

GeneLab: Visualization and VWG - Introduction

National Aeronautics and
Space Administration

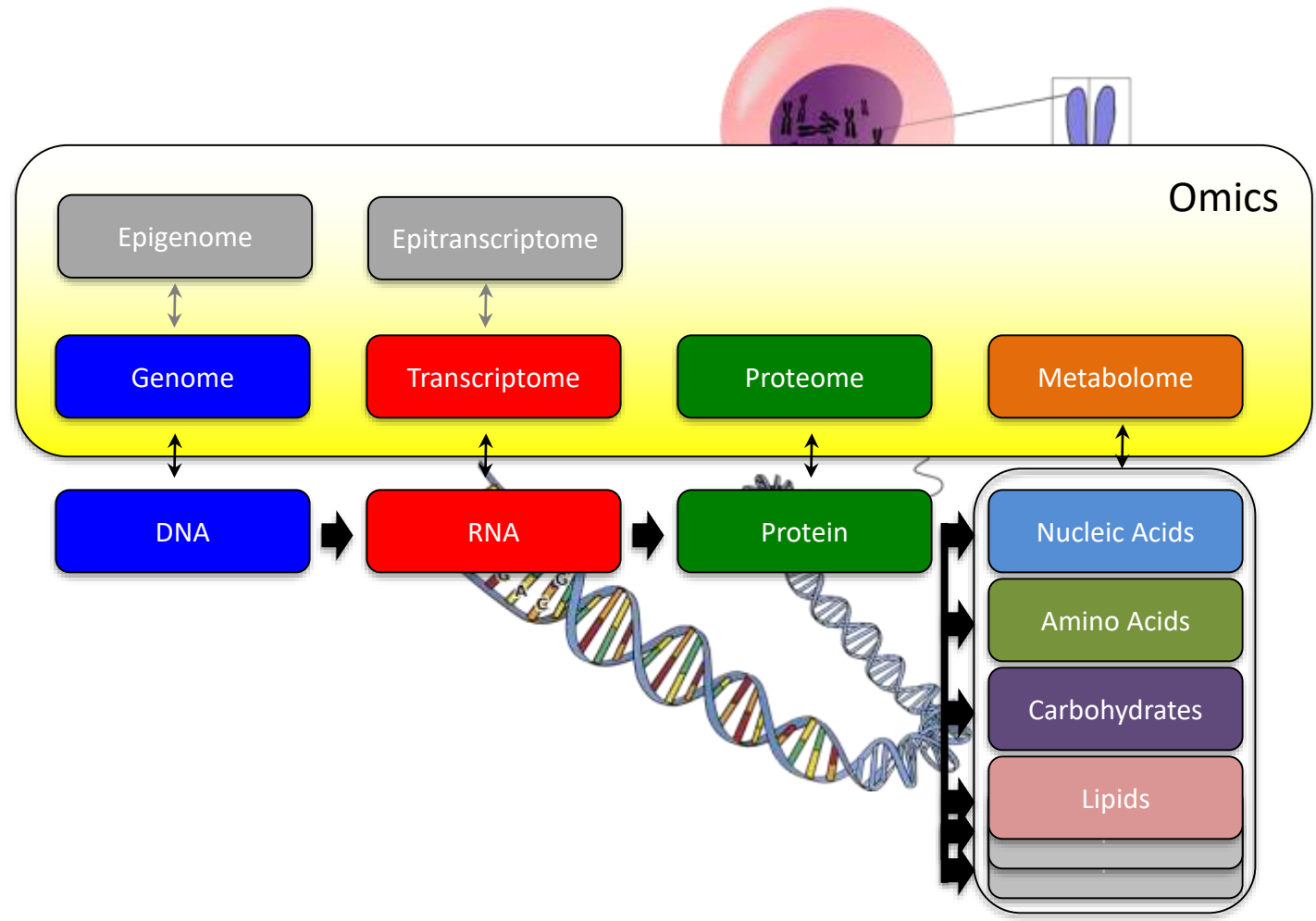


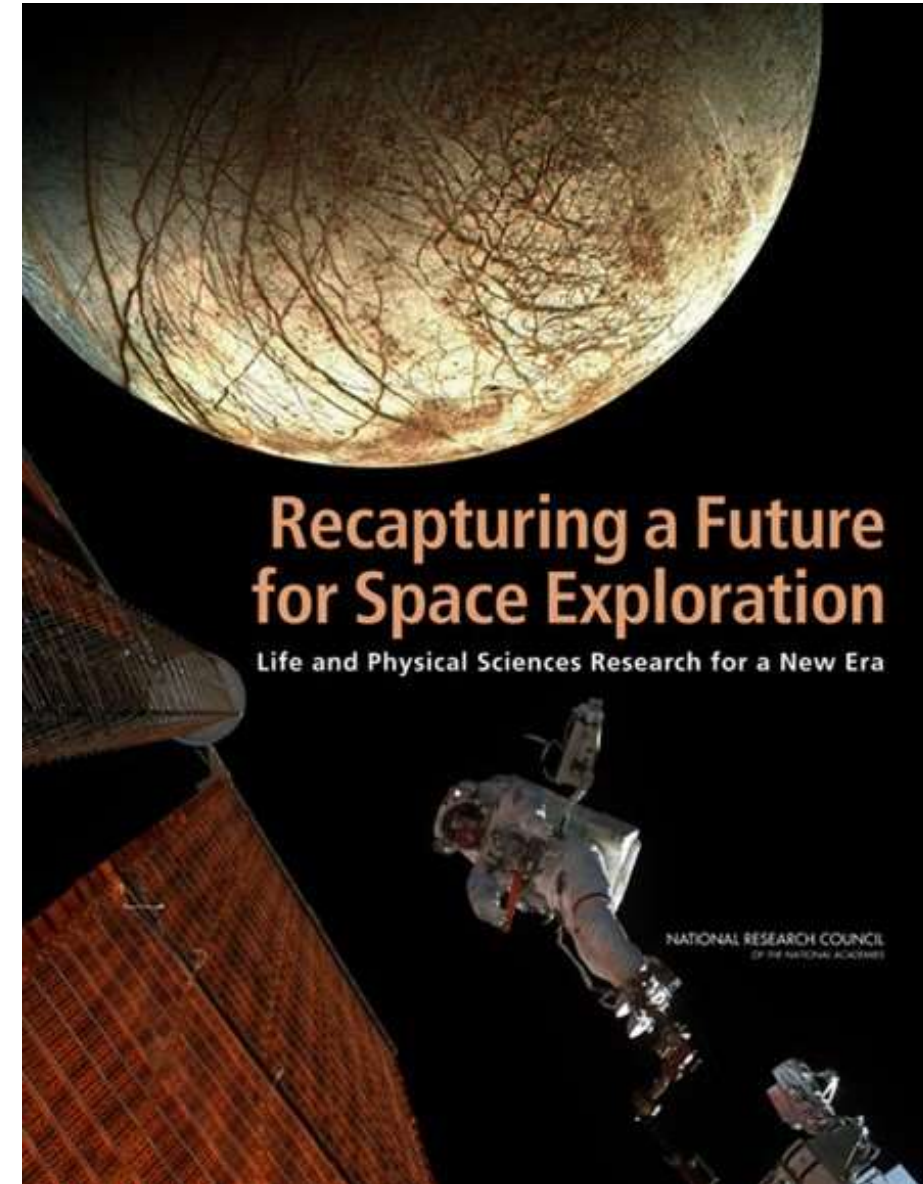
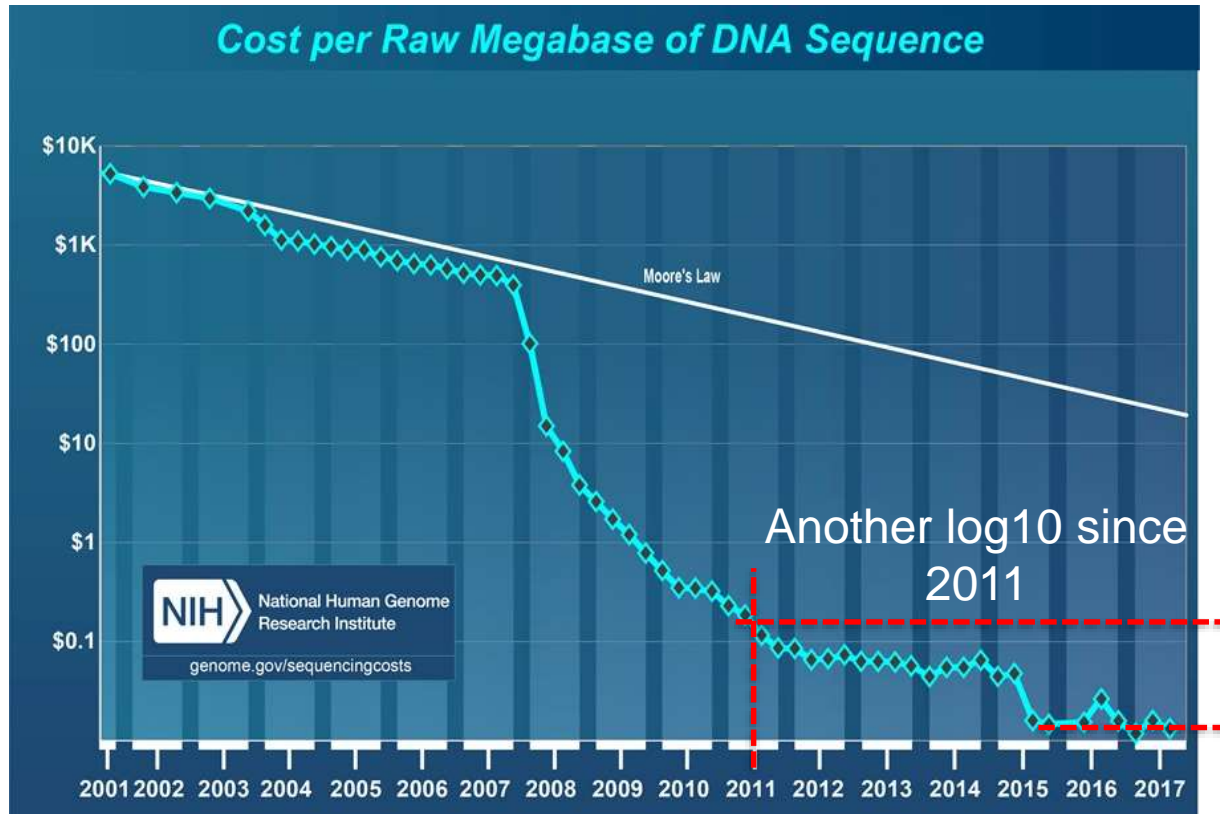
*Disseminate and reuse data, tools, and samples
post-project*

