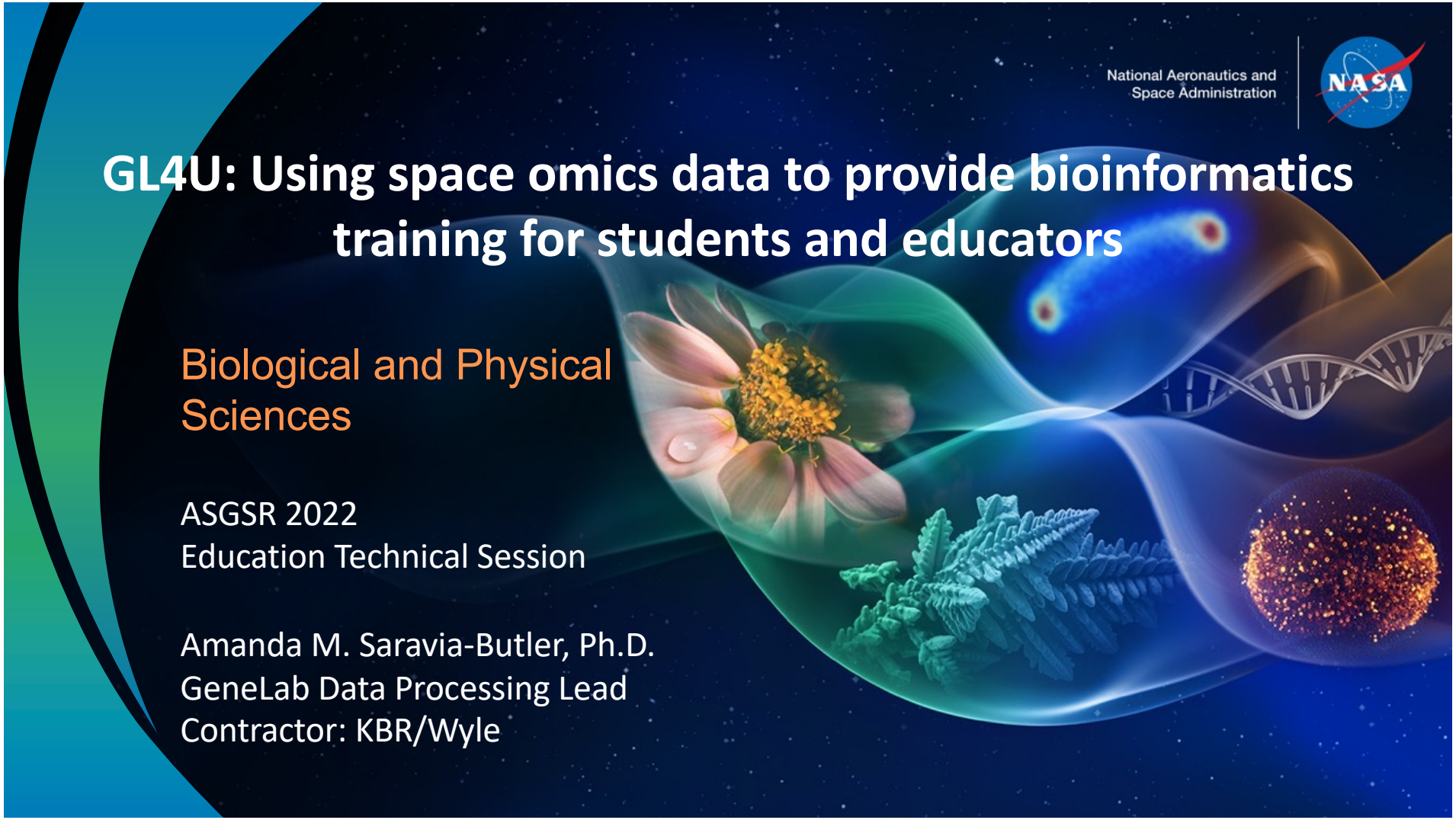


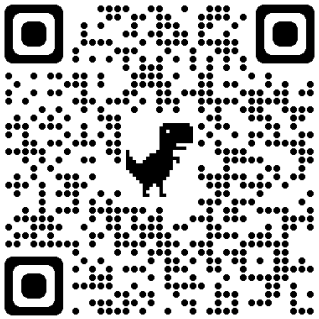
GL4U: Using space omics data to provide bioinformatics training for students and educators

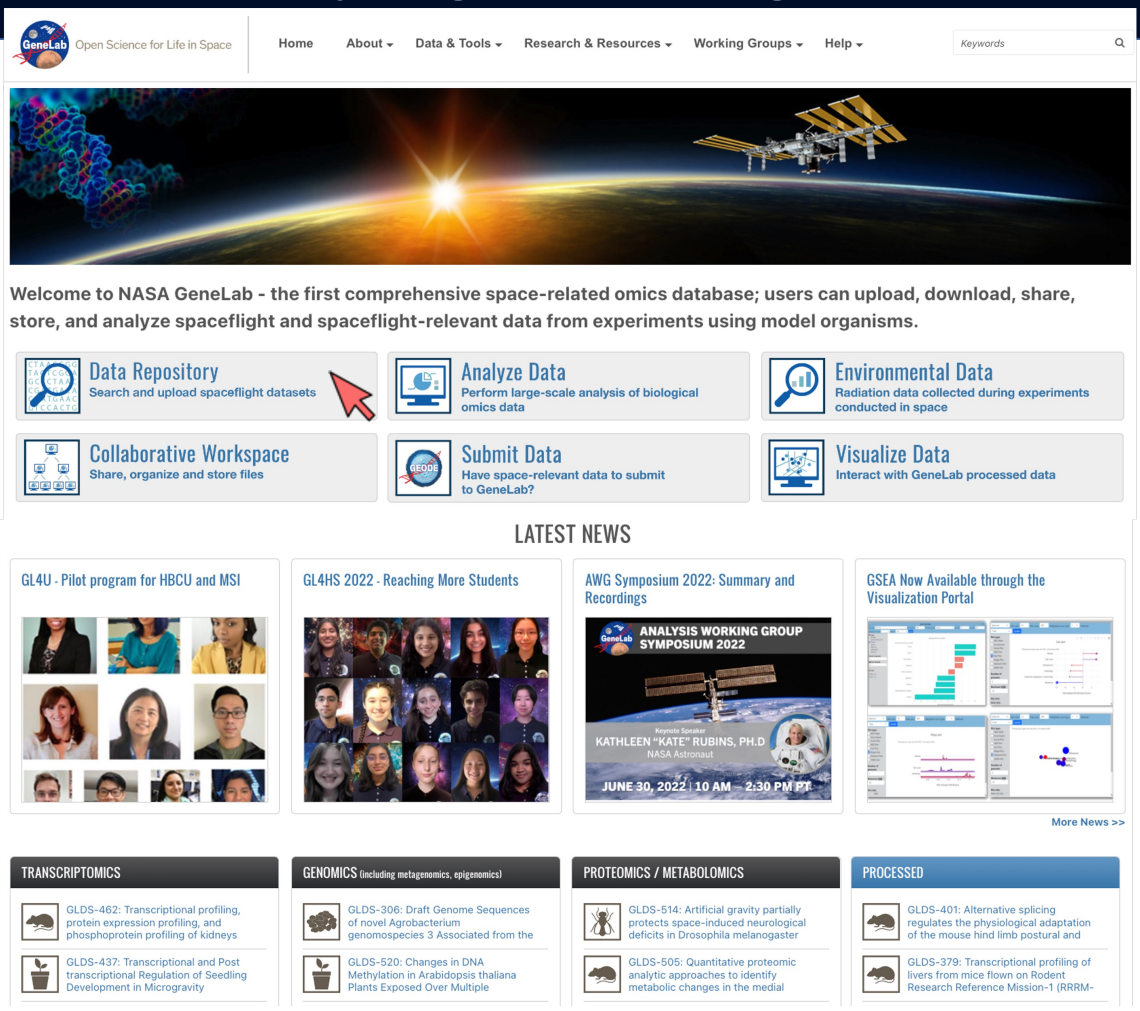
Biological and Physical Sciences

ASGSR 2022
Education Technical Session

Amanda M. Saravia-Butler, Ph.D.
GeneLab Data Processing Lead
Contractor: KBR/Wyle

https://genelab.nasa.gov/





GeneLab Data Repository

Open Science for Life in Space

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

GeneLab Data Repository Search

~390 Datasets, 135TB

Search Datasets

Items per page: 25 | 1 – 25 of 39

Search Filters

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

GeneLab Search Filters

Project Type
- [ ] Ground
- [ ] Spaceflight
- [ ] High Altitude

Assay Type
- [ ] Microarray
- [ ] RNA Sequencing (RNA-Seq)
- [ ] Single-Cell RNA Sequencing
- [ ] miRNA Sequencing (miRNA-Seq)
- [ ] Nucleotide Sequencing

Show more

Tissue
- [ ] Cells
- [ ] Root
- [ ] Seedlings
- [ ] Liver
- [ ] Leaf

Show more

Factor
- [ ] Spaceflight
- [ ] Ionizing Radiation
- [ ] Genotype
- [ ] Time
- [ ] Treatment

Show more

GLDS-524

Ionizing radiation induces transgenerational effects of DNA methylation in zebrafish

Organisms	Factors	Assay Types	Release Date	Description
Danio rerio	Ionizing Radiation Generation	DNA methylation profiling	31-Aug-2023	Ionizing radiation is known to cause DNA damage, yet the mechanisms underlying potential tr exposure have been scarcely studied. Previously, we observed effects in offspring ...

GLDS-520

Changes in DNA Methylation in Arabidopsis thaliana Plants Exposed Over Multiple Generations to Gamma Radiation

Organisms	Factors	Assay Types	Release Date	Description
Arabidopsis thaliana	Generation Ionizing Radiation	DNA methylation profiling	31-Jan-2021	Previous studies have found indications that exposure to ionising radiation (IR) results in DNA plants. However, this phenomenon is yet to be studied across multiple generations...

