

National Aeronautics and
Space Administration

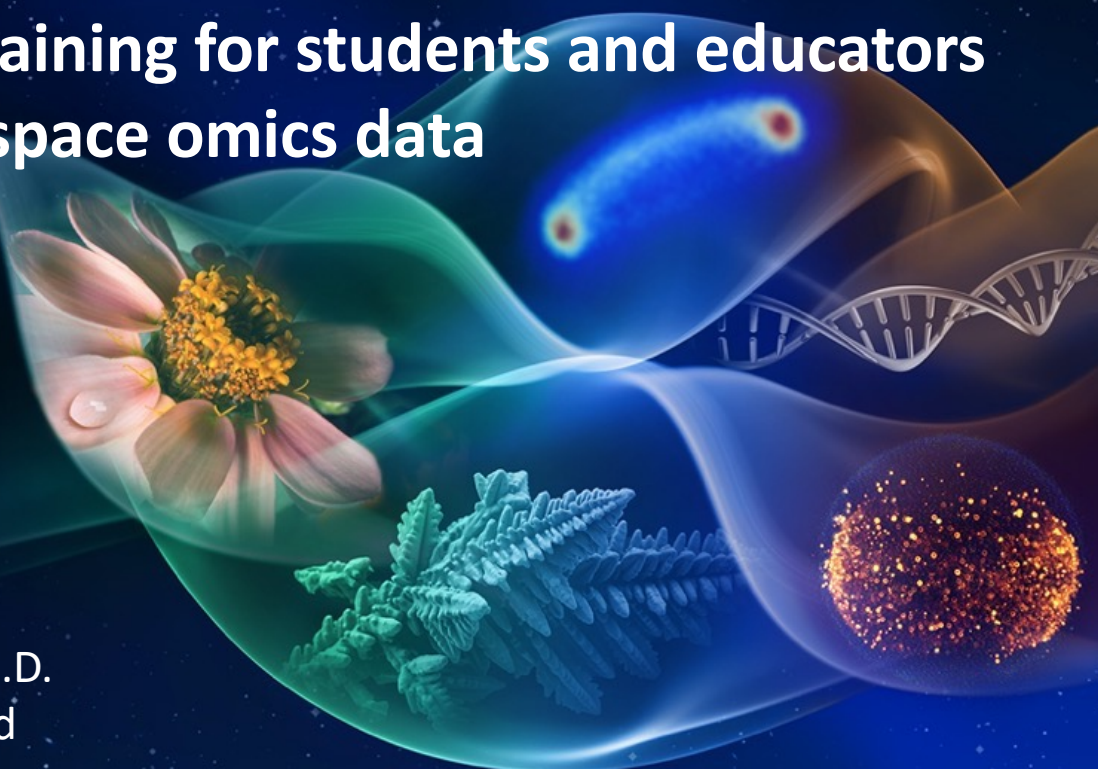


GL4U: Bioinformatics training for students and educators using space omics data

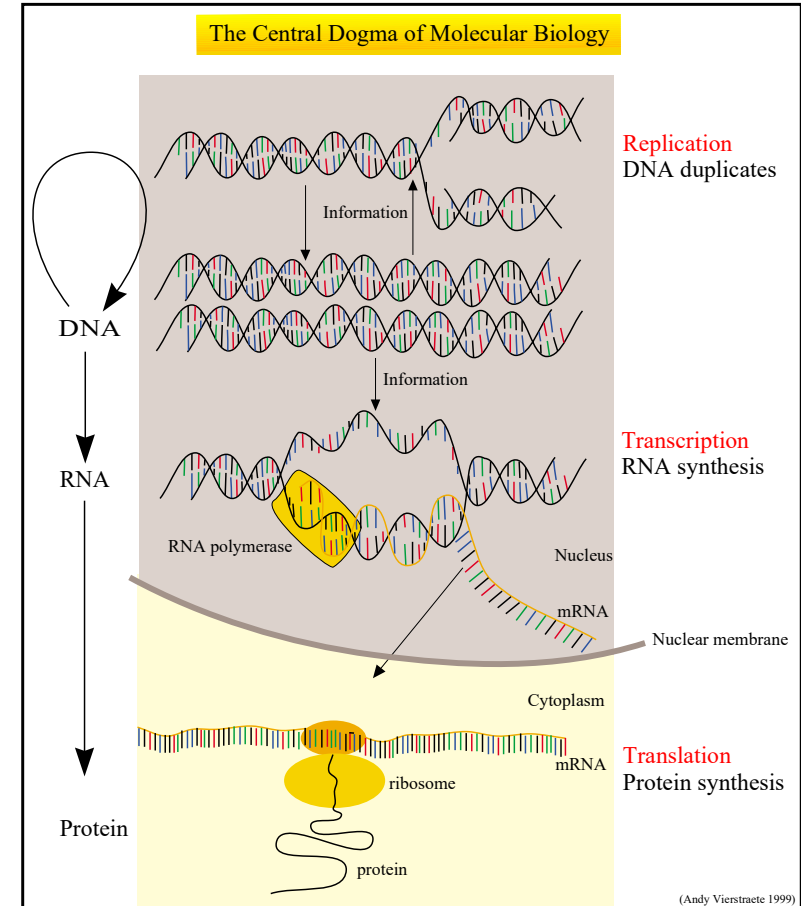