Sylvain V. Costes, PhD
GeneLab Project Manager

Afshin Beheshti, PhD
GeneLab Scientist and VWG Lead



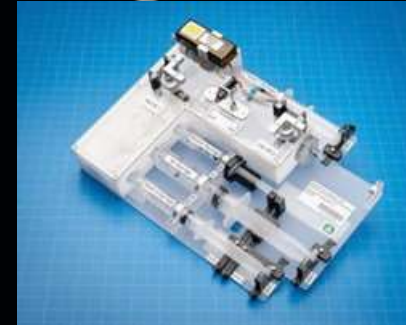
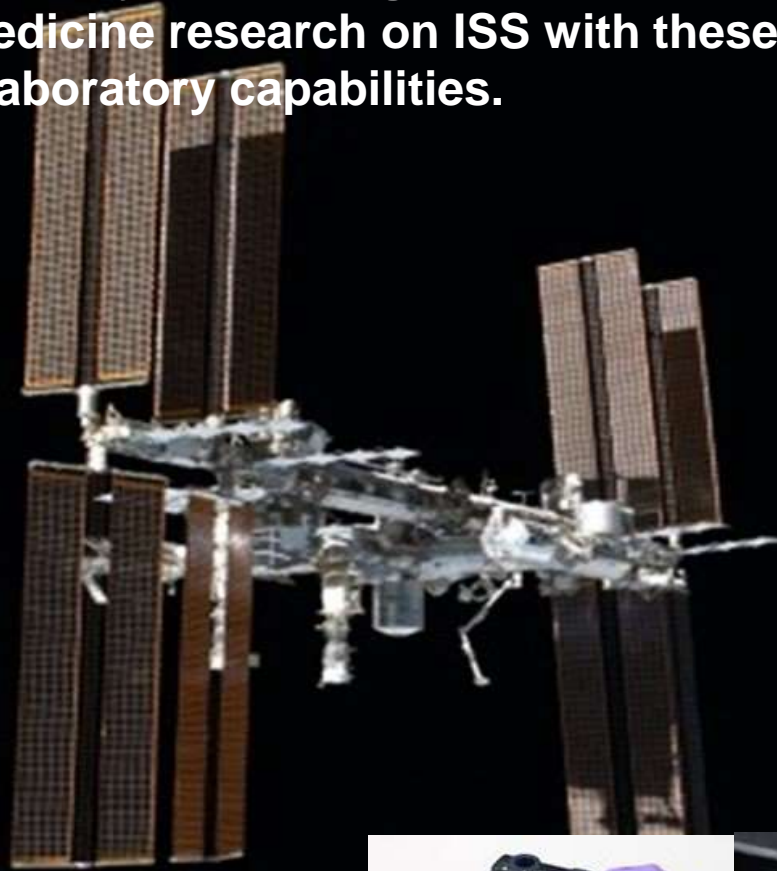




“...genomics, transcriptomics, proteomics, and metabolomics offer an immense opportunity to understand the effects of spaceflight on biological systems...”

*“...Such techniques generate considerable amounts of **data that can be mined and analyzed** for information by multiple researchers...”*

This is truly an exciting time for cellular and molecular biology, omics and biomedicine research on ISS with these amazing additions to the suite of ISS Laboratory capabilities.



Sample Preparation Module



Oxford Nanopore MinION Gene Sequencer

Cepheid Smart Cycler qRT-PCR

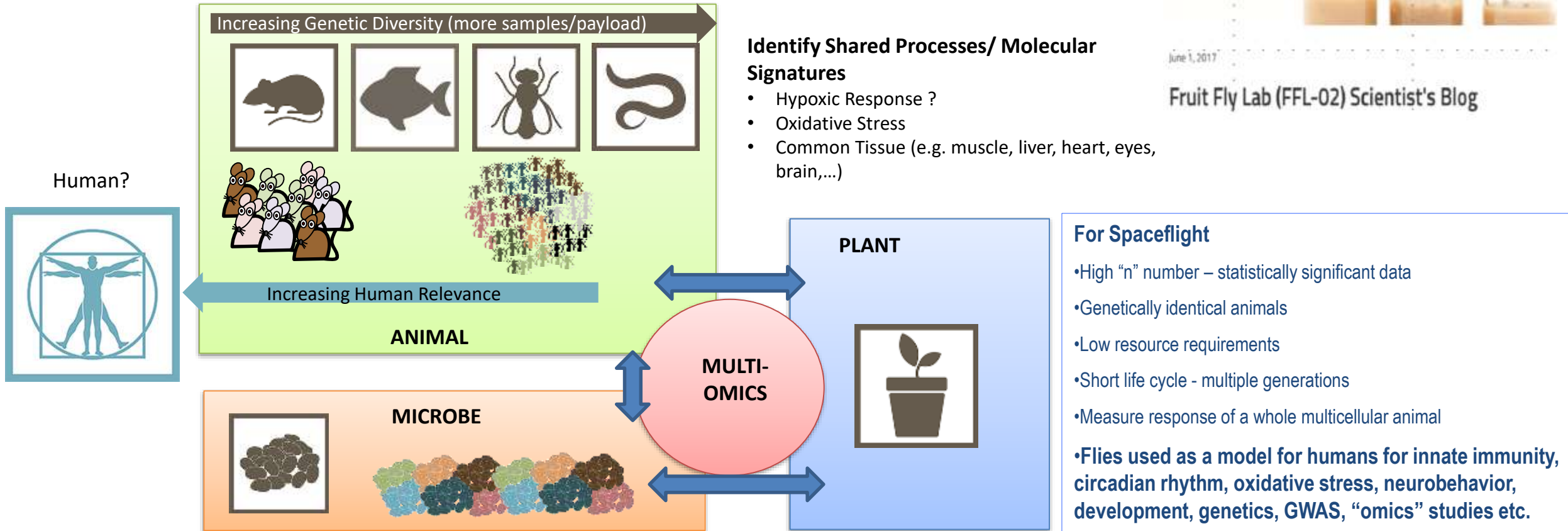
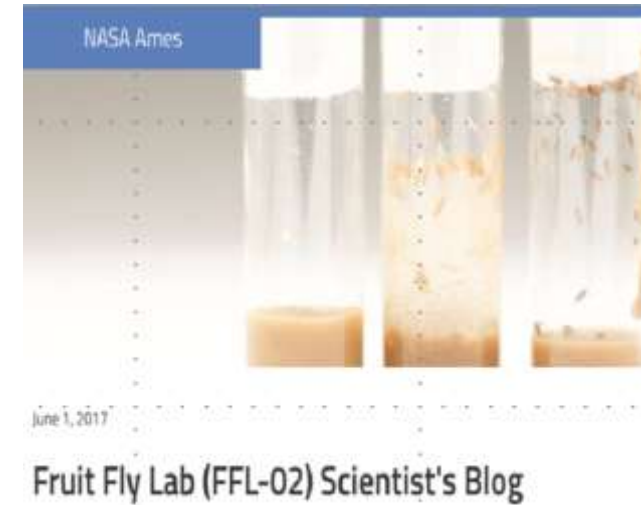


Reaction tube containing lyophilized chemical assay bead (proprietary)

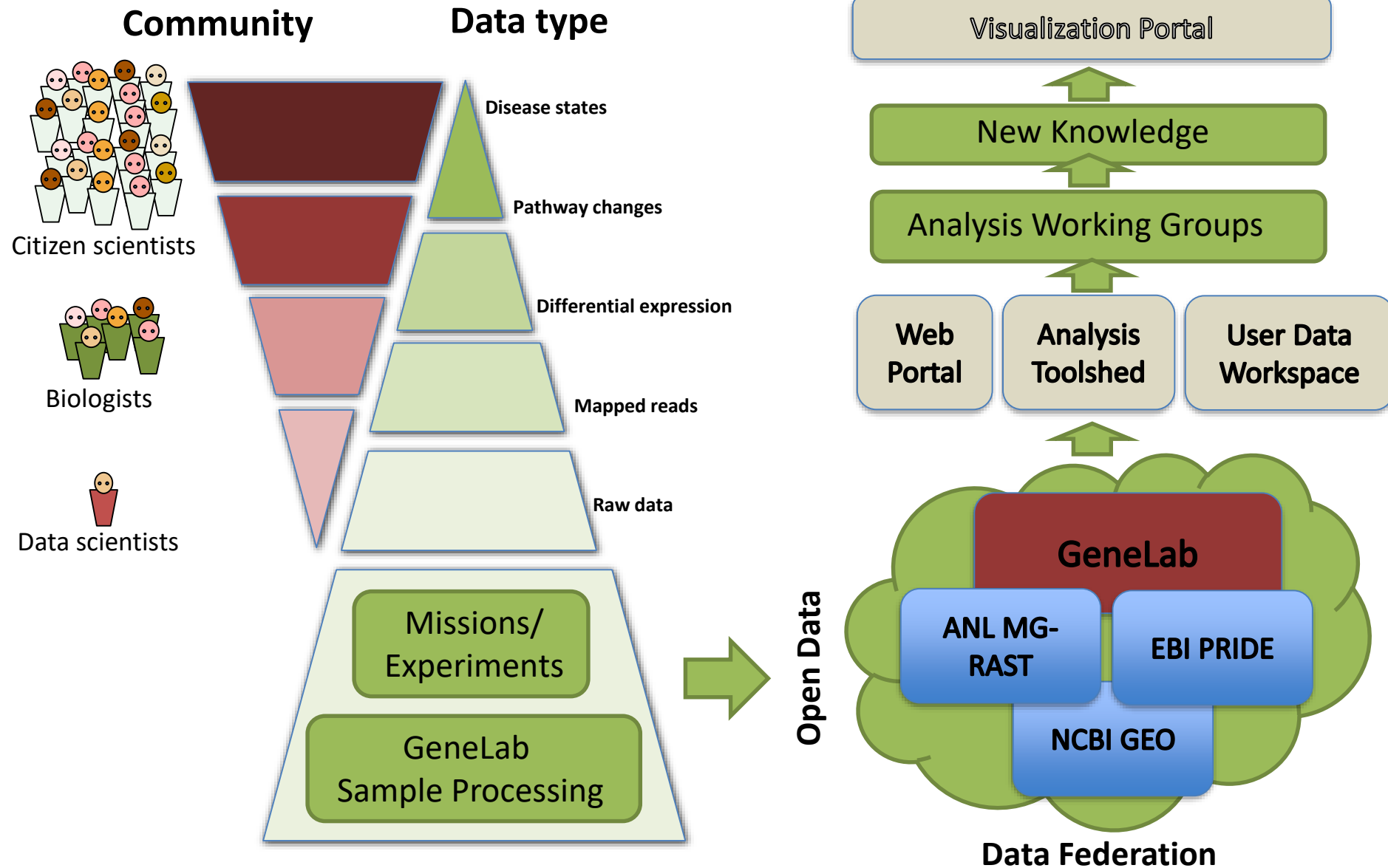


Mini-PCR

- **Sequencing on ISS is still limited in the amount of data generated**
 - Most of the work needs to happen on earth
- **Measurements on human cannot be too invasive and limited in numbers**
 - Usage of animals



GeneLab Data Democratization



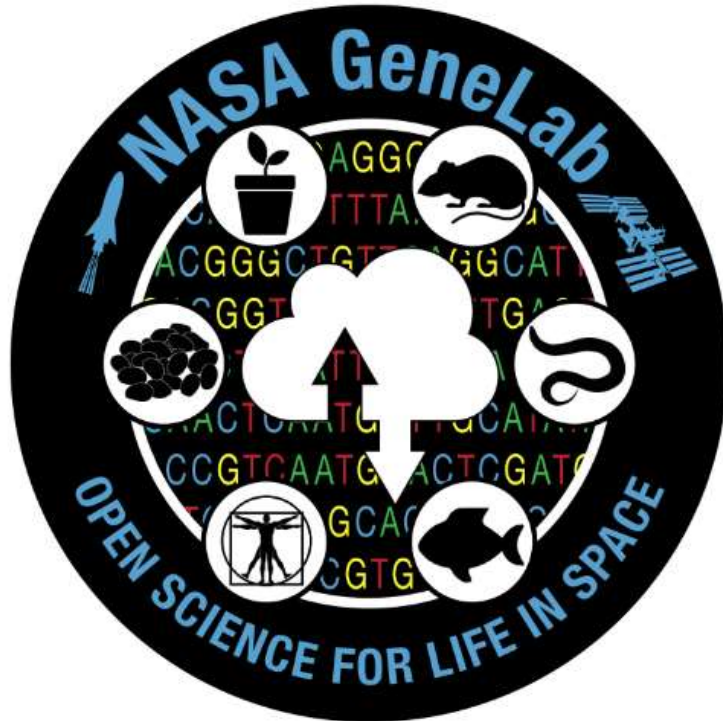


GeneLab

Open Science for Life in Space

[Home](#) [About](#) [Data & Tools](#) [Research & Resources](#) [Help](#)

Keywords



Welcome to NASA GeneLab – the first comprehensive space-related omics database in which users can upload, download, share, store, and analyze spaceflight and corresponding model organism data.