GLDS-519

PUCHI represses early meristem formation in developing lateral roots of Arabidopsis thaliana

Organisms	Factors	Assay Types	Release Date	Description
Arabidopsis thaliana	Genotype Organism Part Treatment Time	transcription profiling	25-Mar-2022	Time-series RNAseq analysis following lateral root induction by gravistimulation. Lateral root o process in the development of a plant's root system and its adaptation to the envi...

GLDS-514

Artificial gravity partially protects space-induced neurological deficits in Drosophila melanogaster

Organisms	Factors	Assay Types	Release Date	Description
Drosophila melanogaster	Sex Spaceflight Altered Gravity	transcription profiling protein expression profiling	08-Sep-2022	Spaceflight poses risks to the central nervous system (CNS), and understanding neurological future missions. We report CNS changes in Drosophila aboard the International Spa...

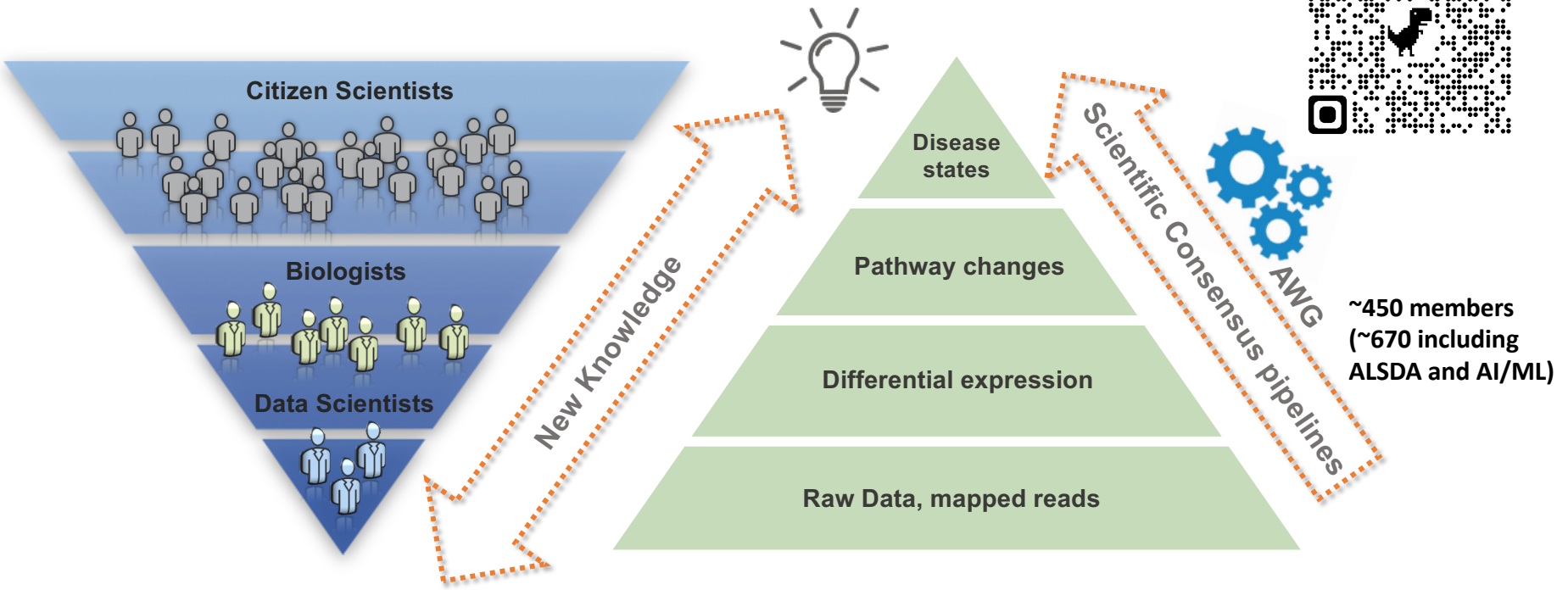
GLDS-512

Transcriptional profiling of adrenal glands from mice flown on the RR-23 mission

Organisms	Factors	Assay Types	Release Date	Description
Mus musculus	Spaceflight Ionizing Radiation	transcription profiling	03-Aug-2022	The objective of the Rodent Research-23 missions (RR-23) was to better understand the effec specifically on the structure and function of the arteries, veins, and lymphat...

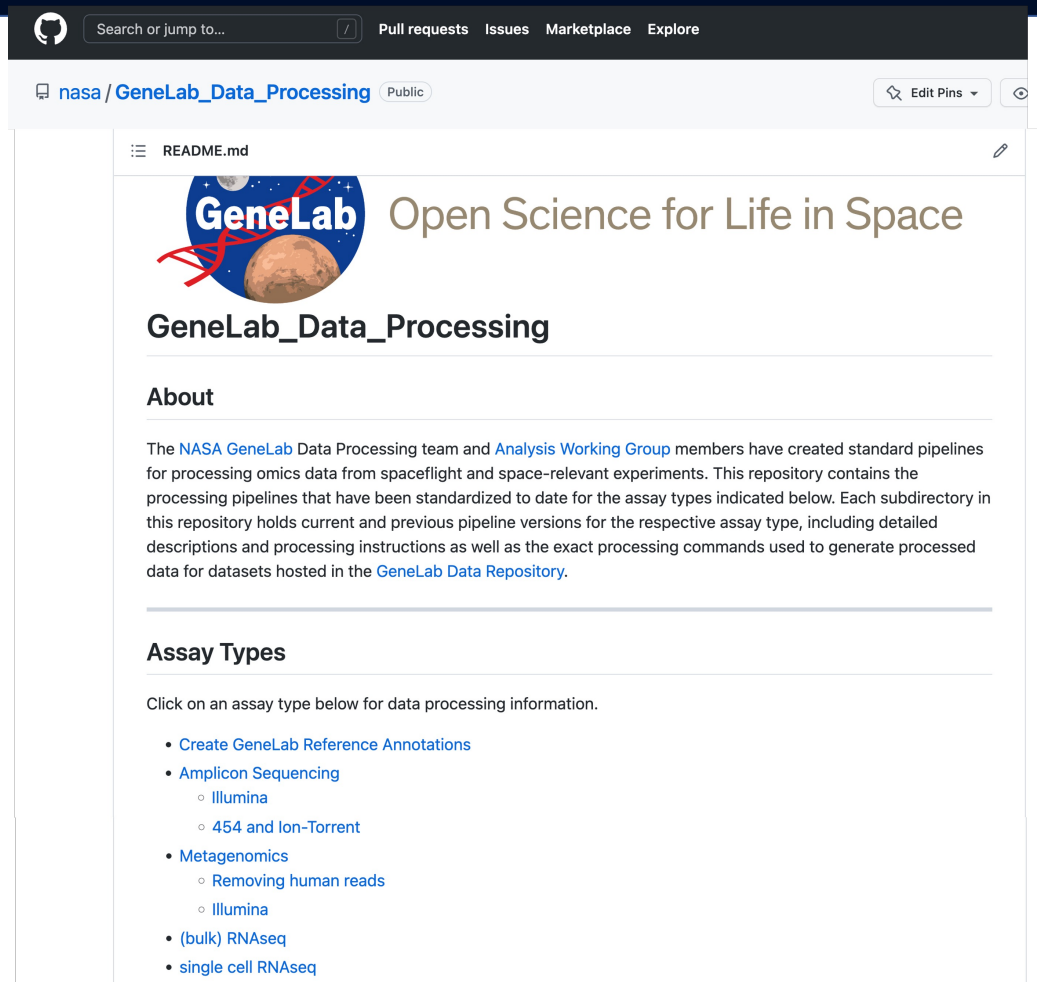
GeneLab Data Democratization

https://genelab.nasa.gov/awg/join



~450 members
(~670 including
ALSDA and AI/ML)

GeneLab Data Processing GitHub Repo



https://github.com/nasa/GeneLab_Data_Processing

nasa / GeneLab_Data_Processing Public

Edit Pins | Watch 5 | Fork 27 | Star 34

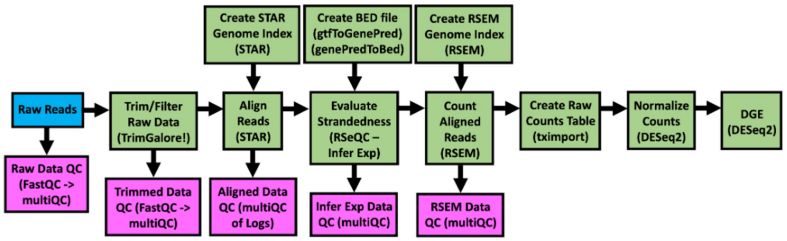
Code | Issues | Pull requests | Actions | Projects | Wiki | Security | Insights

master | GeneLab_Data_Processing / RNAseq / | Go to file | Add file | ...

README.md

GeneLab bioinformatics processing pipeline for Illumina RNA-sequencing data

> The document `GL-DPPD-7101-F.md` holds an overview and example commands for how GeneLab processes RNAseq datasets. See the Repository Links descriptions below for more information. Processed data output files and a GeneLab data processing summary is provided for each GLDS dataset in the GeneLab Data Systems (GLDS) repository.