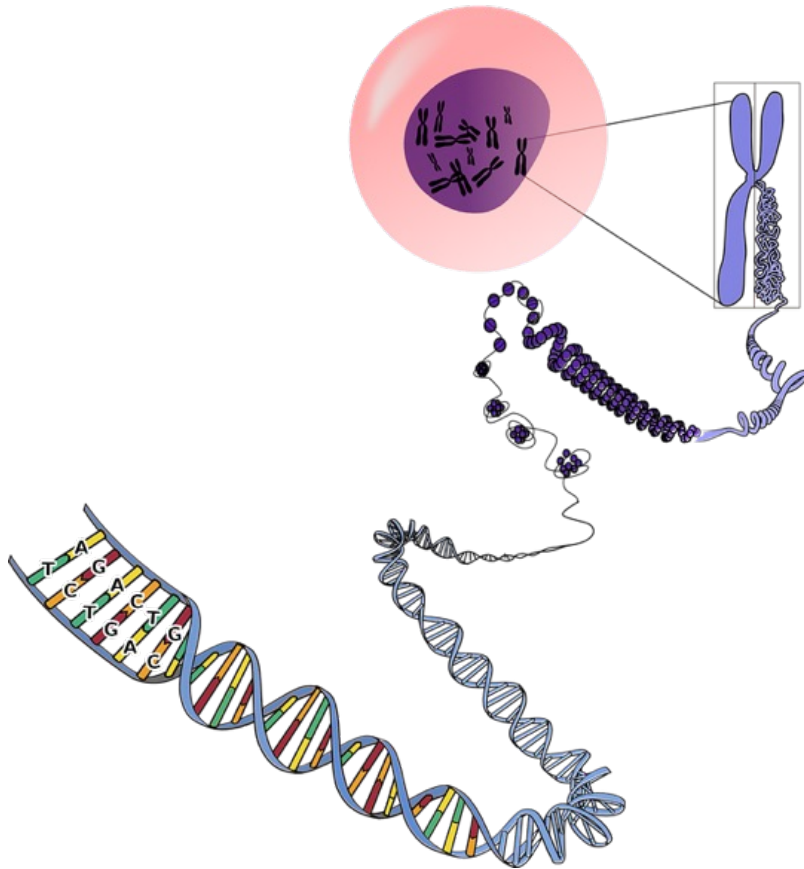
Biological and Physical
Sciences

ISMB 2022
Education COSI

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

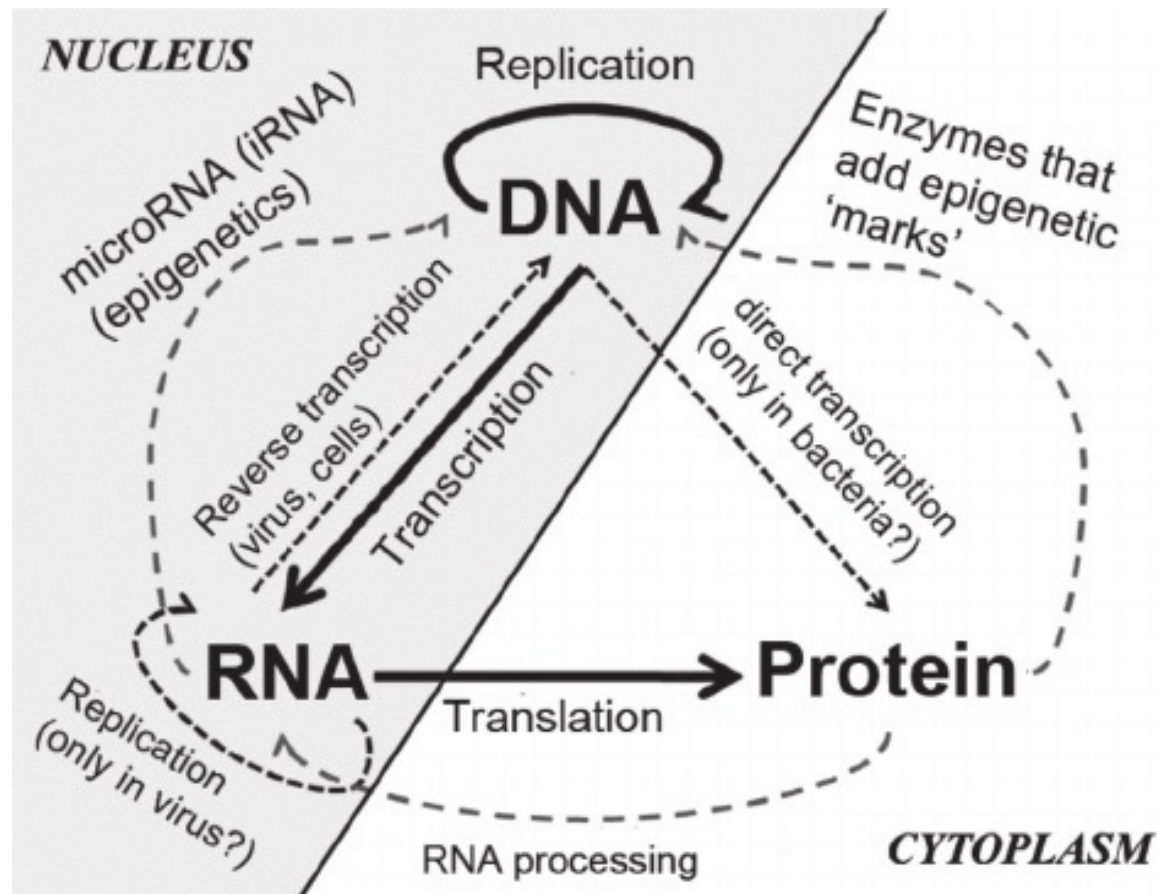


What are 'omics?