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?



Tutorials

New to GeneLab?



Basic Visualization

Quick analysis and visualization of GeneLab Datasets



Advanced Visualization

Visualization and advanced analysis of GeneLab Datasets

GeneLab Visualization	2018			2019									2019-2020
	Oct	Nov	Dec	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sept	
Phase 1: Formed VWG													
Phase 2: Prototype Development-Visualization on GeneLab													
Phase 2: Prototype Development-Notebook based visualization													
Phase 3: Integration & Testing of Visualization with GeneLab													
Phase 4: Maintenance & Future Collaborations													

TCGA
<https://portal.gdc.cancer.gov/>

Harmonized Cancer Datasets Genomic Data Commons Data Portal

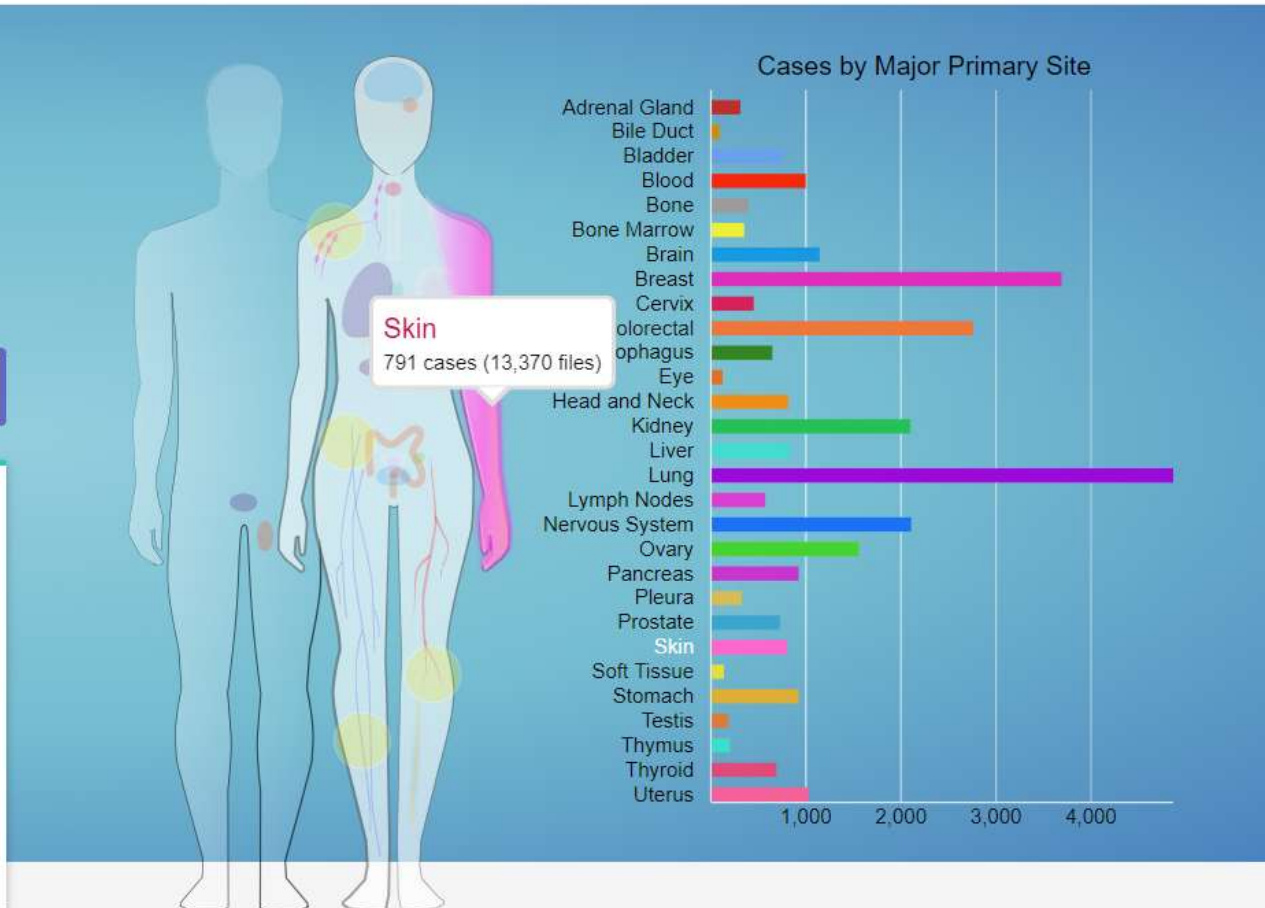
Get Started by Exploring:

Projects
Exploration
Analysis
Repository

Q e.g. BRAF, Breast, TCGA-BLCA, TCGA-A5-A0G2

Data Portal Summary Data Release 15.0 - February 20, 2019

PROJECTS 44	PRIMARY SITES 68	CASES 33,227
FILES 361,076	GENES 22,872	MUTATIONS 3,142,246



GN TP53

LFS1
p53

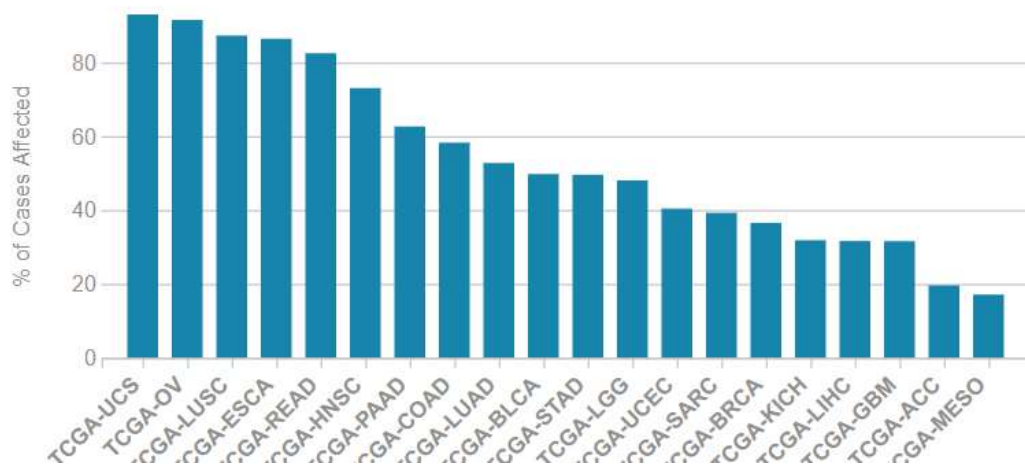
[HGNC:11988](#)
[OMIM:191170](#)
[Ensembl:ENSG00000141510](#)

Type	protein_coding
Location	chr17:7661779-7687550 (GRCh38)
Strand	-
Description	This gene encodes a tumor suppressor protein containing transcriptional activation, DNA binding, and oligomerization domains. The encoded protein responds to diverse cellular stresses to regulate expression of target genes, thereby inducing cell cycl...
Annotation	Cancer Gene Census

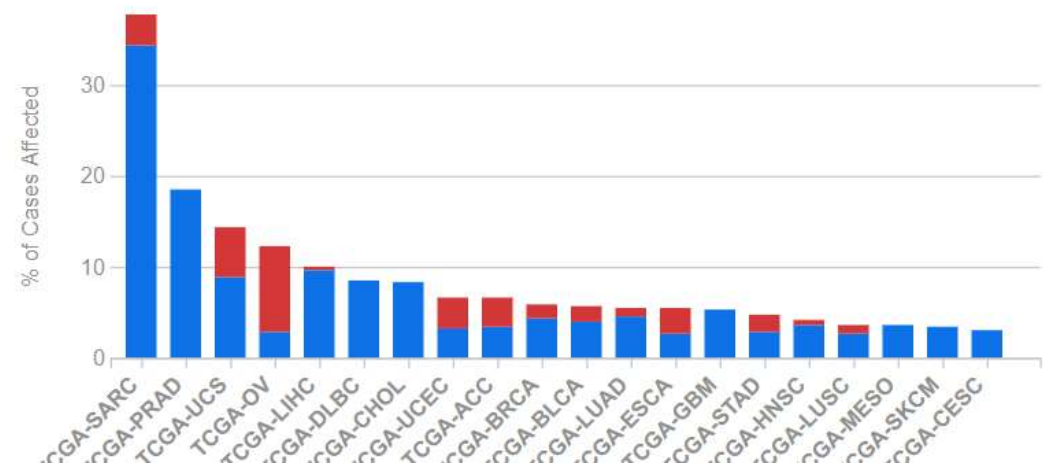
Bar Chart Cancer Distribution

[Open in Exploration](#)

4,008 CASES AFFECTED BY 1,291 MUTATIONS ACROSS 32 PROJECTS



4,251 CASES AFFECTED BY 638 CNV EVENTS ACROSS 31 PROJECTS



Set Operations x
3/4/2019

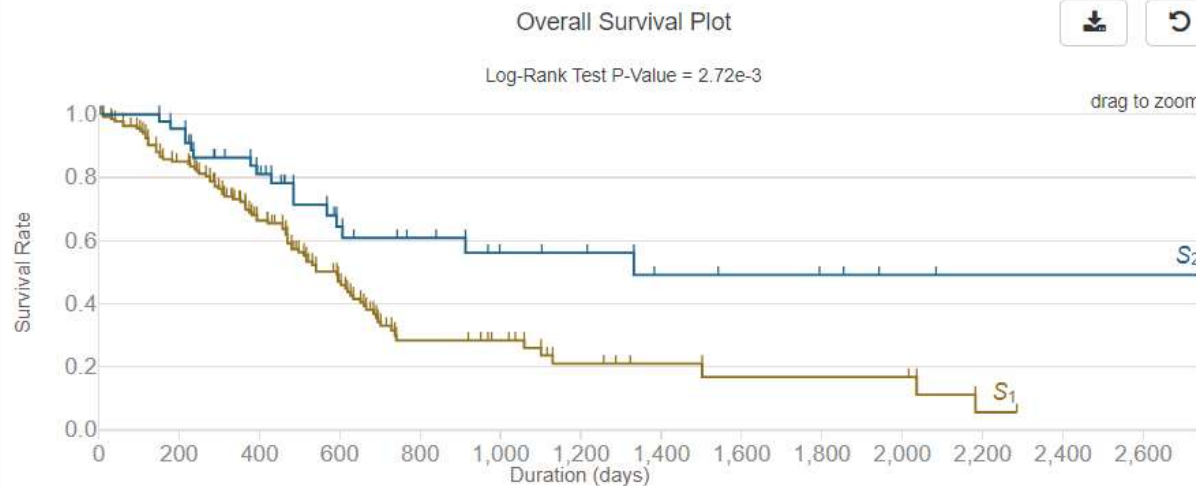
Cohort Comparison x
3/4/2019

Delete All

Cohort Comparison

Demo showing cases with pancreatic cancer with and without mutations in the gene KRAS.