Repository Links

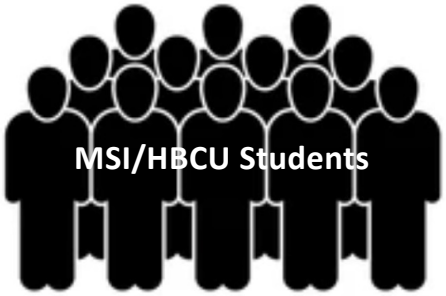
- Pipeline_GL-DPPD-7101_Versions
  - Contains the current and previous GeneLab RNAseq consensus processing pipeline (RCP) versions documentation
- RNAseq_Tool_Install
  - Contains instructions and associated yaml files to install the conda environments containing tools used to process RNAseq datasets using the GeneLab RCP, versions A-E
- GeneLab_Reference_and_Annotation_Files
  - Contains links to the fasta and gtf annotation files used to process RNAseq datasets using the GeneLab RCP, versions A-E
    > Note: Reference annotation files used from RCP version F and onward are documented in GeneLab_Reference_Annotations
- Workflow_Documentation
  - Contains instructions for installing and running the GeneLab RCP workflow
    > Note: The RCP workflow was implemented beginning with RCP version F
- GLDS_Processing_Scripts
  - Contains the exact processing commands and RCP version used for specific GLDS datasets that have been released and are available in the GeneLab Data Systems (GLDS) repository

GeneLab for Colleges and Universities (GL4U)

- GL4U provides space biology-relevant training in bioinformatics to the next generation of scientists through direct and indirect approaches
- The GeneLab team plans to host two annual data processing bootcamps in which participants will learn how to analyze GeneLab's space-relevant omics data types
  - Direct training for college-level students
  - Indirect training for college-level students by training college educators
- During the bootcamp, participants will be introduced to NASA, Space Biology, and GeneLab, and be provided training in experimental design and analysis of omics data using GeneLab's standard processing pipelines
  - Training will consist of a mix of lectures, hands on data processing and analysis via Jupyter Notebooks (JNs), assignments, and questions to gauge understanding and comprehension
  - Educators will receive materials and training to enable them to run the bootcamp at their home institutions or alternatively to adapt the content to implement within existing courses, thereby extending the reach of this initiative

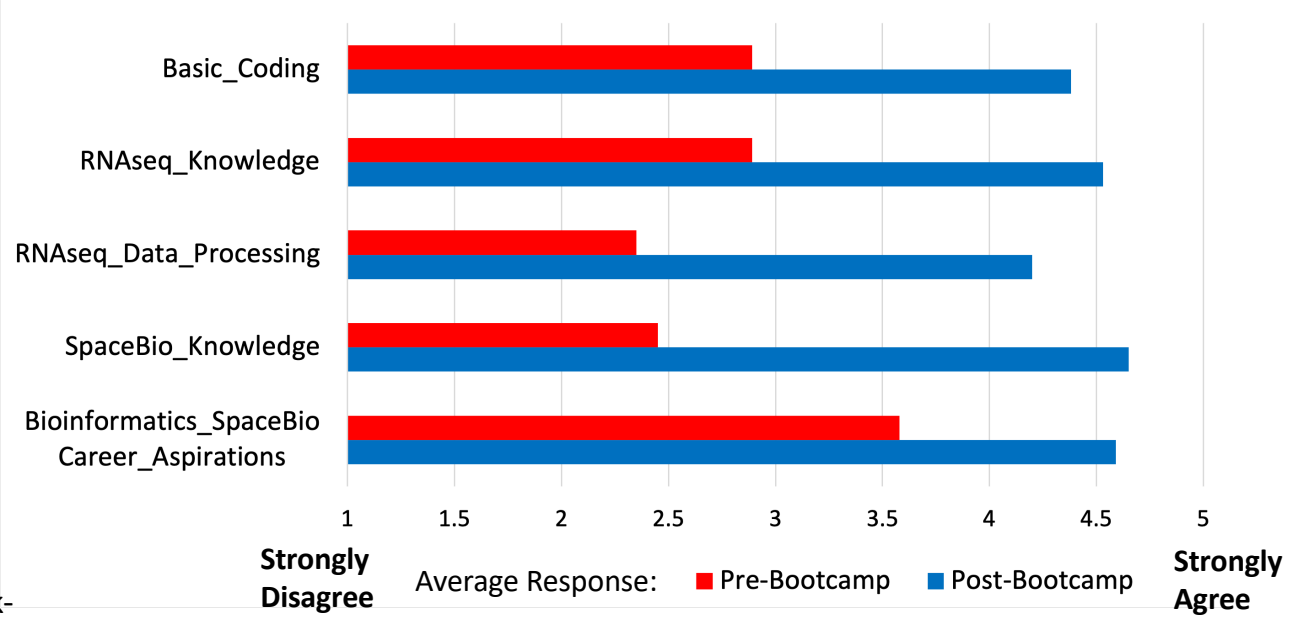
Using GeneLab Data To Train The Next Generation Of Bioinformaticians

GL4U Direct Approach

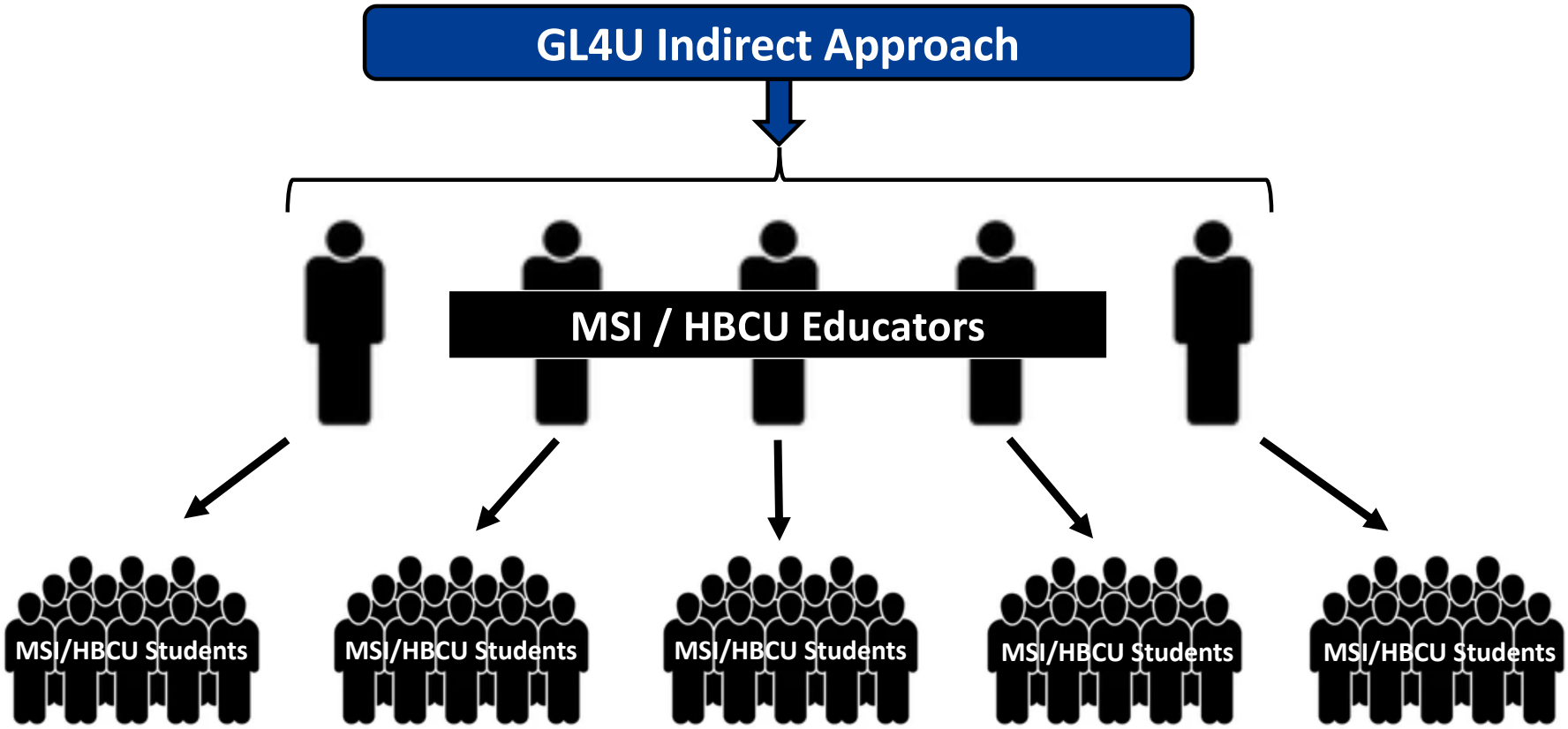


- Pilot was conducted in June 2021 in collaboration with USRA and SJSU
- SJSU students participated in a week-long bootcamp consisting of space biology-specific lectures and hands-on instruction using Jupyter Notebooks to analyze RNA sequence data

Direct Pilot Survey Results, SJSU

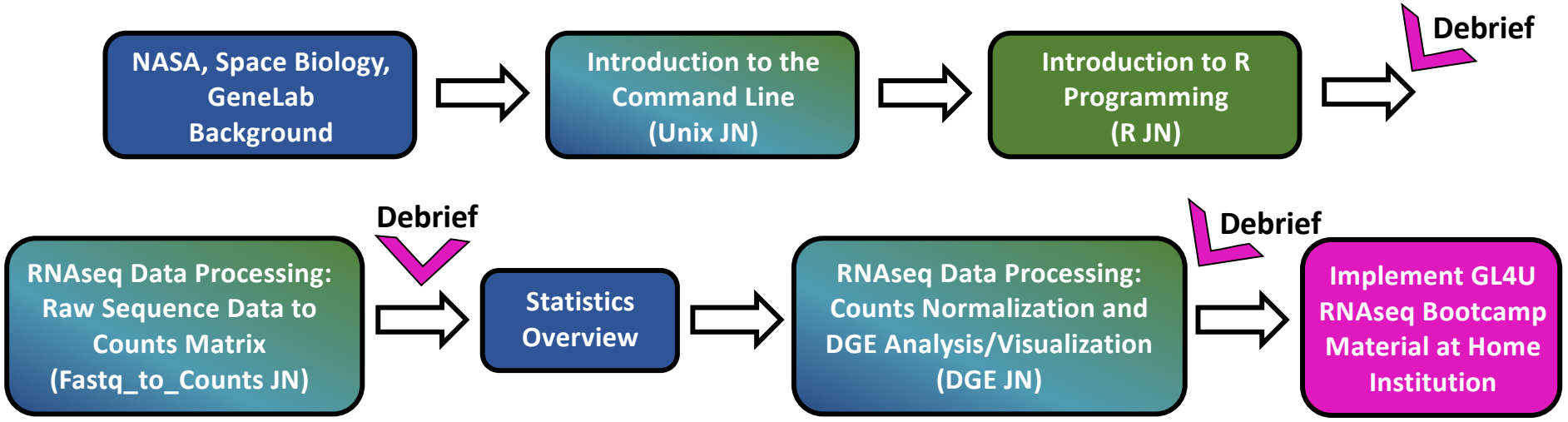


Using GeneLab Data To Train The Next Generation Of Bioinformaticians



Pilot GL4U: RNAseq Educator Bootcamp

- The GL4U indirect (educator) training pilot program was conducted in June 2022 in collaboration with Jet Propulsion Laboratory's (JPL) Planetary Protection Center of Excellence
- During the ~1.5-week long bootcamp, 6 professors and 4 graduate students from 4 HBCUs/MSIs received training via lectures and hands-on instruction using Jupyter Notebooks (JNs) on how to 1) analyze GL RNAseq data and 2) run the RNAseq bootcamp for students at their home institutions



