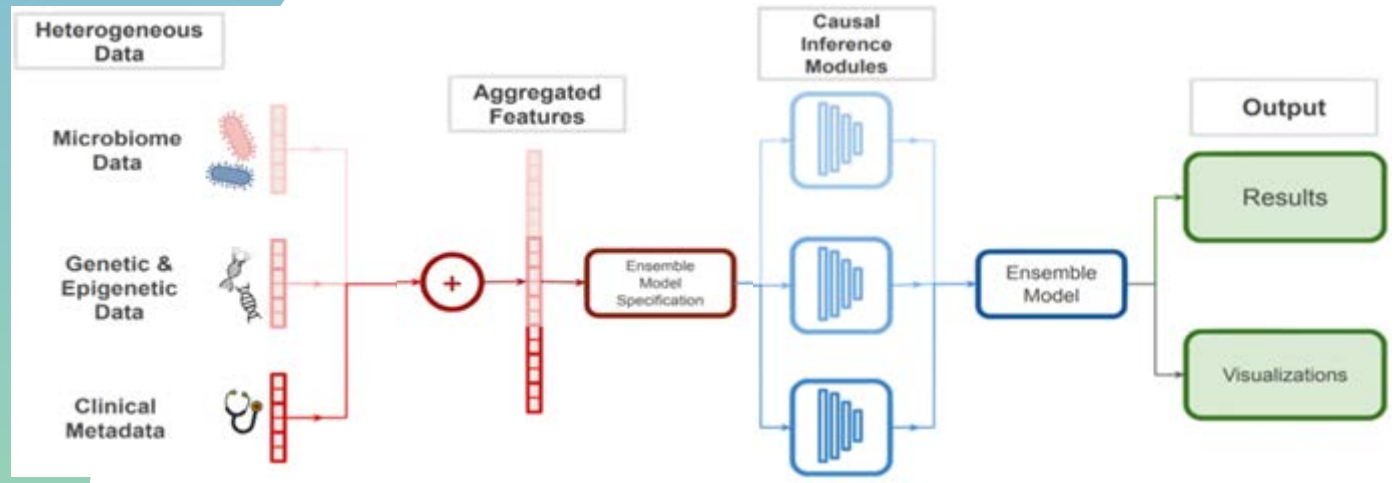
Combined Dataset:

- ~25,000 human-mouse gene homologues
- 455 total samples

- Gamma irradiation or non-irradiated controls:
  - *ex vivo* irradiation of human blood samples
  - *in vivo* irradiated mouse whole blood

Causal inference in complex biological data

CRISP Platform Overview



Generalizable to different data types with the ability to learn causal drivers of any feature within a dataset

Automated dimensionality reduction using machine learning selects the most important input variables for causal analysis

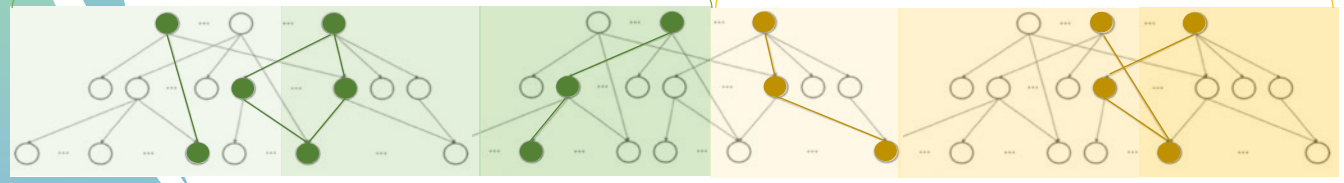
An ensemble of causal inference models "votes" on the selection of causal features

weighted_coefficients

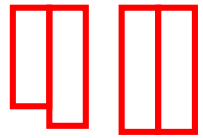
Mouse

Human

6 Datasets



Cancer genes



Top 20 features

TARGET: radiation state (irradiated or control)

CRISP identifies features that are *causal* of a target variable

Developed by FDL Astronaut Health 2020 Team

Co

- **Low dose data are relevant to space flight missions**
- **The NASA Open Science Database system from Space Biology is well suited to ingest low dose DOE data (omics and non-omics)**
- **The usage of FAIR principles for GeneLab and ALSDA (Findable, Accessible, Interoperable and Reusable) facilitates meta-analysis and enables modeling of radiation risk**

NASA Open Science Data Systems

The Division of Biological and Physical Sciences (***BPS)***