Survival Analysis



Cases included in Analysis	# Cases S ₁	%	# Cases S ₂	%
Overall Survival Analysis	139	99%	45	6%

Gender



Cohort

Cases

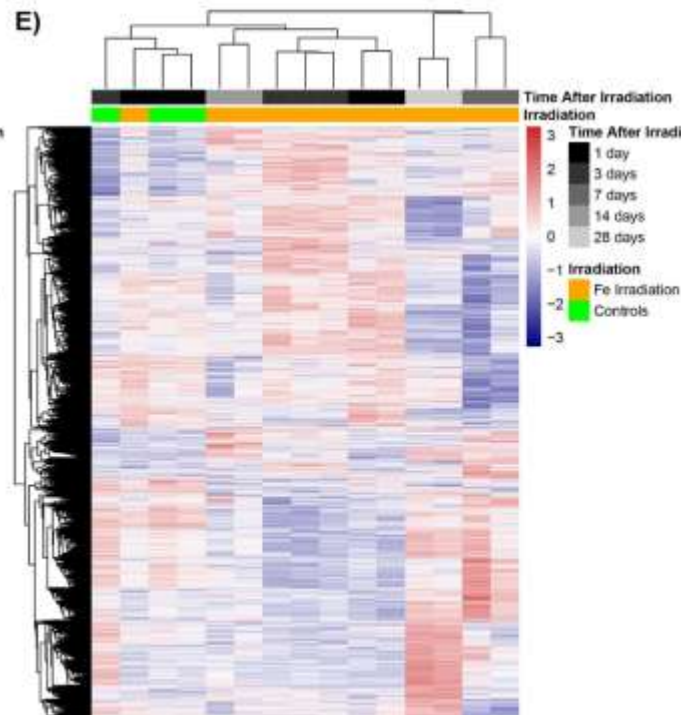
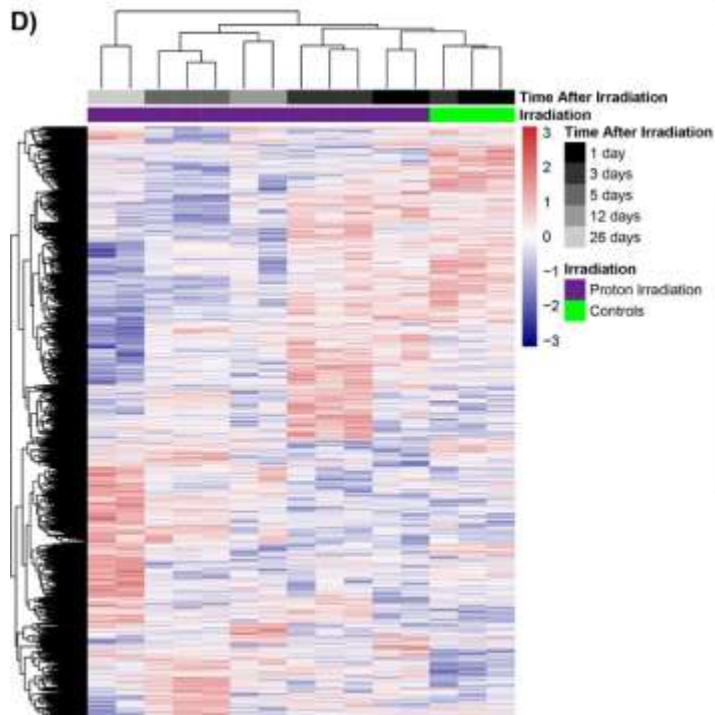
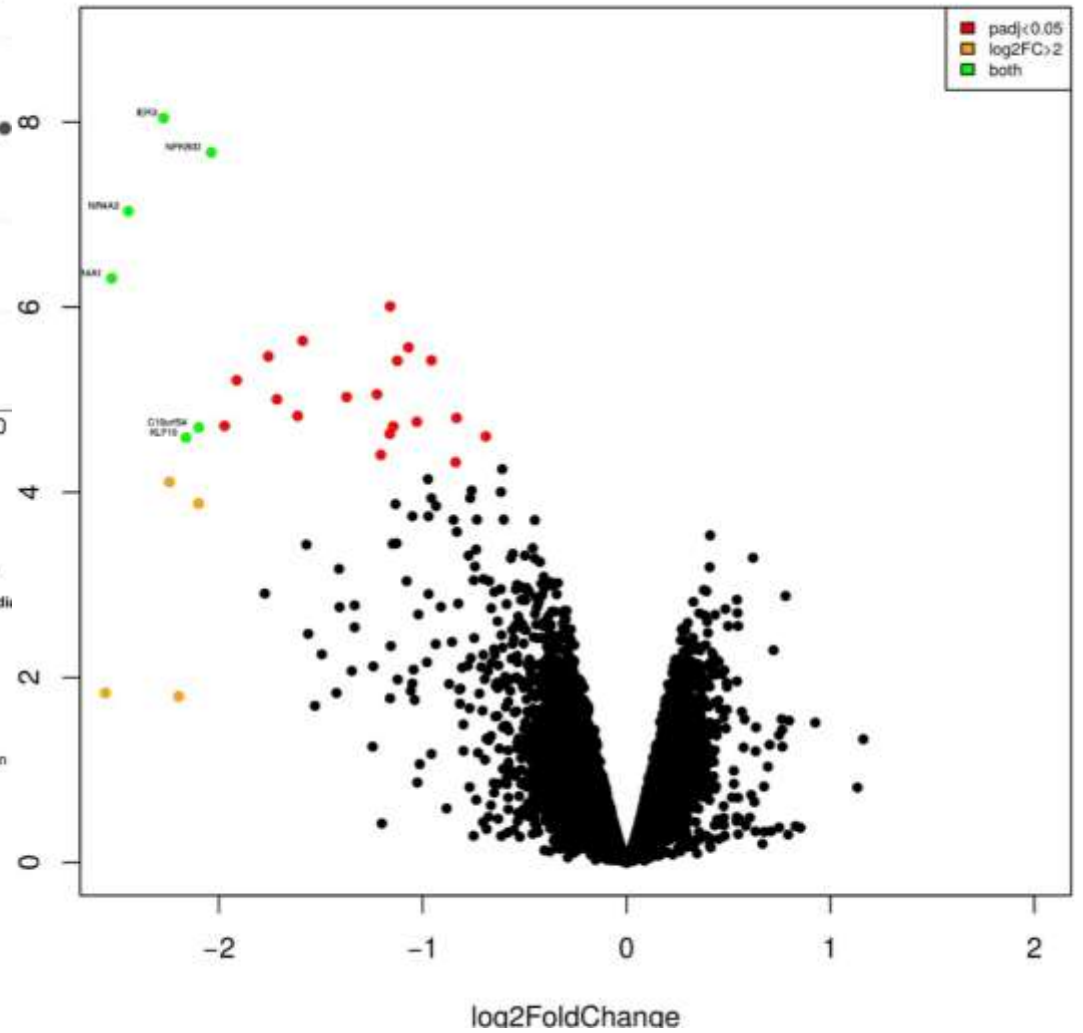
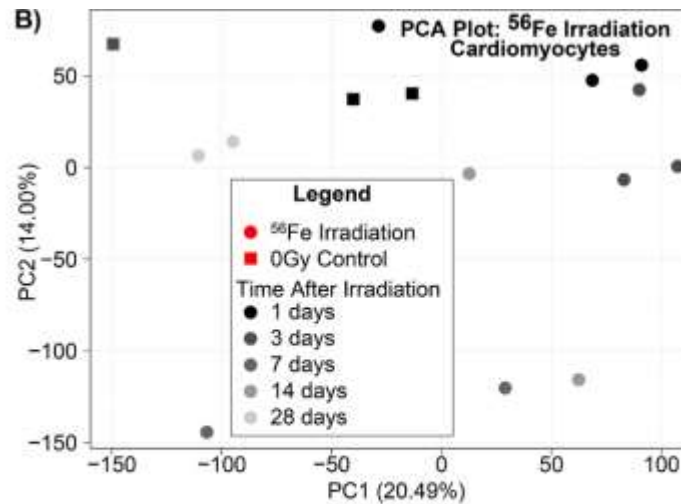
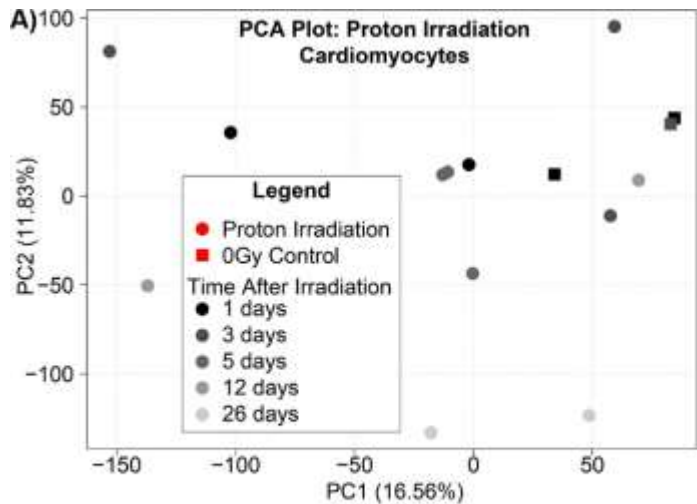
S₁ : Pancreas - KRAS mutated [140](#)

S₂ : Pancreas - KRAS not mutated [774](#)

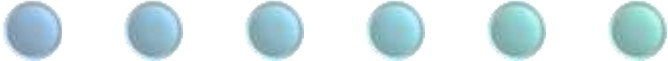
[Open Venn diagram in new tab](#)



- Survival
- Ethnicity
- Gender
- Vital Status
- Race



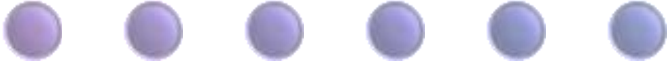
Basic visualization



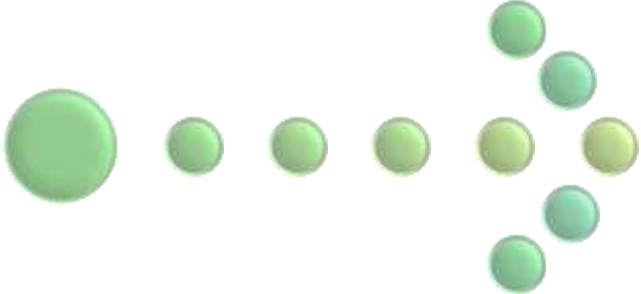
- Fixed Interaction
- Limited Functionality



Advanced Visualization



- Customizable
- Ad Hoc

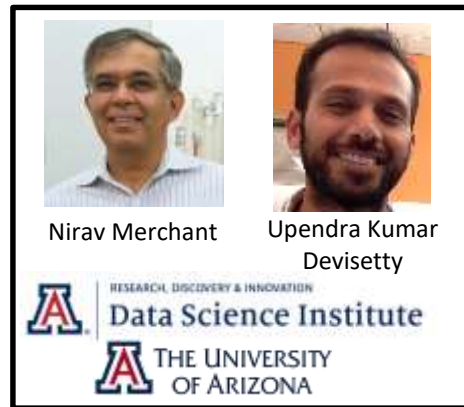


- **Two Components:**
 - One team developing the Visualization Tools
 - One team assisting with API development
- **Both teams are involved on a voluntary basis and VWG designed to be assistance from the global scientific community**
- **Established on 10/2018**