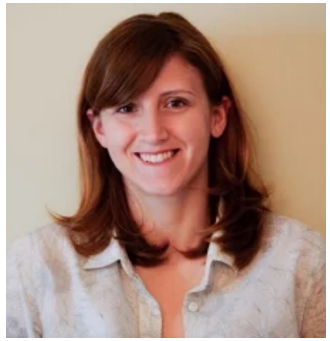
Compute resources compliments of the NASA Center for Climate Simulation (NCCS) Science Managed Cloud Environment (SMCE)

Pilot GL4U: RNAseq Educator Bootcamp Participants

Dr. Tyesha Farmer
Assistant Professor of Genetics
Alabama A&M University

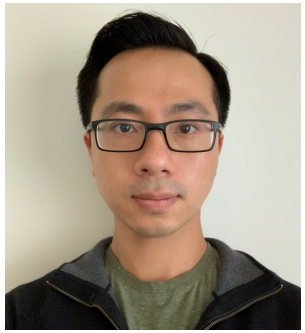
Dr. Elba Serrano
Regents Professor of Biology
New Mexica State University

Dr. Keneshia Johnson
Assistant Professor of Chemistry
Alabama A&M University



Dr. Rachel Mackelprang
Professor of Microbiology
California State University, Northridge

Dr. Wei-Jen Lin
Professor in the Dept. of Biological Sciences
California State Polytechnic University



Dr. Jason Ear
Assistant Professor in the Dept. of Biological Sciences
California State Polytechnic University

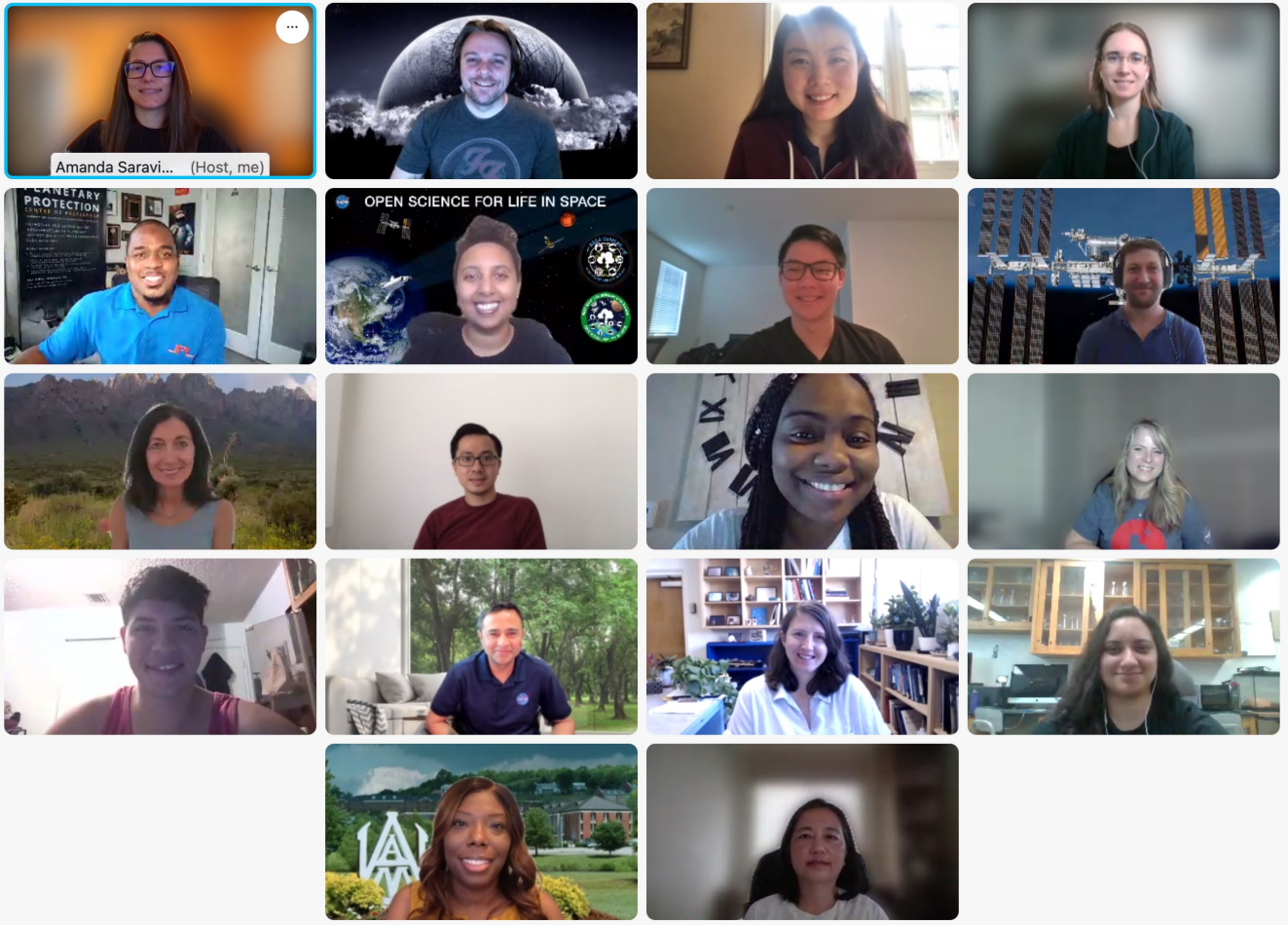
Dr. Joel Steele
Fulbright Future Scholar, Proteomics
Monash University

Chiefe Mo
Master's Student, Biology
California State Polytechnic University

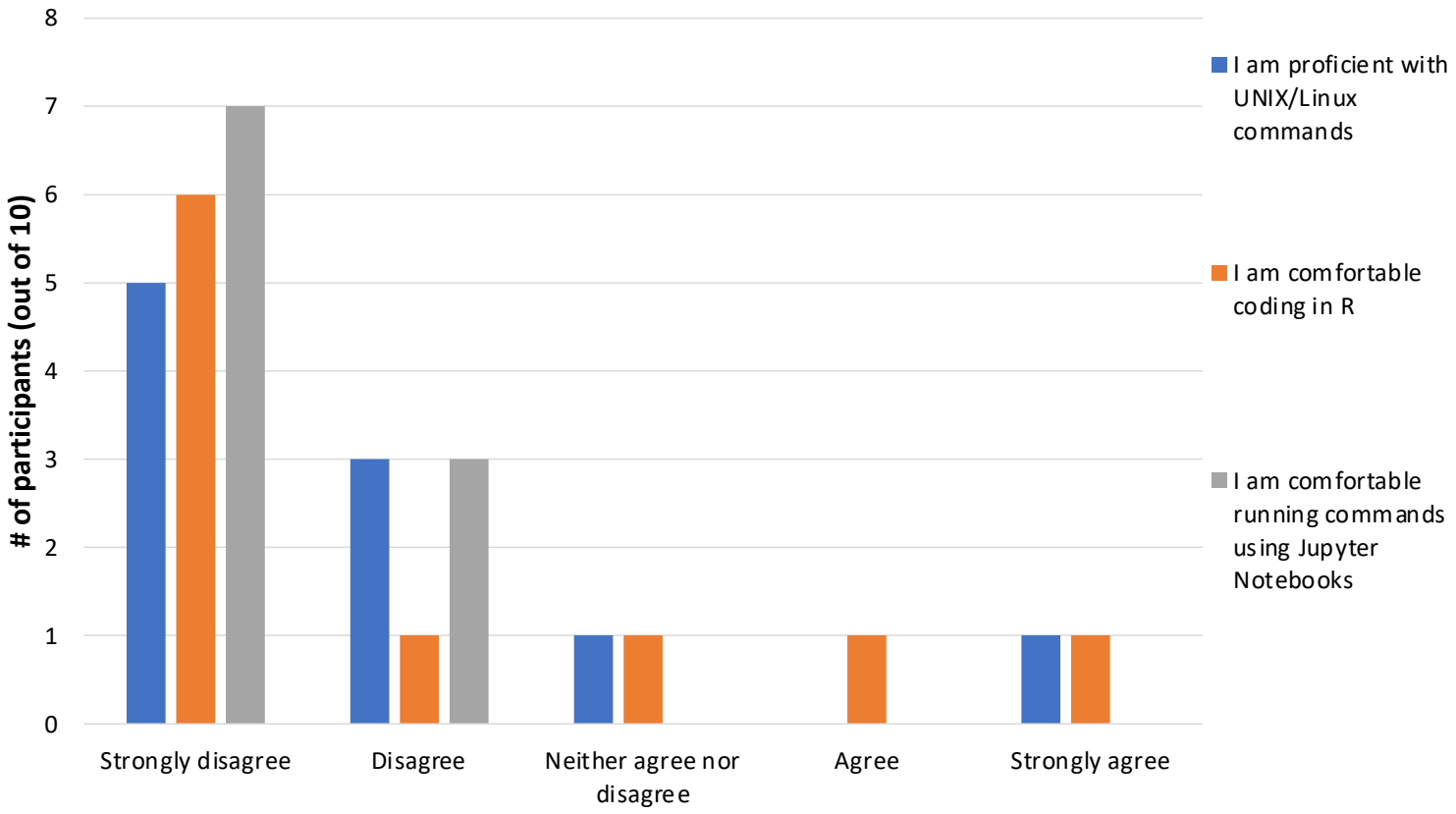
Suzi Arzoumanyan
Teaching Associate, Biology
California State University, Northridge