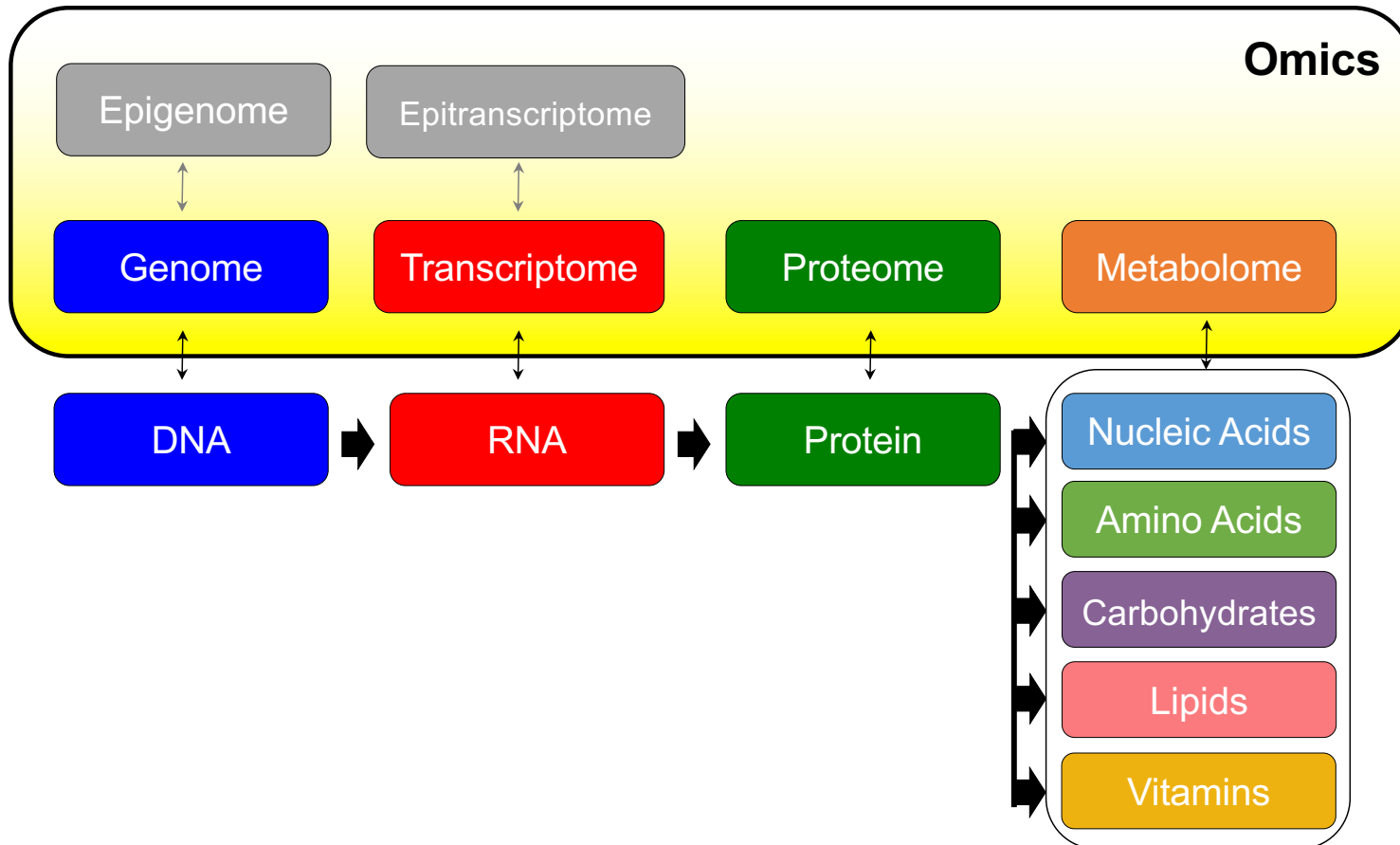
Miller, Christine. 2020. Human Biology.
<https://humanbiology.pressbooks.tru.ca/chapter/4-13-mitosis-and-cytokinesis/>

Beyond the Central Dogma



Gonzalez-Pardo and Alvarez, 2013

What are 'omics?