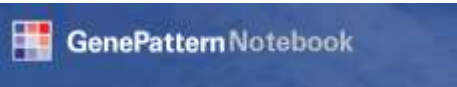
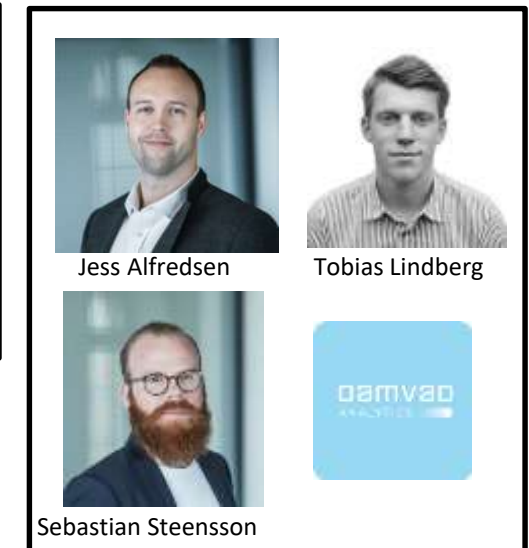
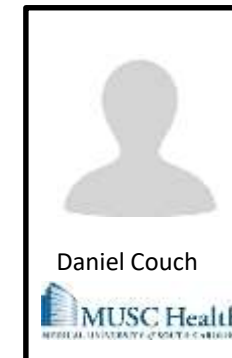
Leading the VWG



Advanced Visualization Team



Basic Visualization Team



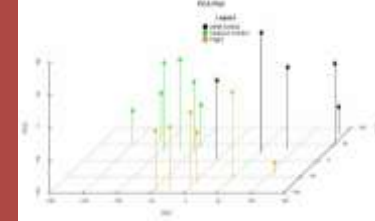
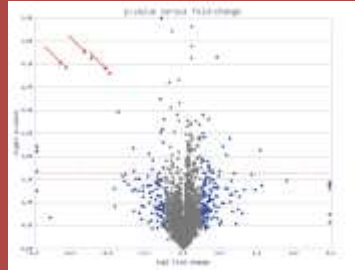
Work Already Done

Future Work



Rien Dijkstra (Astrophysicists/Software Engineer in Amsterdam)

- Interactive PCA plots
- Interactive Volcano Plots
- Interactive Comparison Plots
- Interactive Dendrogram



1st Prototype:
12/21/18
2nd Prototype:
1/11/19
3rd Prototype:
2/8/19

1st Scheduled
Finalized
Prototype for
GeneLab to Test:
4/01/19
UAT: 9/2019



DAMVAD (Software Engineers and Big Data Analysts)

- Interactive Heat map
- Interactive Mouse Diagram integrated with the omics data on GeneLab

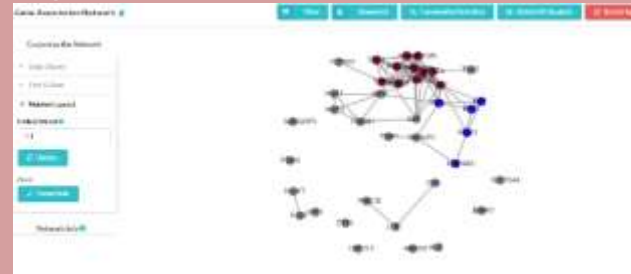
1st Prototype:
12/24/18

1st Scheduled
Finalized
Prototype for
GeneLab to Test:
6/01/19
UAT: 9/2019



Daniel Couch (Systems Programmer)

- Interactive Network tool:
 - Adapt his tool GAIL for GeneLab



Pending
Prototype
(Possible date:
4/1/19)

1st Scheduled
Finalized
Prototype for
GeneLab to Test:
7/01/19
UAT: 9/2019

GeneLab API Visualization Development

API wrapper development and integration with GenePattern Notebook



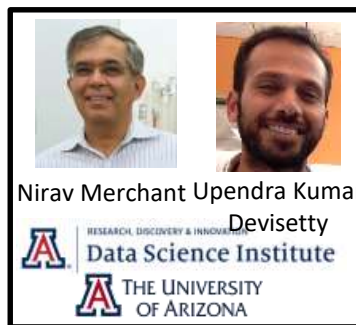
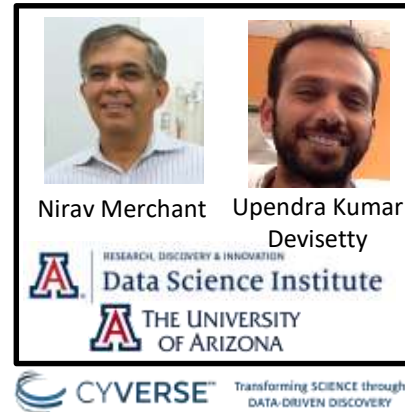
Incorporation of GenePattern/Genelab API into Cyverse

Cyverse to implement a basic GenePattern Notebook/Genelab API interface



Visualization and Advanced Omics Analysis

Utilizing both GenePattern Notebook and Cyverse with the GeneLab API tool, GeneLab pipelines will be made available for advanced omics analysis



Methods to Access the Data

Public Visualization (no account)

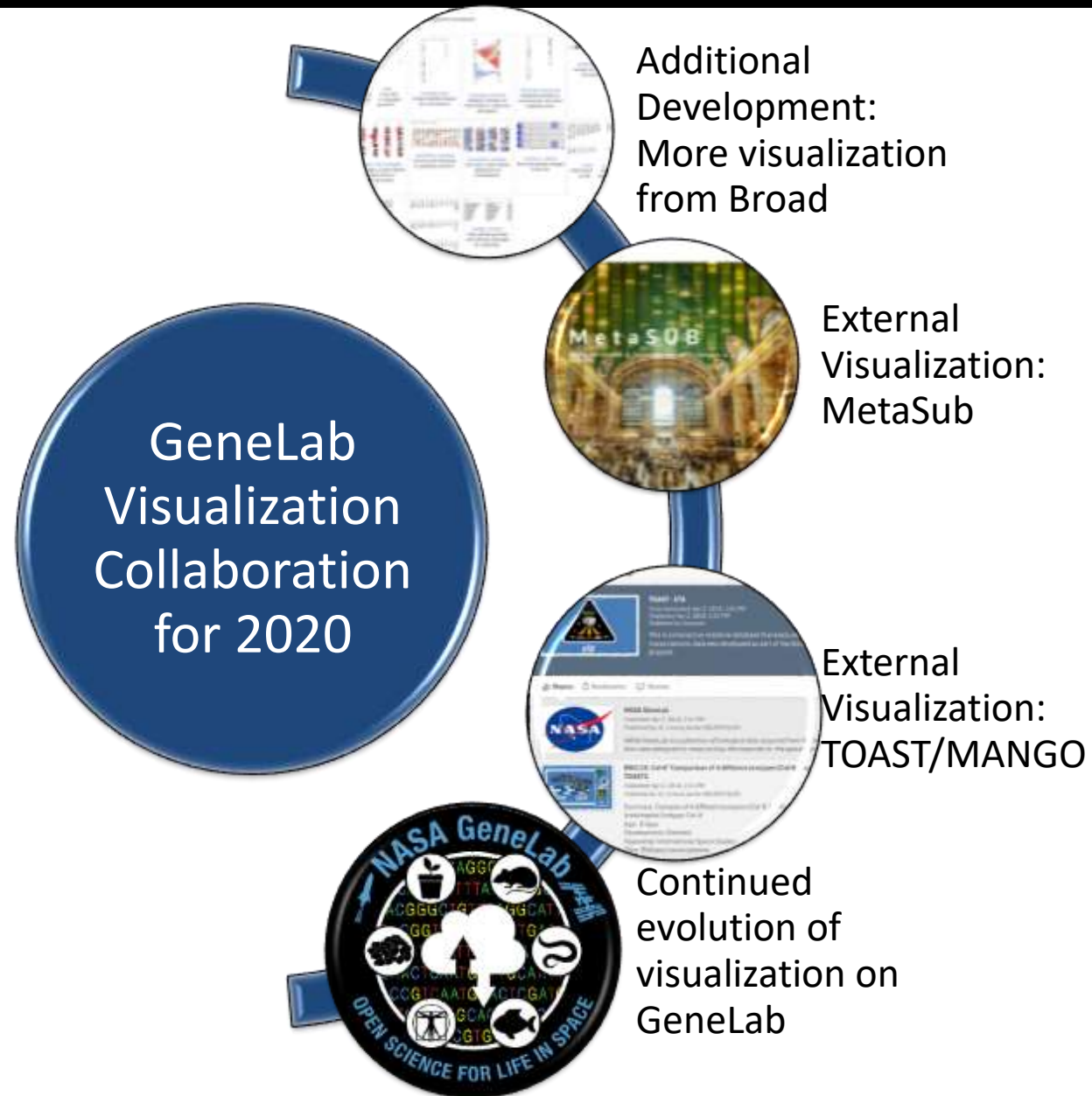


A provided redundant table listing genes and their various attribute



For advanced user, provide intermediate files with specific functions for a workflow that can be modified





setup.py

Removed tqdm

6 days ago

README.md

GeneFab

Installation

GeneFab has been tested under Python 3.5+.

It is recommended to use a user installed [Conda](#) environment.

Currently GeneFab can be installed via `pip` directly from github:

```
pip install -e git://github.com/LankyCyril/genefab.git#egg=genefab
```

For development purposes, it can also be cloned and used like this:

```
$ git clone https://github.com/LankyCyril/genefab
$ cd genefab
$ conda env create --name genefab --file environment-linux.yaml
$ conda activate genefab
$ python
>>> from genefab import get_datasets, GLDS
...

```

Demo

<https://github.com/LankyCyril/genefab/blob/master/genefab-demo.ipynb>

Description

The current iteration of GeneFab supports GeneLab datasets processed according to API version 2.1. As of April 2, 2019, these are only the datasets GLDS-4, GLDS-30, and GLDS-42.

```
GeneLabDataSet(accession, storage_prefix=".genelab", index_by="Sample Name", verbose=False)
```

Initializes a GeneLabDataSet instance corresponding to `accession`.

Each GeneLabDataSet contains references to assays performed in the study; they are stored under the field `assays` as a list of `Assay` instances.





Kirill's API GeneLab API Development



```

[] rows x 8 columns

In [9]: assay.factors
Out[9]: Factor
Sample Name          Factor Value: Microgravity Simulation
Mnus_2T3cells_1G_RWV_CTRL_3days_Rep1  1G
Mnus_2T3cells_1G_RWV_CTRL_3days_Rep2  1G
Mnus_2T3cells_1G_RWV_CTRL_3days_Rep3  1G
Mnus_2T3cells_uG_RWV_3days_Rep1      uG
Mnus_2T3cells_uG_RWV_3days_Rep2      uG
Mnus_2T3cells_uG_RWV_3days_Rep3      uG

In [10]: assay.processed_data[0]
Out[10]: Sample Name  Mnus_2T3cells_1G_RWV_CTRL_3days_Rep1  ---  Mnus_2T3cells_uG_RWV_3days_Rep3
ID
NM_017477  9.429224  ---  9.953417
NM_013477  9.424985  ---  10.253487
NM_091042484  11.133064  ---  11.210010

[] rows x 8 columns

In [11]: glds = GeneLabDataSet("GL20-30")

GeneLabDataSet(self, accession, verbose=False,
storage_prefix='genelab', assay_class=Assay, index_by='Sample Name')

Stores GLDS metadata associated with an accession number
[F4] V1 (INSERT) 2088/2097 [F3] History [F6] Paste mode
Kirill@PC192789

```

```

In [13]: assay.factors
Out[13]: Factor
Hybridization Assay Name
GSM105147  1G
GSM105148  1G
GSM105149  1G
GSM105150  uG
GSM105151  uG
GSM105152  uG

In [14]: assay.processed_data[0]
Out[14]: Hybridization Assay Name  GSM105147  GSM105148  GSM105149  GSM105150  GSM105151  GSM105152
ID
NM_017477  9.429224  10.186013  10.186013  8.762991  9.709973  9.953417
NM_013477  9.424985  10.176255  10.234763  9.727475  10.807745  10.253487
NM_091042484  11.133064  11.183789  12.244323  11.132884  11.212148  11.210010

In [15]: assay.fields
Out[15]: {'Sample Name': {'a1000000samplename'}, 'Factor Value: Microgravity Simulation': {'a100001factorvaluemicrogravitiesimulation'}, 'Extract Name': {'a100000extractname'}, 'Protocol REF': {'a100011protocolref'}, 'a100005protocolref'}, 'Accession Type': {'a100000accesstype'}, 'Labelled Extract Name': {'a100009labelledextractname'}, 'Label': {'a100010label'}, 'Array Design REF': {'a100011arraydesignref'}, 'Array Data File': {'a100012arraydatafile'}, 'Derived Array Data Matrix File': {'a100017derivedarraydatamatrixfile'}, 'Comment: General Processed Array Data Files': {'a100018commentgeneralprocesseddatafiles'}, 'Comment: Raw Data Images': {'a100020commentrawdataimages'}, 'Comment: Raw Data Plots': {'a100021commentrawdataplots'}, 'Comment: Supplemental Materials': {'a100022commentssupplementalmaterials'}, 'Comment: Normalized Data Files': {'a100023commentnormalizeddatafiles'}, 'Comment: Normalized Annotated Data Files': {'a100024commentnormalizeddataannotateddatafiles'}, 'Comment: Normalized Data Plots': {'a100025commentnormalizeddataplots'}}

In [16]: assay.metadata[["*plots.*"]]

```

Welcome to NASA GeneLab – the first comprehensive space-related omics database in which users can upload, download, share, store, and analyze spaceflight and corresponding model organism data.