Mark Ortiz
Master's Student, Biology
California State Polytechnic University

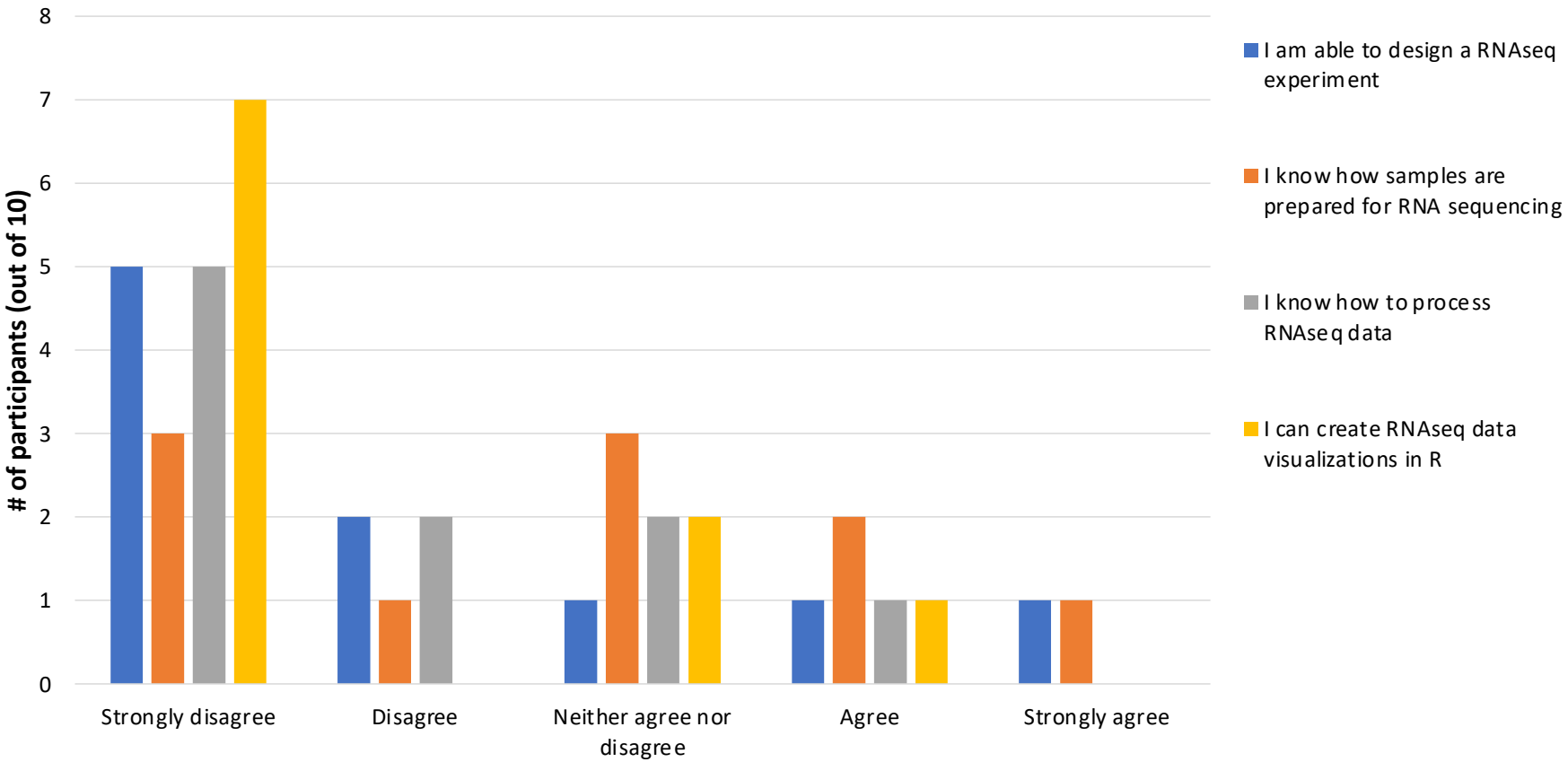
Pilot GL4U: RNAseq Educator Bootcamp Participants



Pre-Survey Results: Basic Coding

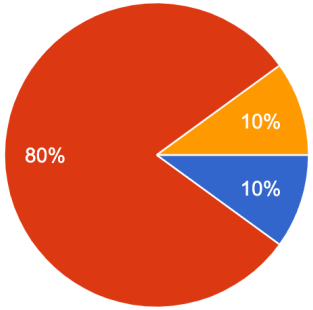


Pre-Survey Results: RNAseq Knowledge

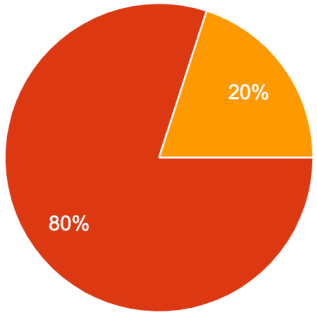


Pre-Survey Results: Home Institution Curriculum

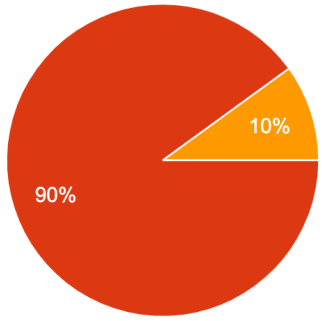
I currently teach UNIX/Linux commands in one (or more) of my courses



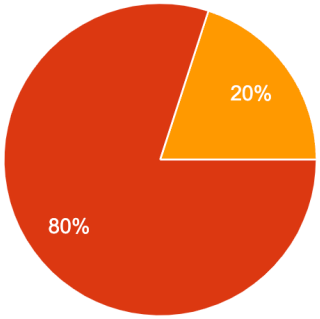
I currently teach R in one (or more) of my courses



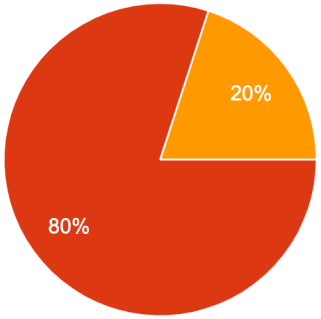
I currently use Jupyter in one (or more) of my courses



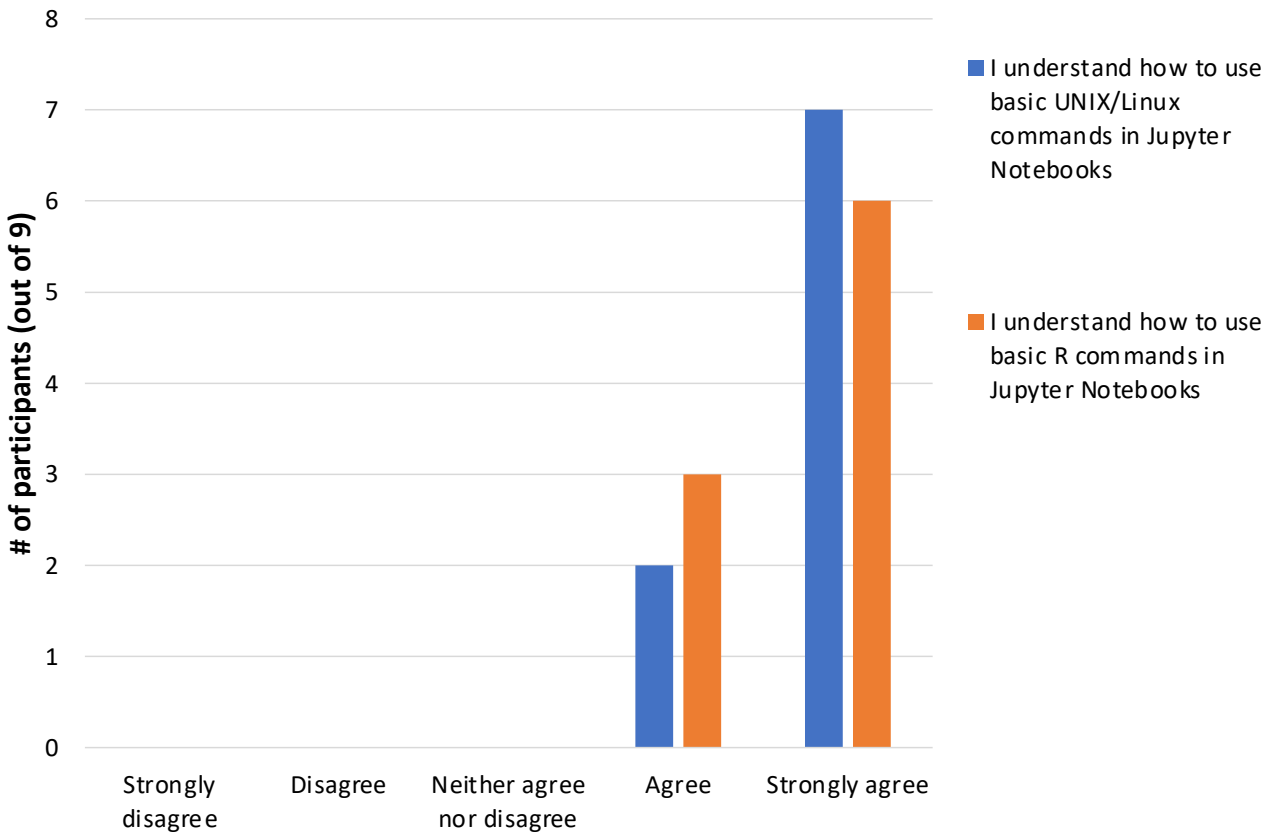
I currently teach a bioinformatics course