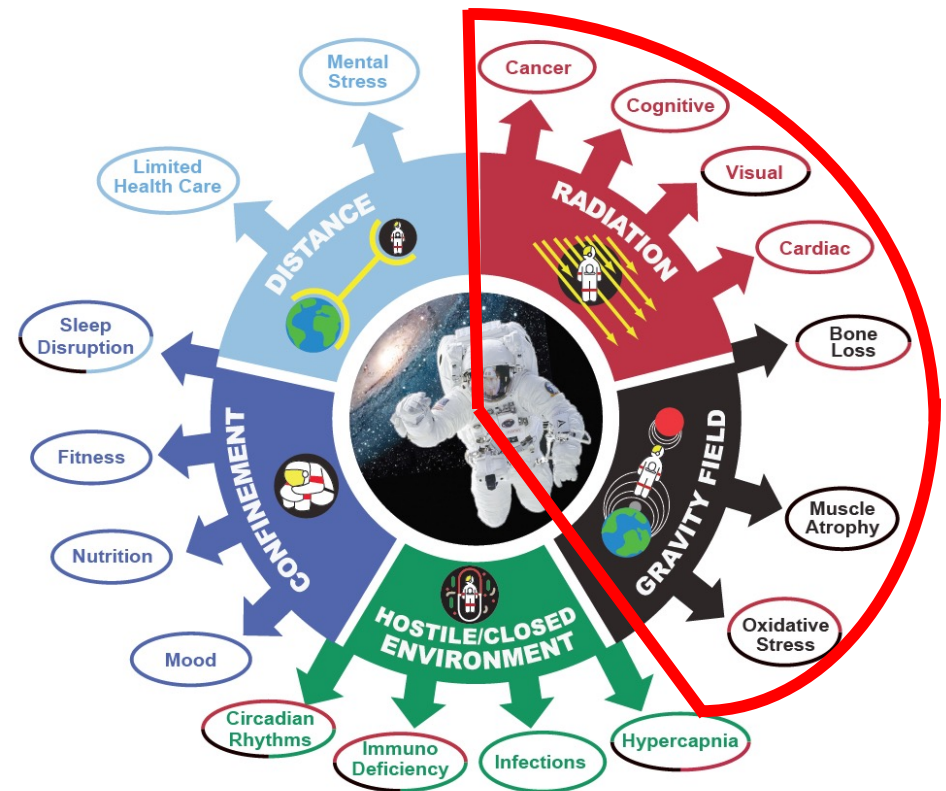


Why is studying omics important for spaceflight?



- 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

Primary Stressors of Spaceflight



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



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

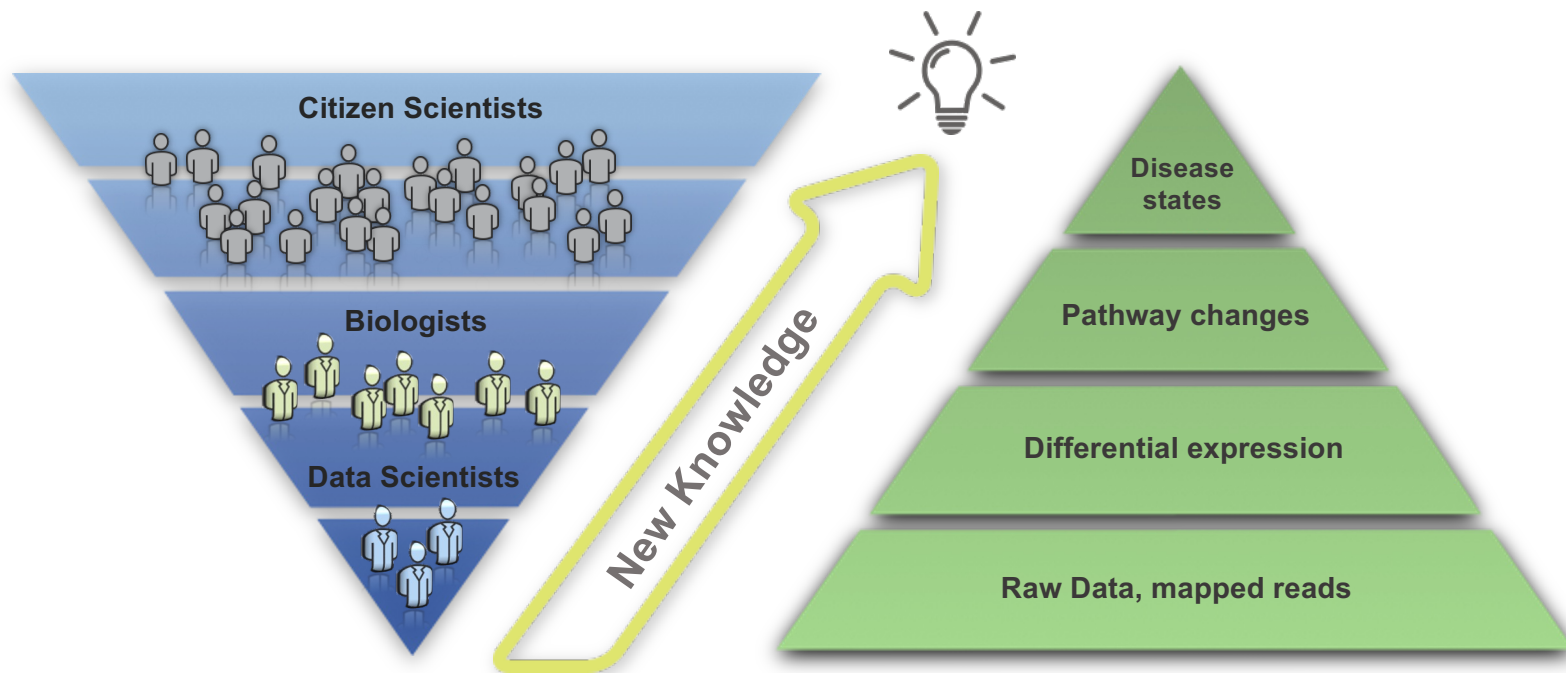
Have space-relevant data to submit to GeneLab?