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?	Tutorials New to GeneLab?
Basic Visualization Quick analysis and visualization of GeneLab Datasets	Advanced Visualization Visualization and advanced analysis of GeneLab Datasets

LATEST DATA RELEASES

TRANSCRIPTOMICS

GLDS-208: Comparative gene expression analysis in the Arabidopsis thaliana root zone using RNA-seq

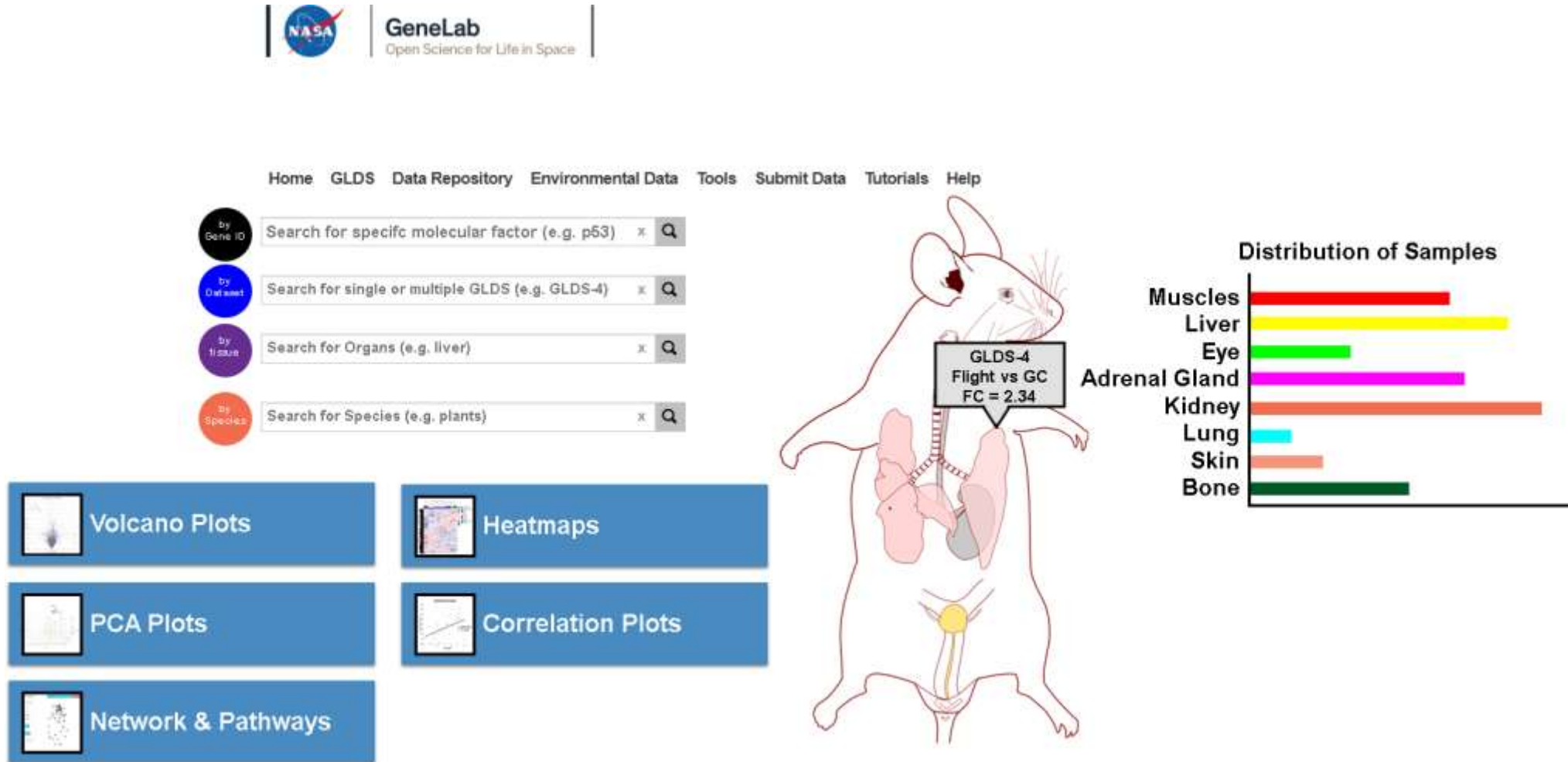
GENOMICS

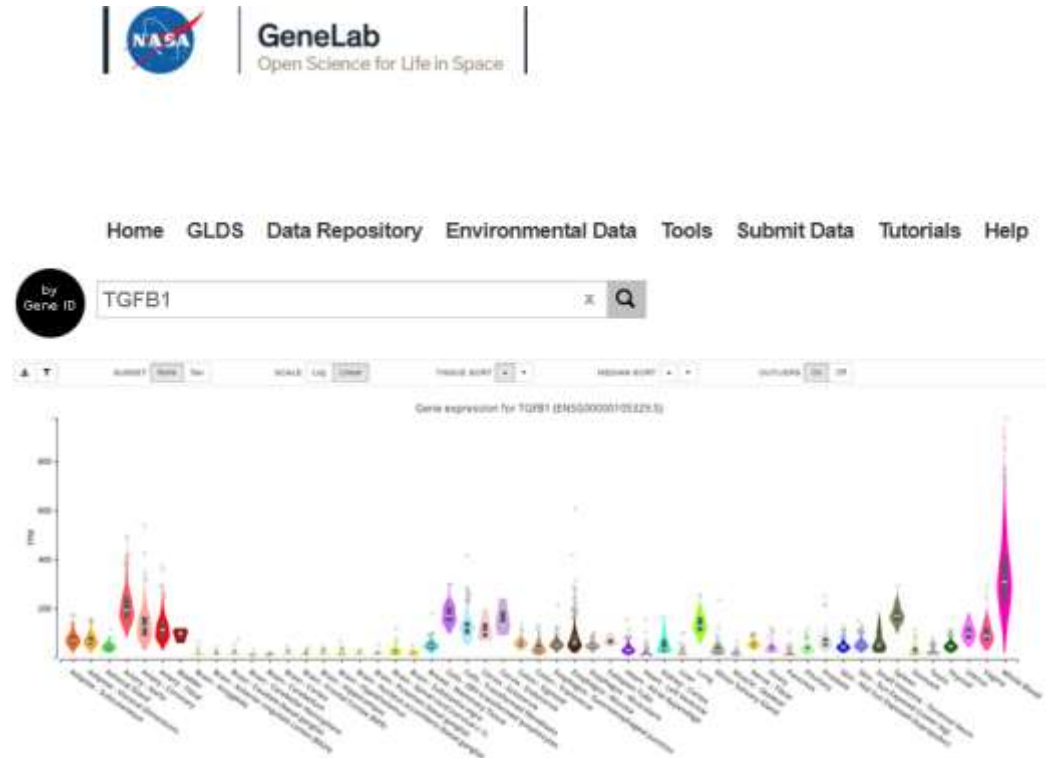
GLDS-203: Low dose (0.4 Gy) irradiation (LDR) and biofluids endpoints (S111) experiment in mice (P382)

PROTEOMICS / METABOLOMICS

GLDS-209: Re-Adaption on Earth after Spaceflights About the Mouse Line Database








Visualize with one of the following options:

 Volcano Plots

 Heatmaps

 PCA Plots

 Correlation Plots

 Network & Pathways

 | **GeneLab**
Open Science for Life in Space |  | 

[Home](#) [GLDS](#) [Data Repository](#) [Environmental Data](#) [Tools](#) [Submit Data](#) [Tutorials](#) [Help](#)

GenePattern Login





GenePattern Server
Broad Institute

GenePattern Username
Username

GenePattern Password
Password

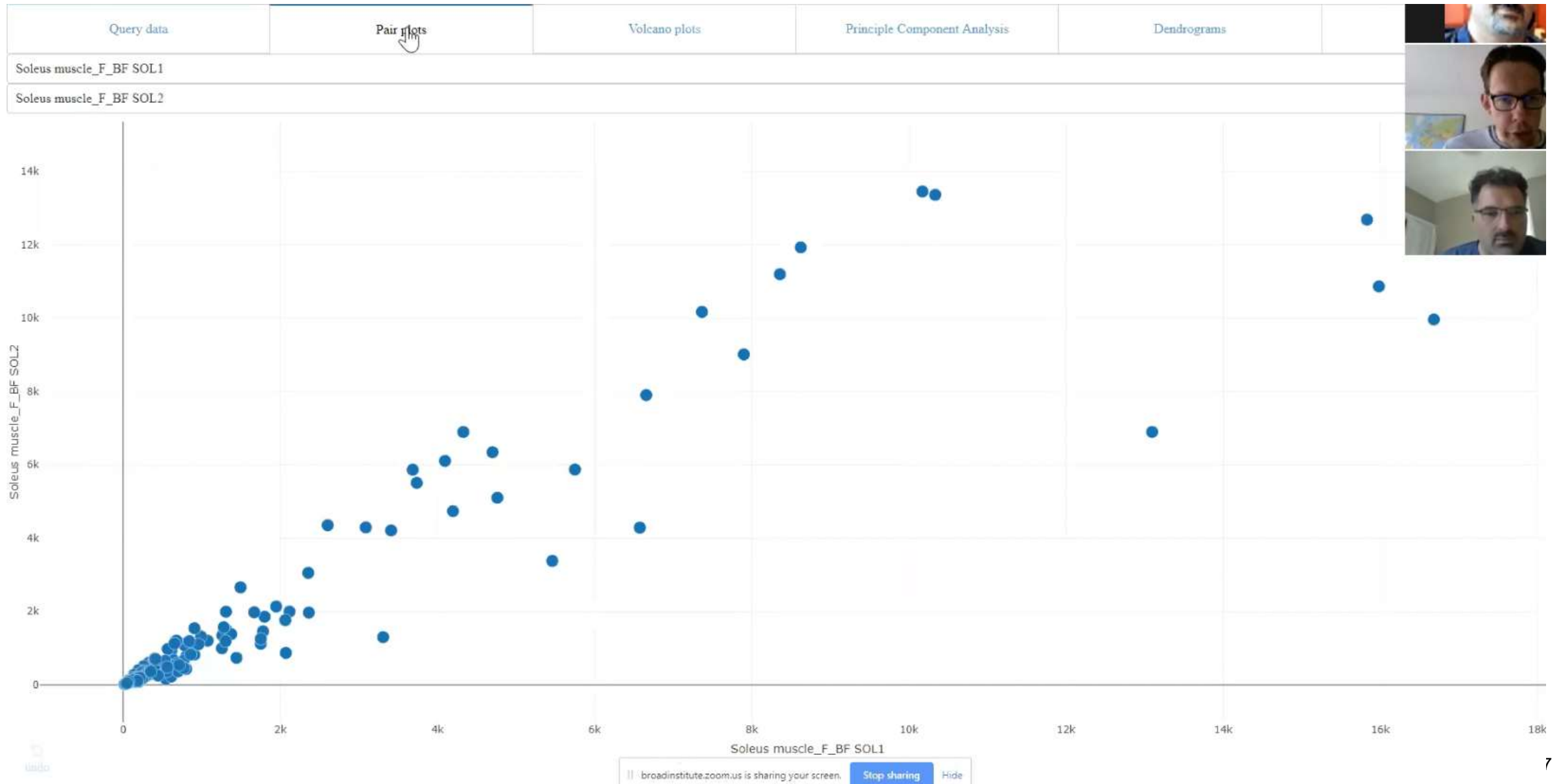
[Log into GenePattern](#) [Register an Account](#)