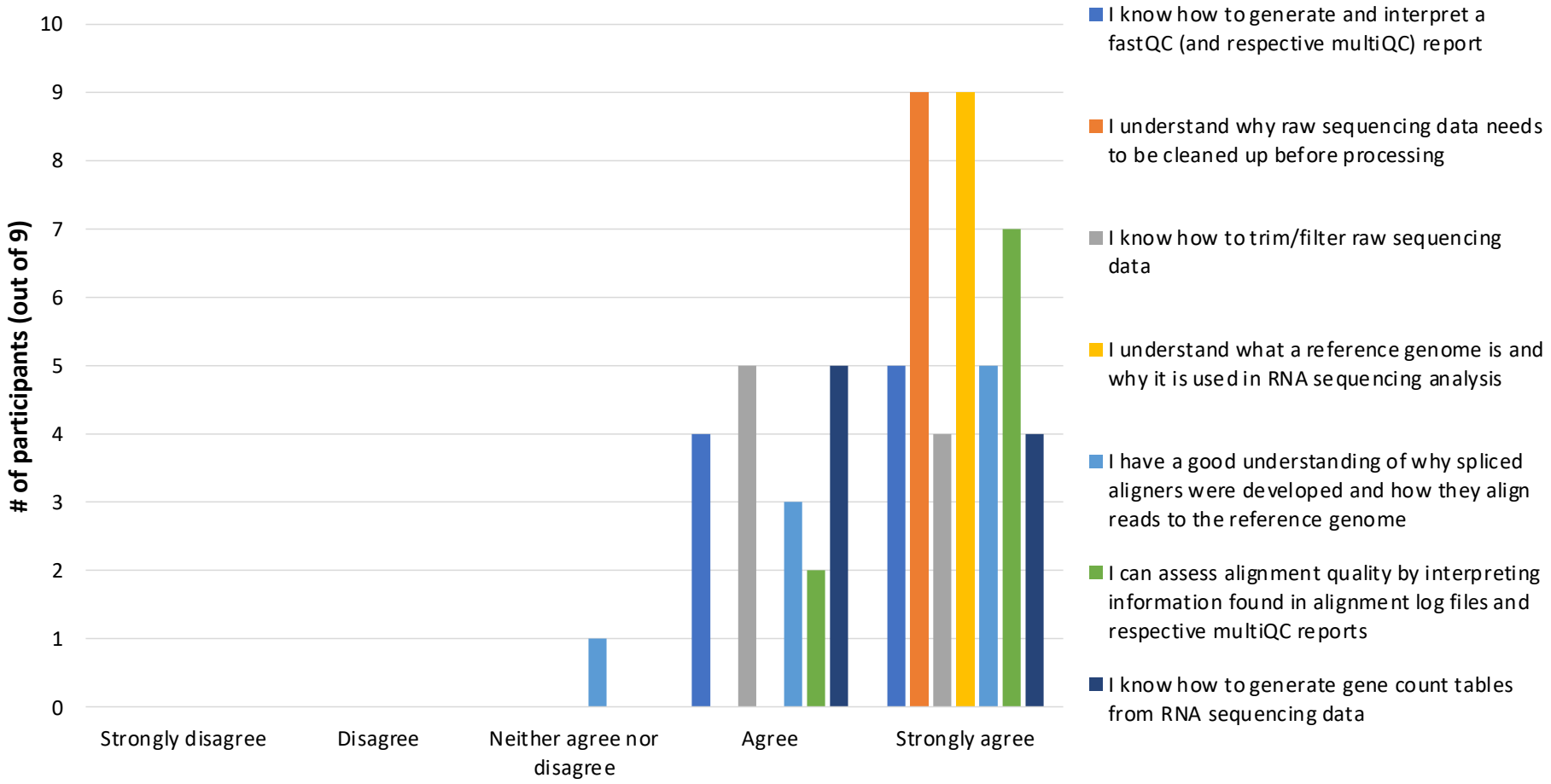
I currently teach RNA sequence analysis in one (or more) of my courses



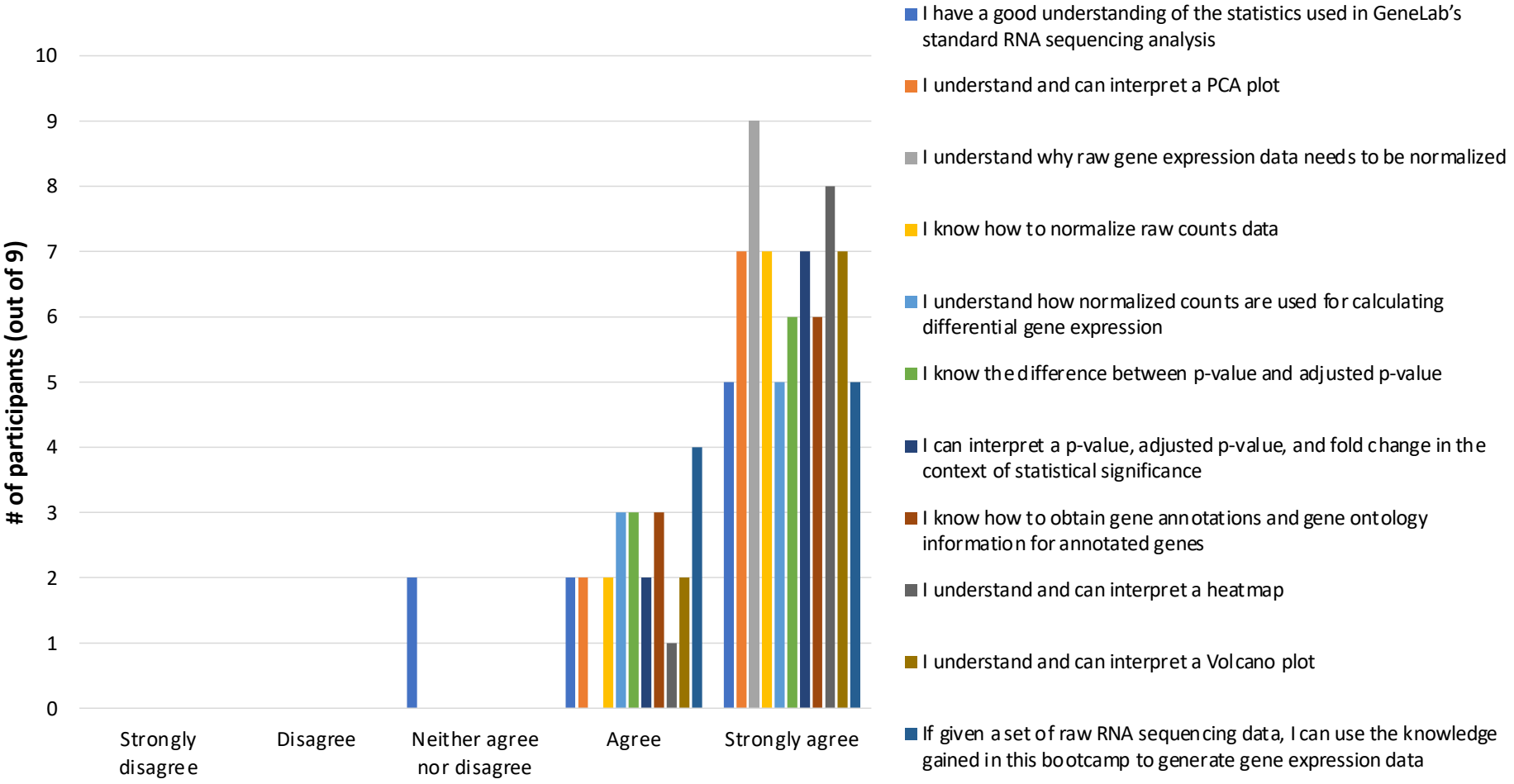
Post-Survey Results: Basic Coding



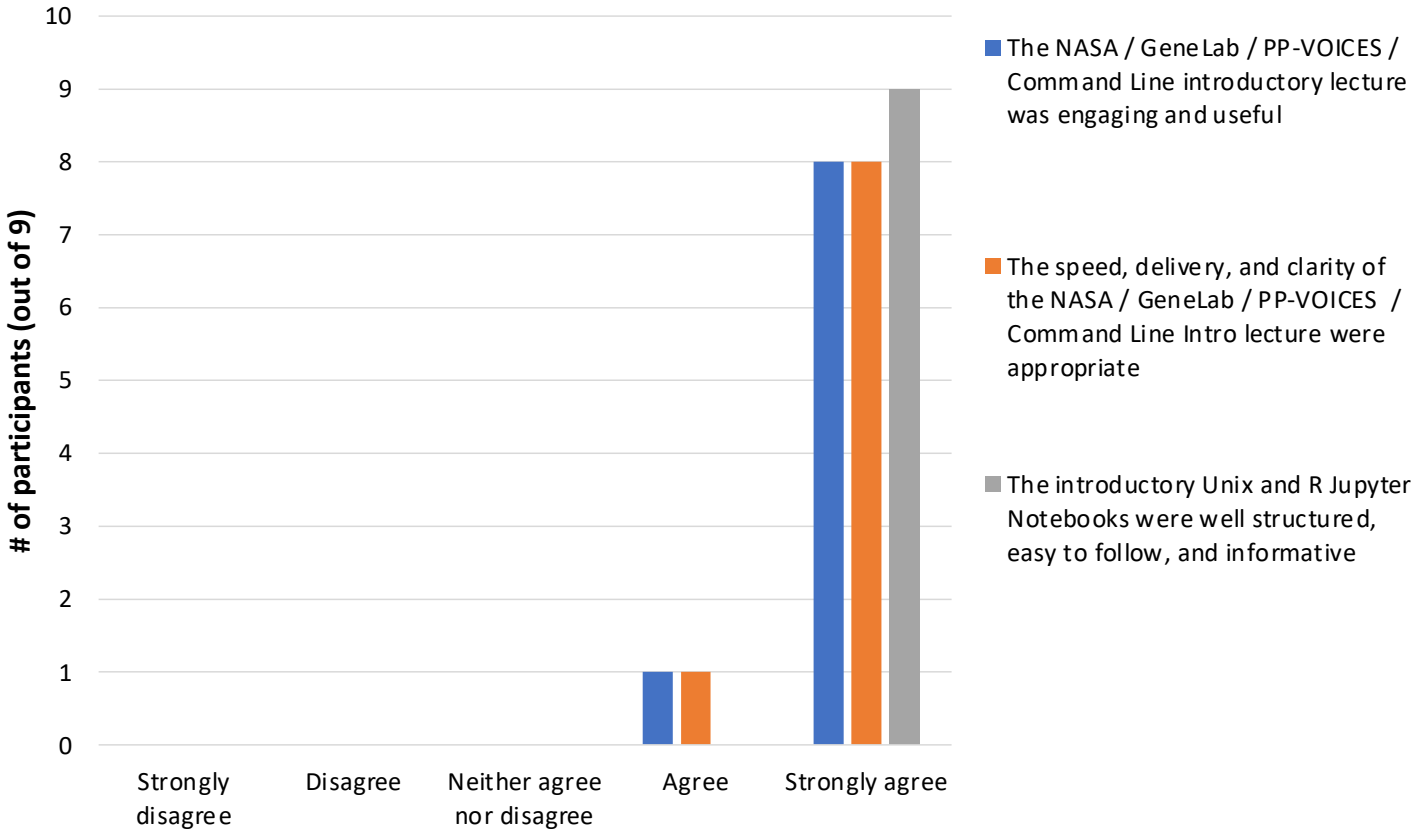
Post-Survey Results: RNAseq (Fastq to Counts)