Visualize Data

Interact with GeneLab processed data

GeneLab Data Democratization



GeneLab Standard Data Processing Pipelines



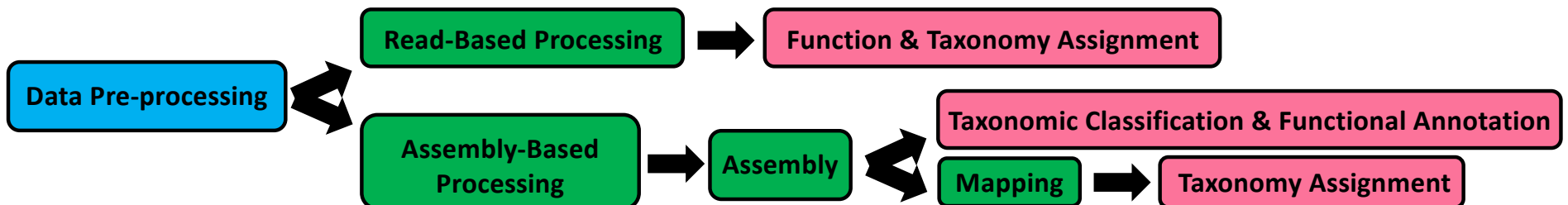
- **RNAseq** – Evaluate gene expression changes by quantifying RNA expression in samples exposed to the space environment vs. control



- **Amplicon Sequencing** – Identify microbial changes in organisms and environments exposed to spaceflight conditions by evaluating sequence changes in specific genomic regions



- **Metagenomics** – Identify microbial changes within organisms and environments exposed to spaceflight conditions by evaluating complete genome sequence changes



GeneLab for Colleges and Universities (GL4U)

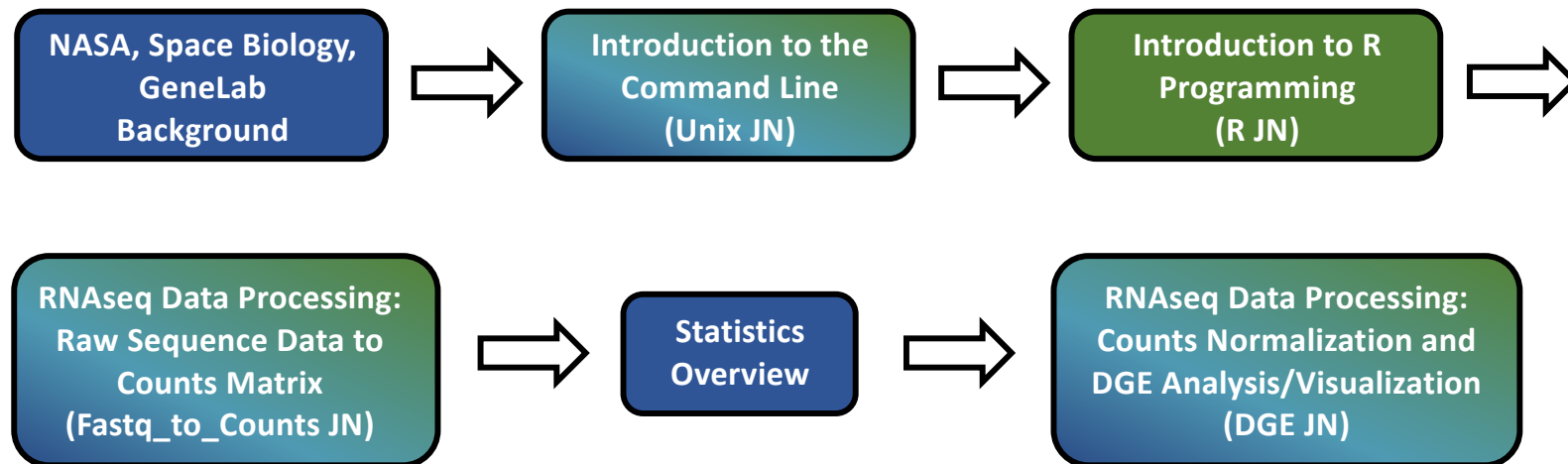


- GL4U provides space biology-relevant training in bioinformatics to the next generation of scientists through direct and indirect approaches
- The GeneLab team plans to host two annual data processing bootcamps in which participants will learn how to analyze GeneLab's space-relevant omics data types
 - 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

Pilot GL4U: RNAseq Student Bootcamp



- The GL4U direct training pilot program was conducted in June 2021 in collaboration with Universities Space Research Association (USRA) and San Jose State University (SJSU)
- During the pilot, 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