**User Interface
without adjustments**

-  by Gene ID Search for specific molecular factor (e.g. p53) x Q
-  by Dataset Search for single or multiple GLDS (e.g. GLDS-4) x Q
-  by Tissue Search for Organs (e.g. liver) x Q
-  by Species Search for Species (e.g. plants) x Q

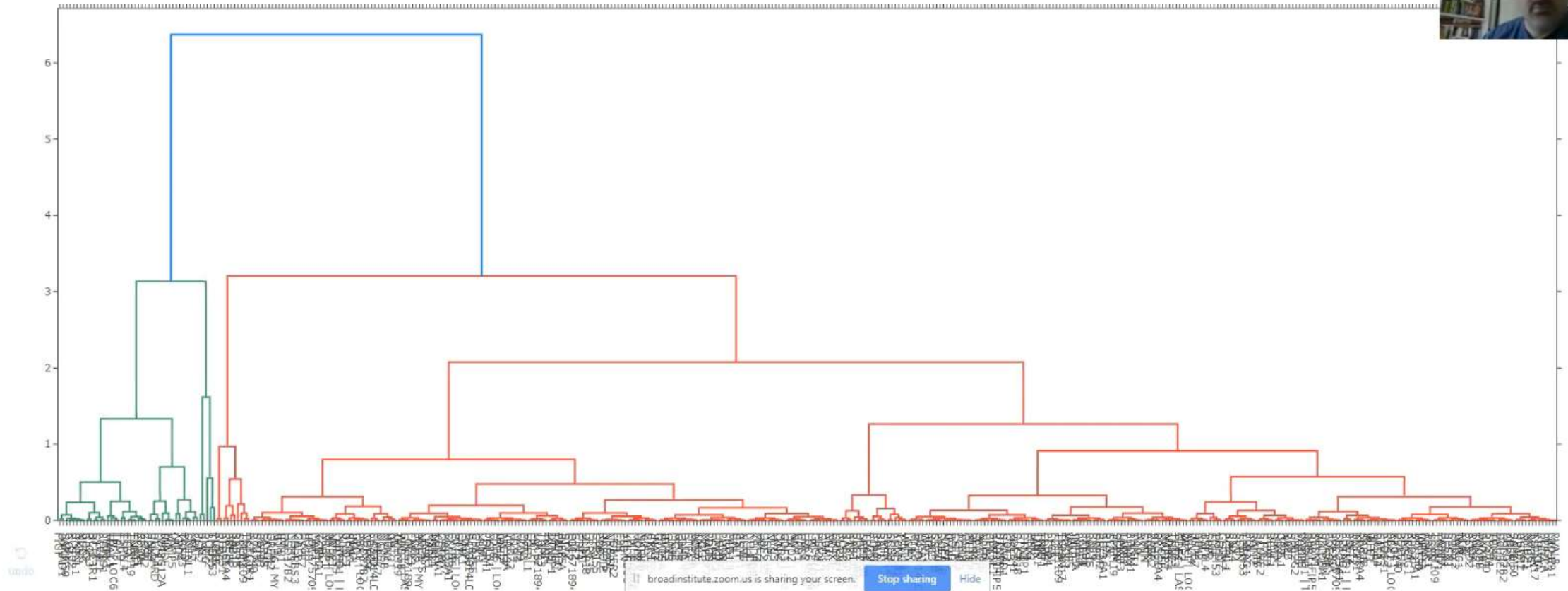
Query data	Pair plots	Volcano plots	Principle Component Analysis	Dendrograms
GeneLab Study				
Organism				
Tissue Type				
Gene Name				
Comparison Type				

GeneLab Study	Organism	Tissue Type	Gene Name	Comparison Type	Fold Change (Log2)
GLDS-111	mouse	Extensor digitorum muscle	0610042C05RIK	F vs GC	-0.4887913788371339
GLDS-111	mouse	Extensor digitorum muscle	0610042C05RIK	F vs SMG	-0.5816786559450451
GLDS-111	mouse	Extensor digitorum muscle	0610042C05RIK	SMG vs GC	0.09288727710791107
GLDS-111	mouse	Extensor digitorum muscle	1110013L07RIK	F vs GC	-0.321165860205428
GLDS-111	mouse	Extensor digitorum muscle	1110013L07RIK	F vs SMG	-0.12584214734109933
GLDS-111	mouse	Extensor digitorum muscle	1110013L07RIK	SMG vs GC	-0.1953237128643288
GLDS-111	mouse	Extensor digitorum muscle	1110019N10RIK	F vs GC	-0.4581764992642475
GLDS-111	mouse	Extensor digitorum muscle	1110019N10RIK	F vs SMG	-0.2940955606236057
GLDS-111	mouse	Extensor digitorum muscle	1110019N10RIK	SMG vs GC	-0.16408093864064185
GLDS-111	mouse	Extensor digitorum muscle	1110058L19RIK	F vs GC	0.0009118229516197414
GLDS-111	mouse	Extensor digitorum muscle	1110058L19RIK	F vs SMG	-0.016582993188499186
GLDS-111	mouse	Extensor digitorum muscle	1110058L19RIK	SMG vs GC	0.01749481614011892
GLDS-111	mouse	Extensor digitorum muscle	1810015C04RIK	F vs GC	1.112762287749198
GLDS-111	mouse	Extensor digitorum muscle	1810015C04RIK	F vs SMG	0.948963506381427
GLDS-111	mouse	Extensor digitorum muscle	1810015C04RIK	SMG vs GC	0.1637987813677712
GLDS-111	mouse	Extensor digitorum muscle	2010315L10RIK	F vs GC	0.074744107876769
GLDS-111	mouse	Extensor digitorum muscle	2010315L10RIK	F vs SMG	0.2166768235598912
GLDS-111	mouse	Extensor digitorum muscle	2010315L10RIK	SMG vs GC	-0.14193271568312243
GLDS-111	mouse	Extensor digitorum muscle	2400001E08RIK	F vs GC	-0.42270257656608934
GLDS-111	mouse	Extensor digitorum muscle	2400001E08RIK	F vs SMG	-0.23755516143965874
GLDS-111	mouse	Extensor digitorum muscle	2400001E08RIK	SMG vs GC	-0.1851474151264307
GLDS-111	mouse	Extensor digitorum muscle	2400001E08RIK LOC433216	F vs GC	-0.08367188308301321
GLDS-111	mouse	Extensor digitorum muscle	2400001E08RIK LOC433216	F vs SMG	-0.16508960464339414
GLDS-111	mouse	Extensor digitorum muscle	2400001E08RIK LOC433216	SMG vs GC	0.08141772156038099
GLDS-111	mouse	Extensor digitorum muscle	2610018G03RIK	F vs GC	-0.49107700944892374
GLDS-111	mouse	Extensor digitorum muscle	2610018G03RIK	F vs SMG	0.06138994651388862
GLDS-111	mouse	Extensor digitorum muscle	2610018G03RIK	SMG vs GC	-0.5524669559628123
GLDS-111	mouse	Extensor digitorum muscle	2610208M17RIK	F vs GC	0.16867090574058852
GLDS-111	mouse	Extensor digitorum muscle	2610208M17RIK	F vs SMG	0.10444698472176682
GLDS-111	mouse	Extensor digitorum muscle	2610208M17RIK	SMG vs GC	0.06422322101993157
GLDS-111	mouse	Extensor digitorum muscle	2700049H19RIK	F vs GC	

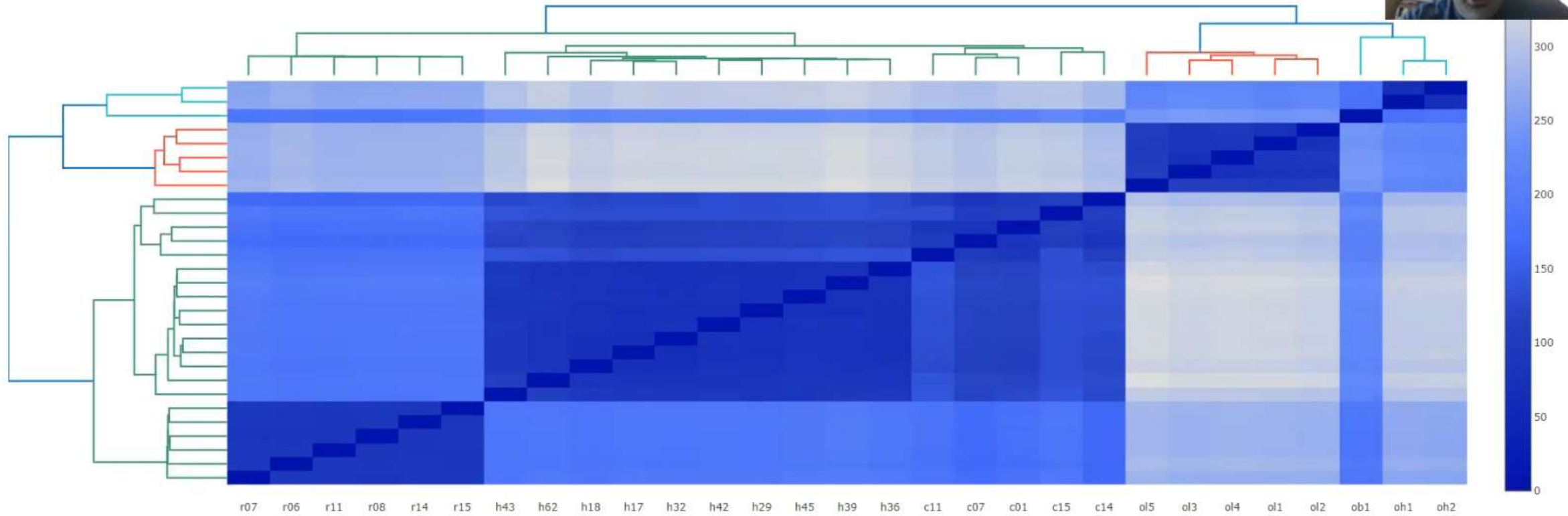


Visualization in Development: Basic Visualization examples

Query data	Pair plots	Volcano plots	Principle Component Analysis	Dendrograms
F vs GC				
Bottom dendrogram				
Hide heatmap				



Query data	Pair plots	Volcano plots	Principle Component Analysis	Dendrograms	
F vs GC					
Left dendrogram					
Hide heatmap					



Location: Boston Marriott Cambridge, 50 Broadway, Cambridge, MA 02142

Purpose: This is a workshop with all members who are involved with the VWG and are assisting GeneLab with developing our visualization capabilities. This workshop will allow all members to present what they are working on, what stage the visualization is at, and future development of the visualization on GeneLab. Lastly, this workshop will allow for communication between all VWG members and allow hands-on demos of all tools and development of how to fully integrate all visualization components with the GeneLab for the next phase of development.