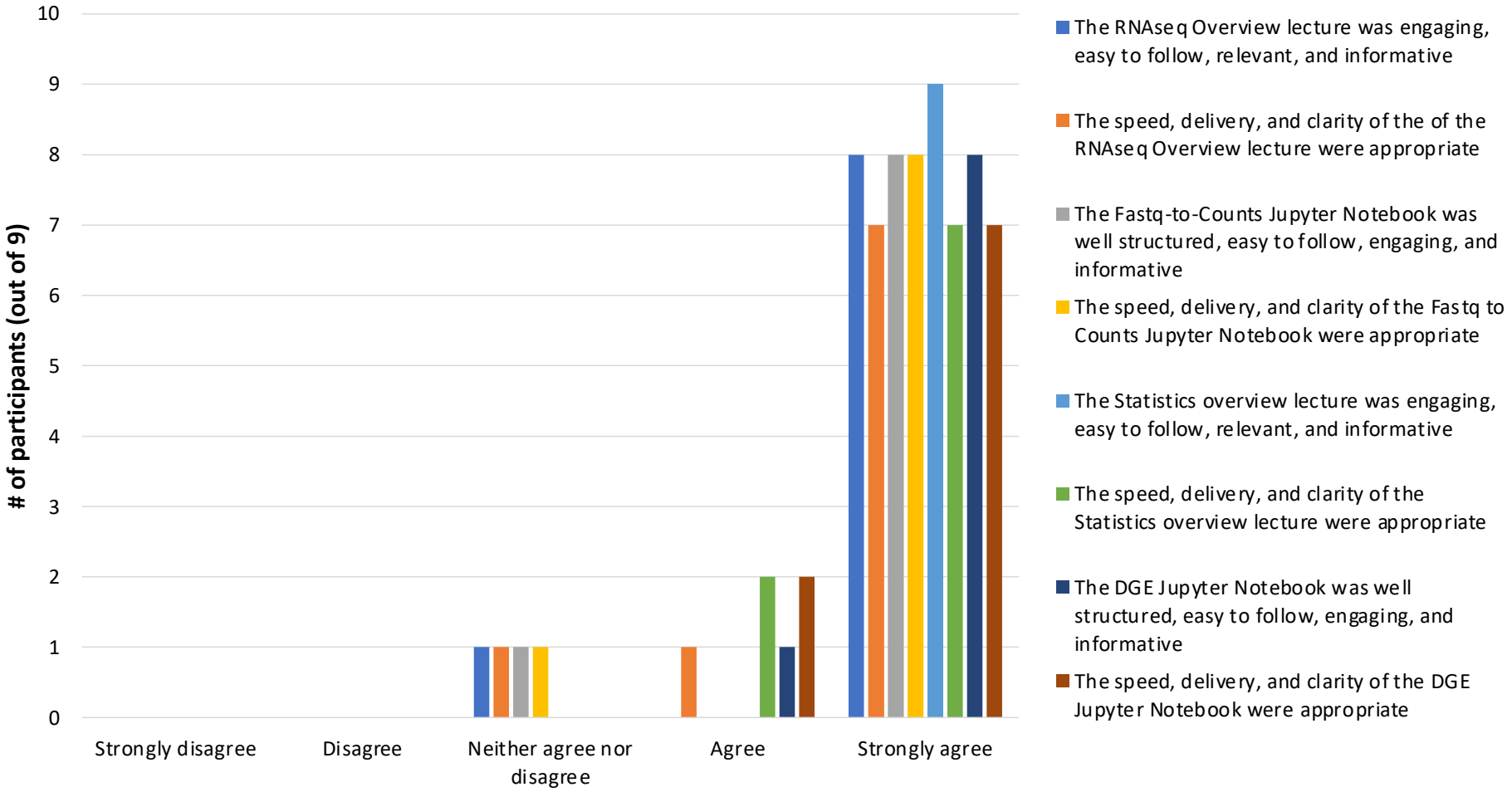
Post-Survey Results: RNAseq (DGE and Visualizations)



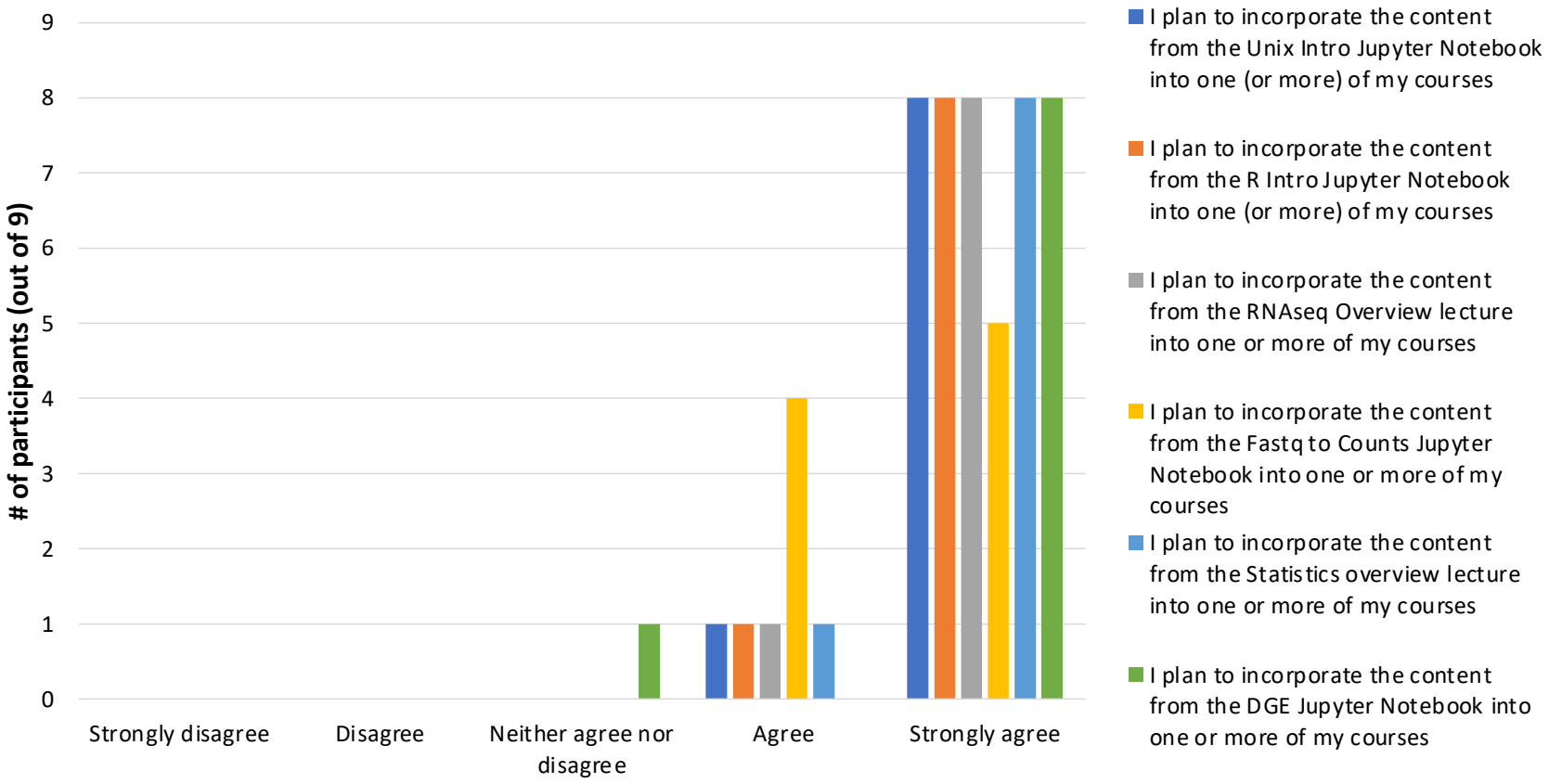
Post-Survey Results: Intro/Coding Content Feedback