Pilot GL4U: RNAseq Student Bootcamp Participants



Pilot Student Bootcamp Pre-Survey Results



■ Strongly Disagree ■ Disagree ■ Somewhat agree ■ Agree ■ Strongly agree

1. I am proficient with UNIX/Linux commands.



2. I am comfortable coding in Python.



3. I am comfortable coding in R.



4. I am comfortable running commands using Jupyter Notebooks.



5. I am able to design a RNAseq experiment.



6. I know how samples are prepared for RNA sequencing.



7. I understand sequencing by synthesis.



Pilot Student Bootcamp Pre-Survey Results Cont'd



■ Strongly Disagree ■ Disagree ■ Somewhat agree ■ Agree ■ Strongly agree

8. I know how to analyze RNAseq data.



9. I can create RNAseq data visualizations in R.



10. I have statistics experience.



11. I know how to interpret p-values, adjusted p-values, and log fold change in gene expression data.



12. I know about and understand space biology.



13. I am considering a career in bioinformatics.



14. I am considering a career in space biology.



100%

0%

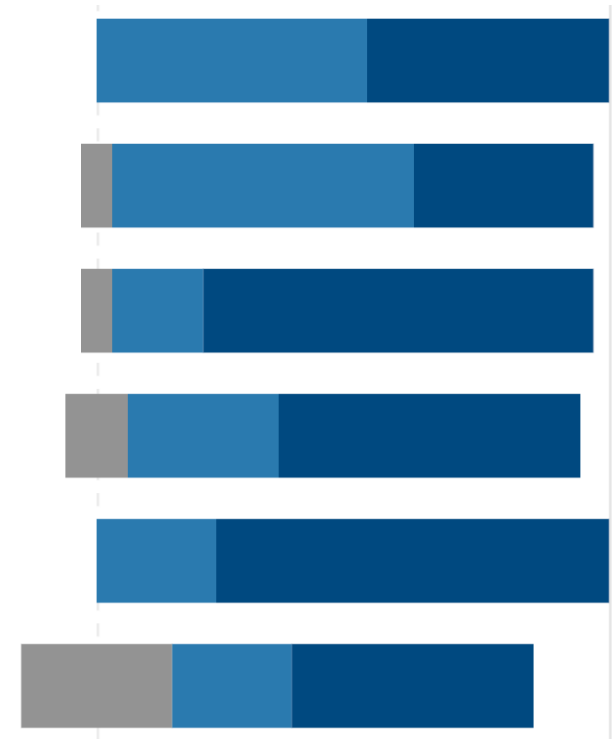
100%

Pilot Student Bootcamp Post-Survey Results: Intro/Coding



■ Strongly Disagree ■ Disagree ■ Somewhat agree ■ Agree ■ Strongly agree

1. I understand how to use basic UNIX/Linux commands in Jupyter Notebooks.
2. I understand how to use basic R commands in Jupyter Notebooks.
3. The NASA / GeneLab / Command Line introductory lecture was engaging and useful.
4. The speed, delivery, and clarity of the of the NASA / GeneLab / Command Line Intro lecture were...
5. The introductory Unix and R Jupyter Notebooks were well structured, easy to follow, and informative.
6. I know how to prepare an output directory structure for use in RNA sequencing analysis.

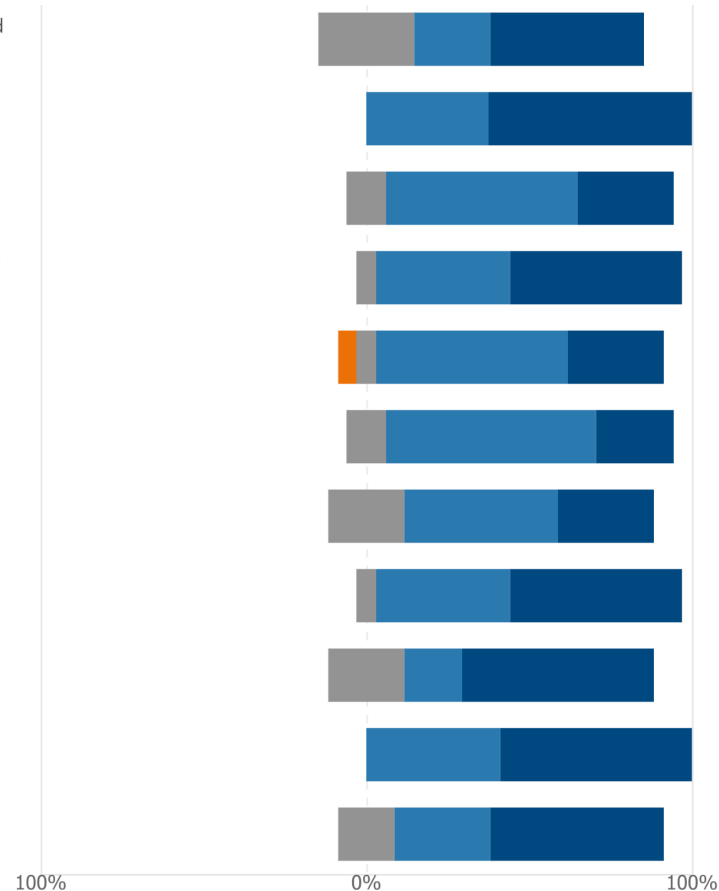


Bootcamp Post-Survey Results: RNAseq (Fastq to Counts)