Each presentation from VWG members should be interactive and should include the following in their presentations:

1. Brief overview of the tool and platform. Should include in your presentation the architecture of your platforms, demo on how to use it, and any technical hurdles that we can discuss.
2. Demo on the tool. If possible please provide a way that the VWG members might be able to use the tool on their own computers.
3. The code for your tool made available earlier or during the presentation. If you want you can post your tool on our GitHub page (<https://github.com/abehesht/GeneLab-Visualization>)
4. Also allow for time during your presentation to have people interact with you and allow for brainstorming on how your tool can easily be integrated into the overall visualization flow that we will implement on GeneLab.

Agenda for Monday, 4/22/19:

8:00am – 8:40am	Introduction & GeneLab visualization: Sylvain and Afshin
8:40am – 9:10am	GeneLab Data Systems framework and deployment: Dan Berrios
9:10am – 9:30am	GeneLab Metadata & Processed data: Sam Gebre
9:30am – 10:30am	Cyverse: Sarah Roberts and Nirav Merchant
10:30am – 10:45am	Coffee break
10:45am – 11:45am	API Presentation and Discussion (GeneLab API wrapper: Kirill Grigorev)
11:45am – 12:00pm	Framework summary
12:00pm – 1:30pm	Lunch (on your own)
1:30pm – 2:15pm	GenePattern Notebook: Ted Lefield
2:15pm – 2:45pm	Interactive Plots for GeneLab: Rien Dijkstra
2:45pm – 3:30pm	Interactive Visualization Tools for GeneLab: Jess Alfredsen, Sebastian Steensson
3:30pm – 3:45pm	Coffee break
3:45pm – 4:45pm	Network tool: Daniel Couch
4:45pm – 5:00pm	Discussion and wrap up for the day

Agenda for Tuesday, 4/23/19:

- 8:00am – 8:10am Other visualization tools Introduction:
Afshin Beheshti
- 8:10am – 9:00am TOAST/MANGO: Richard Barker,
Simon Gilroy
- 9:00am – 10:30am Interactive discussion on how to
integrate the API with Basic
Visualization tools
- 10:30am – 10:45am Coffee break
- 10:45am – 12:30pm Interactive discussion on how to
integrate all visualization tools together
in one package
- 12:30pm – 2:00pm Lunch (on your own)
- 2:00pm – 5:00 pm Working Session
- 5:00pm – 5:50pm Break before going to Broad Institute
- 5:50pm – 5:55pm Meet on Hotel lobby and walk over to
Broad

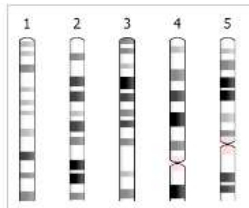
Tuesday Evening meetup with the Broad Institute at the Broad Institute:

- Brief Presentations from GeneLab and VWG:
- 6:00pm – 6:10pm Afshin Beheshti: Brief introduction of who
we are at GeneLab and a very brief
presentation of what we do and why are
holding this meetup
- 6:10pm – 6:15pm Rien Dijkstra
- 6:15pm – 6:20pm Jess Alfredsen, Sebastian Steensson
- 6:20pm – 6:25pm Cyverse: Sarah Roberts and Nirav
Merchant
- 6:25pm – 6:30pm TOAST/MANGO: Richard Barker
- Brief presentations from people at Broad Institute for potential
future visualization collaborations:
- 6:30pm – 6:40pm Eric Weitz, Broad Institute: Ideogram
- 6:50pm – 7:00pm Jen Lapan, Broad Institute: Data
Donation Platform
- 7:00 pm – 7:30pm Social interaction between everyone

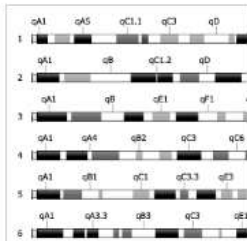
Overview | Ideogram

Ideogram.js is a JavaScript library for chromosome visualization. Ideogram supports drawing and animating genome-wide datasets.

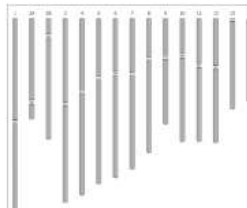
In the examples below, genomic data is fetched from servers and rendered on the client using D3.js.



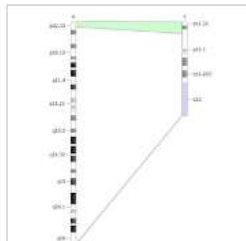
Human
Chromosomes 1-22, X and Y rendered vertically.



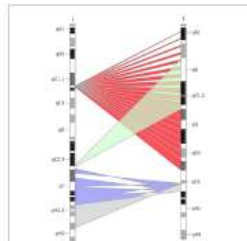
Mouse
Chromosomes 1-19, X and Y rendered horizontally, with labeled bands.



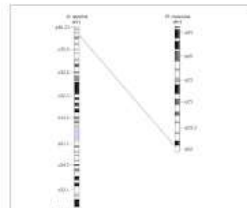
Eukaryotes
Genomes for many organisms, using data fetched via API.



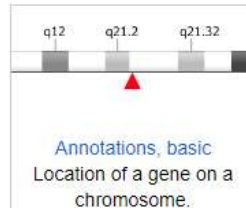
Homology, basic
Simple relations between two chromosomes.



Homology, advanced
Relations between two chromosomes, using color and opacity.



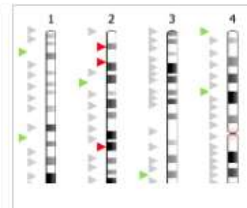
Homology, interspecies
Relations between two chromosomes from a different taxon.



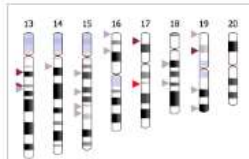
Annotations, basic
Location of a gene on a chromosome.



Annotations, overlaid
Overlaid features on chromosomes.



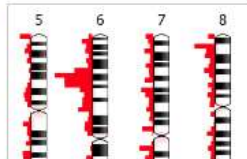
Annotations, tracks
Stacked features on chromosomes.



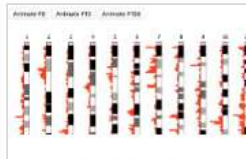
Annotations, external data
Data fetched from a third party API, annotated by shape and color.



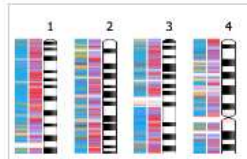
Annotations, file URL
Features parsed from standard file formats, e.g. BED, loaded via URL.



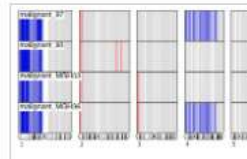
Annotations, histogram
Use shape to show feature distributions on chromosomes.



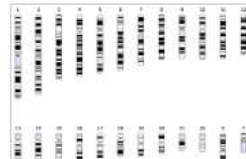
Annotations, animated
Genome-wide transitions in expression over time.



Annotations, heatmap
Use color to show feature distributions on chromosomes.



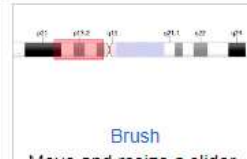
Geometry, collinear
All chromosomes arranged in one line.



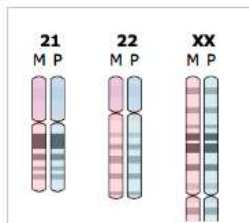
Layout, small
Chromosomes surrounded by other content.



Layout, tabs
Tab containers and scrolling for large ideograms.



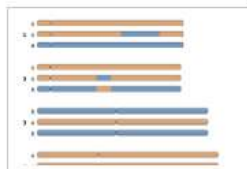
Brush
Move and resize a slider-like widget to change genomic coordinates in focus.



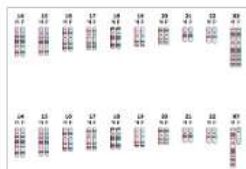
Ploidy, basic
Diploid genome of a



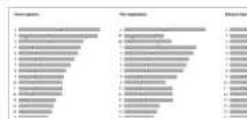
Ploidy, rearrangements
Loss and gain of cytogenetic arms and chromatids.



Ploidy, recombination
Gene conversion and double chromosome crossover among



Multiple, primates
Three primate genomes, with orthologs annotated for comparison.



Multiple, primates
Three primate genomes, with orthologs annotated for comparison.

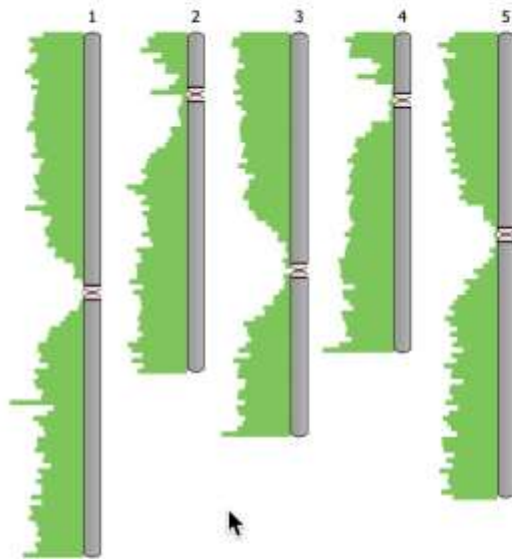


Eric Weitz, PhD

Plant gene expression in space | Ideogram

[Overview](#) | [Previous](#) | [Next](#) | [Source](#)

This ideogram shows the distribution of all *Arabidopsis thaliana* genes throughout the genome. Explore how this plant's genes changed expression while in outer space by applying filters below.



Gene type

- Protein coding
- ncRNA
- lncRNA
- tRNA
- rRNA
- snRNA
- snoRNA

Significant expression change

- Yes
- No

Broad's engineering effort to advance ubiquitous, global, direct-to-participant genomic research. DDP is in a learning-and-prep mode now, anticipating a broad range of user demands to visualize ingested D2P data and eventual, analytical outputs

The Data Donation Platform (DDP): A software stack that enables direct participant engagement, including consent and recontact, via intuitive web and mobile interfaces. DDP provides the underlying infrastructure for disease-specific registries such as the Angiosarcoma Project, the Rare Genomes Project, and the Global A-T Family Data Platform.

BROAD INSTITUTE Careers · Blog · Search

ABOUT US · PEOPLE · SCIENCE · DATA AND TOOLS · CENTERS · COMMUNITY · CONTACT · NEWS AND MEDIA

HOME » SCIENCE » TECHNOLOGY AREAS

DATA SCIENCES PLATFORM

The life sciences are in the midst of a data revolution. Cheap and accurate genome sequencing is a reality, advanced imaging is routine, and clinical data is increasingly stored in electronic formats. These innovations — and the massive data sets they produce — have brought us to the threshold of a new era in medicine, one where the data sciences hold the potential to propel our understanding and treatment of human disease.

The Broad Data Sciences Platform (DSP) is a methods development and software engineering group dedicated to maximizing the impact of the data sciences on the life sciences. DSP engineers, analysts, and designers build applications and capabilities to serve the Broad and beyond.

The DSP is organized around four principal components:

AREAS OF FOCUS

- [Workbench](#)
- [Analytical Tools](#)
- [User Interfaces](#)
- [Production Data Processing](#)

SPEAKER SERIES

[Leaders in Data Science II DSP](#)



Jen Lapan, PhD