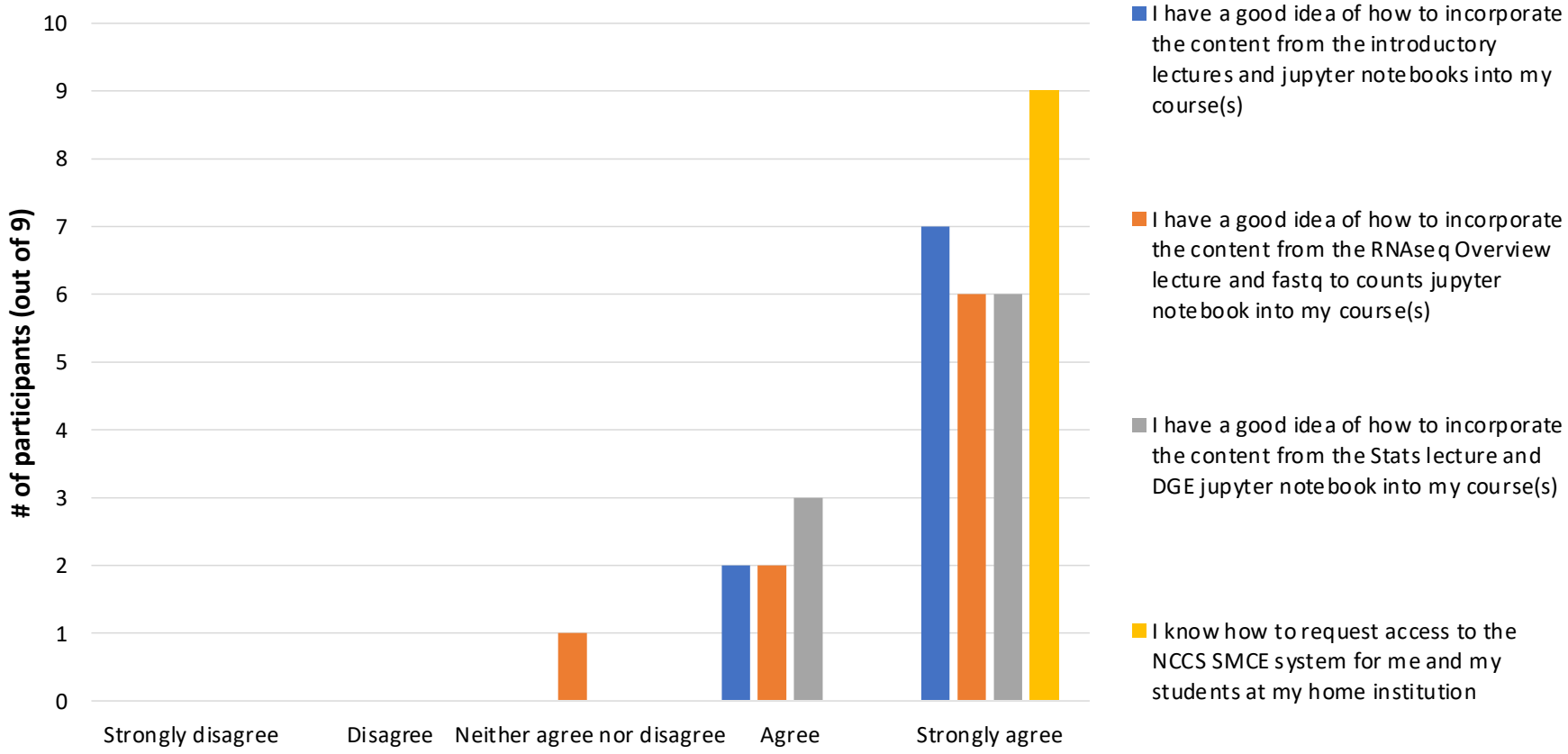
Post-Survey Results: RNAseq Content Feedback



Post-Survey Results: Home Institution Planned Curriculum



Post-Survey Results: Home Institution Teaching Preparedness



Educator Feedback

"**The camp was exceptionally well structured has gone over content a lot better than other courses I have participated in. The ability to run the code and understand it is much more valuable than plug and play GUI environments**, black boxes do not help science, as such things like galaxy or other pipelines really pull down the ability of scientists to approach a dataset and make meaningful ground breaking impacts."

"Thank you! The content and delivery was easy to understand. The instructors are knowledgeable and were considerate of the differences in people's expertise… **I felt very comfortable with the material, I think it was the right amount of information for college-level students.** I appreciate your hard work and time invested in the bootcamp."

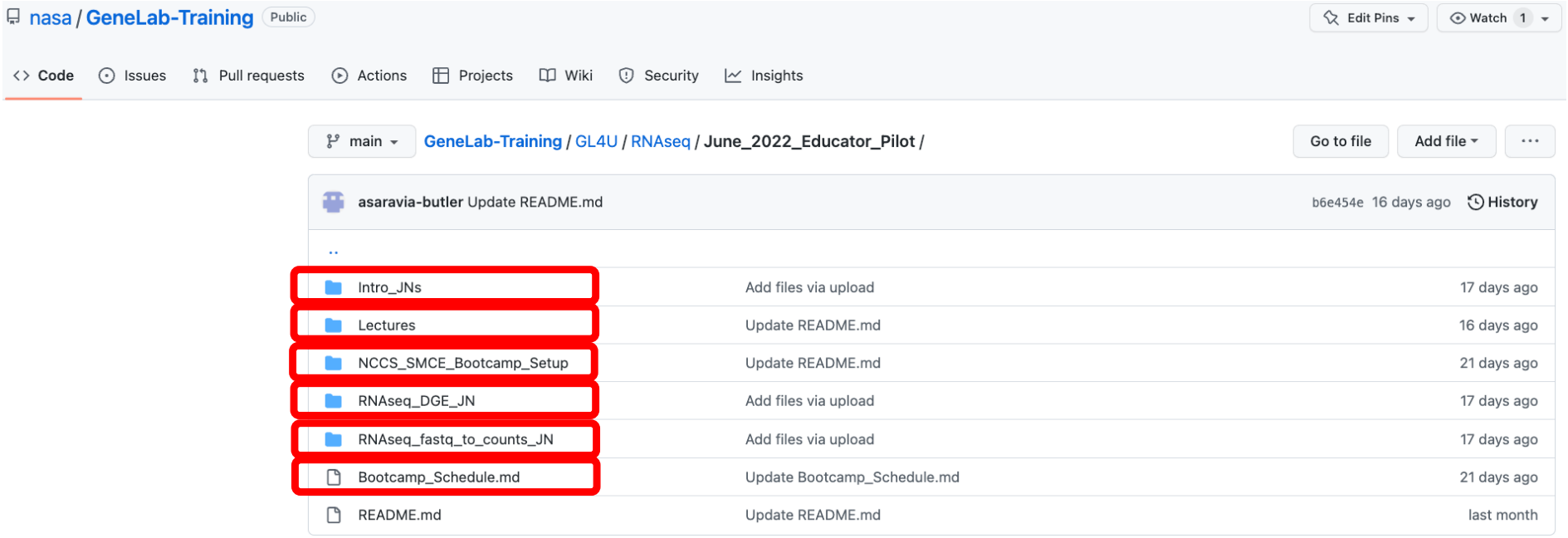
"I thought this was really great. One issue I struggle with is developing material to teach bioinformatics to my students. **The material provided here is fantastic and will greatly improve how I teach bioinformatics and the scope of what I'll be able to teach**..."

"It was super informative and the instructors were extremely patient and passionate."

"**This was an amazing workshop and the absolute best bioinformatics bootcamp that I've ever participated in**. The material was easy to follow and the instructors did a fantastic job facilitating the discussions and providing thorough answers. I also appreciated how willing they were to review content several times if necessary. The areas where I am not confident are not a reflection of weaknesses in the workshop; rather, I need/want to review that content again (and again) to gain the confidence needed to teach it… **I look forward to developing ways to implement the things learned for the benefit of students that I teach.**"

Pilot GL4U: RNAseq Educator Bootcamp Is Available on GitHub

https://github.com/nasa/GeneLab-Training/tree/main/GL4U/RNAseq/June_2022_Educator_Pilot



What's Next?

- Design a bootcamp to teach each of GeneLab's standard processing pipelines
- GL4U: AstroAmpSeq Bootcamp
  - Next GL4U undergraduate-level course designed to teach students how to process amplicon sequence data using a group-based independent learning approach
  - Initiated by NASA intern, Katie Blackwell, under the guidance of Jessica Lee and GeneLab's Mike Lee and Amanda Saravia-Butler
- Update bootcamp content annually to keep consistent with GeneLab Data Processing pipeline versions (https://github.com/nasa/GeneLab_Data_Processing)
- Implement GL4U bootcamps as open-source curriculum accessible worldwide?

Acknowledgements

- **NASA GeneLab**
  - PM: Sylvain Costes, PhD
  - DPM: Samrawit Gebre
  - DP: Lauren Sanders, PhD
  - DP: Amanda Saravia-Butler, PhD
  - DP: Mike Lee, PhD
  - GeneLab Team
- **JPL Planetary Protection**
  - PM: Alvin L. Smith II, PhD
  - Engineer: Lisa Guan
  - Scientist: Arman Seuylemezian
  - JPL PP Team
- **Universities Space Research Association (USRA)**
  - Saba Hussain
  - Tristyn Acasio
  - Rachel Gilbert
- **San Jose State University**
  - Philip Heller, PhD
  - Steven Boring
- **NASA Space Biology**
  - BPS PS: Lisa Carnell
  - PM: Robert Vik
  - PS: Parag Vaishampayan, PhD
  - SB ADC: David Smith

Funding

- Compute resources were made available through the NASA Science Managed Cloud Environment (SMCE) funded by SMD AWS Space Act Agreement
- GeneLab is funded by the NASA Space Biology program within the NASA Science Mission Directorate's (SMD) Biological and Physical Sciences (BPS) Division
- JPL Engineering and Science Directorate (ESD) HBCU/MSI Internal Funding Award

Questions?

https://github.com/nasa/GeneLab-Training/tree/main/GL4U

Sign up for the GL4U mailing list

- Stay up-to-date on future GL4U events / bootcamps: Send and e-mail to GL4U-join@lists.nasa.gov with the Subject: **subscribe**