■ Strongly Disagree
 ■ Disagree
 ■ Somewhat agree
 ■ Agree
 ■ Strongly agree

- 7. I know how to generate and interpret a fastQC (and respective multiQC) report.
- 8. I understand why raw sequencing data needs to be cleaned up before processing.
- 9. I know how to trim/filter raw sequencing data.
- 10. I understand what a reference genome is and why it is used in RNA sequencing analysis.
- 11. I have a good understanding of why spliced aligners were developed and how they align reads t...
- 12. I can assess alignment quality by interpreting information found in alignment log files and...
- 13. I know how to generate gene count tables from RNA sequencing data.
- 14. The RNAseq Overview lecture was engaging, easy to follow, relevant, and informative.
- 15. The speed, delivery, and clarity of the of the RNAseq Overview lecture were appropriate.
- 16. The Fastq-to-Counts Jupyter Notebook was well structured, easy to follow, engaging, and informative.
- 17. The speed, delivery, and clarity of the Fastq to Counts Jupyter Notebook were appropriate.



Bootcamp Post-Survey Results: RNAseq (Stats/DGE)



■ Strongly Disagree ■ Disagree ■ Somewhat agree ■ Agree ■ Strongly agree

18. I have a good understanding of the statistics used in GeneLab's standard RNA sequencing analysis.



19. I understand and can interpret a PCA plot.



20. I understand why raw gene expression data needs to be normalized.



21. I know how to normalize raw counts data.



22. I understand how normalized counts are used for calculating differential gene expression.



23. I understand the concept of hypothesis testing and what a null hypothesis is.



24. I know the difference between p-value and adjusted p-value.



25. I can interpret a p-value, adjusted p-value, and fold change in the context of statistical significance.



26. I know how to obtain gene annotations and gene ontology information for annotated genes.



Bootcamp Post-Survey Results: RNAseq (Stats/DGE/Visualizations)



■ Strongly Disagree ■ Disagree ■ Somewhat agree ■ Agree ■ Strongly agree

27. I understand and can interpret a heatmap.

28. I understand and can interpret a Volcano plot.

29. The Statistics overview lecture was engaging, easy to follow, relevant, and informative.

30. The speed, delivery, and clarity of the Statistics overview lecture were appropriate.

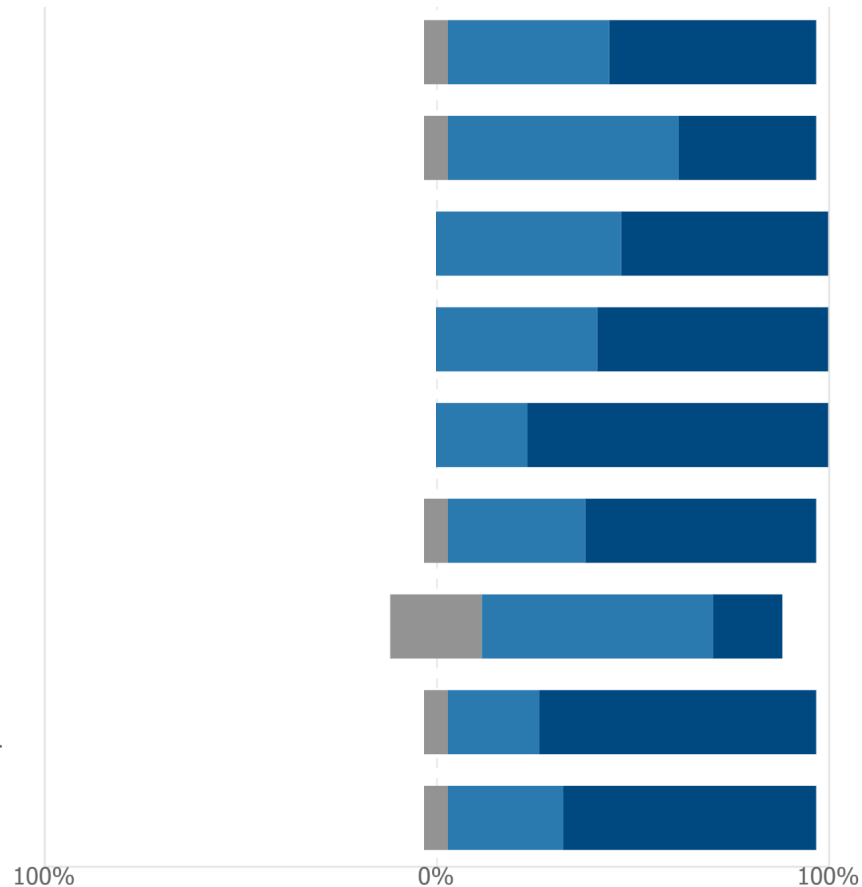
31. The DGE Jupyter Notebook was well structured, easy to follow, engaging, and informative.

32. The speed, delivery, and clarity of the DGE Jupyter Notebook were appropriate.

33. If given a set of raw RNA sequencing data, I can use the knowledge gained in this bootcamp to...

34. I would be interested in exploring NASA GeneLab data further using the knowledge that I gained durin...

35. This bootcamp and its content are relevant to my interests and/or to my degree coursework.



SJSU Student Feedback



“The bootcamp was very informative and lectures were very well structured. I just wished that it was distributed into more than 5 days as it was a lot of information to digest. But I am very glad that the lectures and materials were shared so I can refer to parts that I was not clear of.... Overall, the bootcamp experience was amazing, loved the fact that we had guest speakers as well. Thank you so much, especially to Amanda and Lauren for being so helpful and patient in explaining difficult concepts.”

“This was such a great experience, I learned lots of new information and I can't wait to explore more into Unix and R and their functionalities!”

“I found the Bootcamp extremely well organized and its content very thorough.”

“The bootcamp was great. I definitely appreciated the depth of the JNs because they helped reinforce the learning from the lectures really well. The RNAseq overview lecture was a lot for me to take in.... I will definitely save **the RNAseq lecture material** to study in the future because it is **bursting with great information.**”

“Considering the amount of content they had to teach us in a week, I thought the speakers did an amazing job.... I think I will have to go through the lectures a few more time to really absorb everything now, but I still learned so much throughout the week and the speakers were very engaging. I appreciated how they asked questions for us to answer and answered our questions very thoroughly. **This is an experience I would recommend to anyone** who is willing to put the time and effort into learning about RNA sequencing.”

Pilot GL4U: RNAseq Student Bootcamp Is Available on GitHub



https://github.com/nasa/GeneLab-Training/tree/main/GL4U/RNAseq/June_2021_Student_Pilot

nasa / GeneLab-Training Public

Edit Pins Watch 1

<> Code Issues Pull requests Actions Projects Wiki Security Insights

main GeneLab-Training / GL4U / RNAseq / June_2021_Student_Pilot / Go to file Add file ...

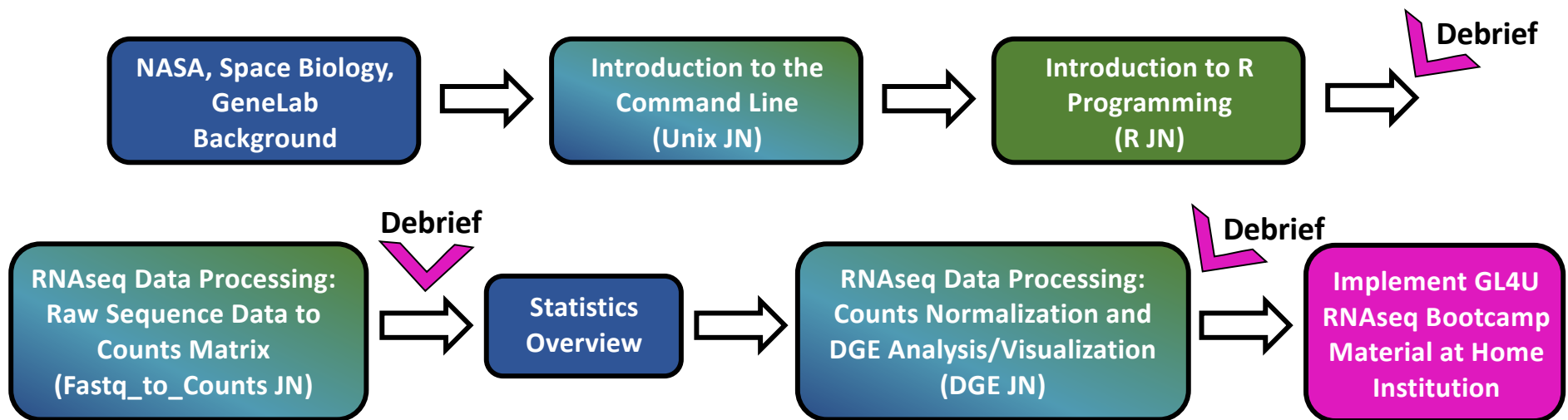
asaravia-butler Changing the name of the Pilot GL4U student bootcamp directory to ind... b92b264 on Apr 21 History

..		
Intro_JNs	Changing the name of the Pilot GL4U student bootcamp directory to ind...	2 months ago
JupyterHub_Installation	Changing the name of the Pilot GL4U student bootcamp directory to ind...	2 months ago
Lectures	Changing the name of the Pilot GL4U student bootcamp directory to ind...	2 months ago
RNAseq_DGE_JN	Changing the name of the Pilot GL4U student bootcamp directory to ind...	2 months ago
RNAseq_fastq_to_counts_JN	Changing the name of the Pilot GL4U student bootcamp directory to ind...	2 months ago
Set-Up_RNAseq_Processed_Data	Changing the name of the Pilot GL4U student bootcamp directory to ind...	2 months ago
Bootcamp_Schedule.md	Changing the name of the Pilot GL4U student bootcamp directory to ind...	2 months ago
README.md	Changing the name of the Pilot GL4U student bootcamp directory to ind...	2 months ago

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 Mexico State University



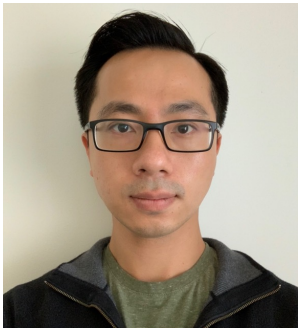
Dr. Keneshia Johnson
Assistant Professor of
Chemistry
Alabama A&M University



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

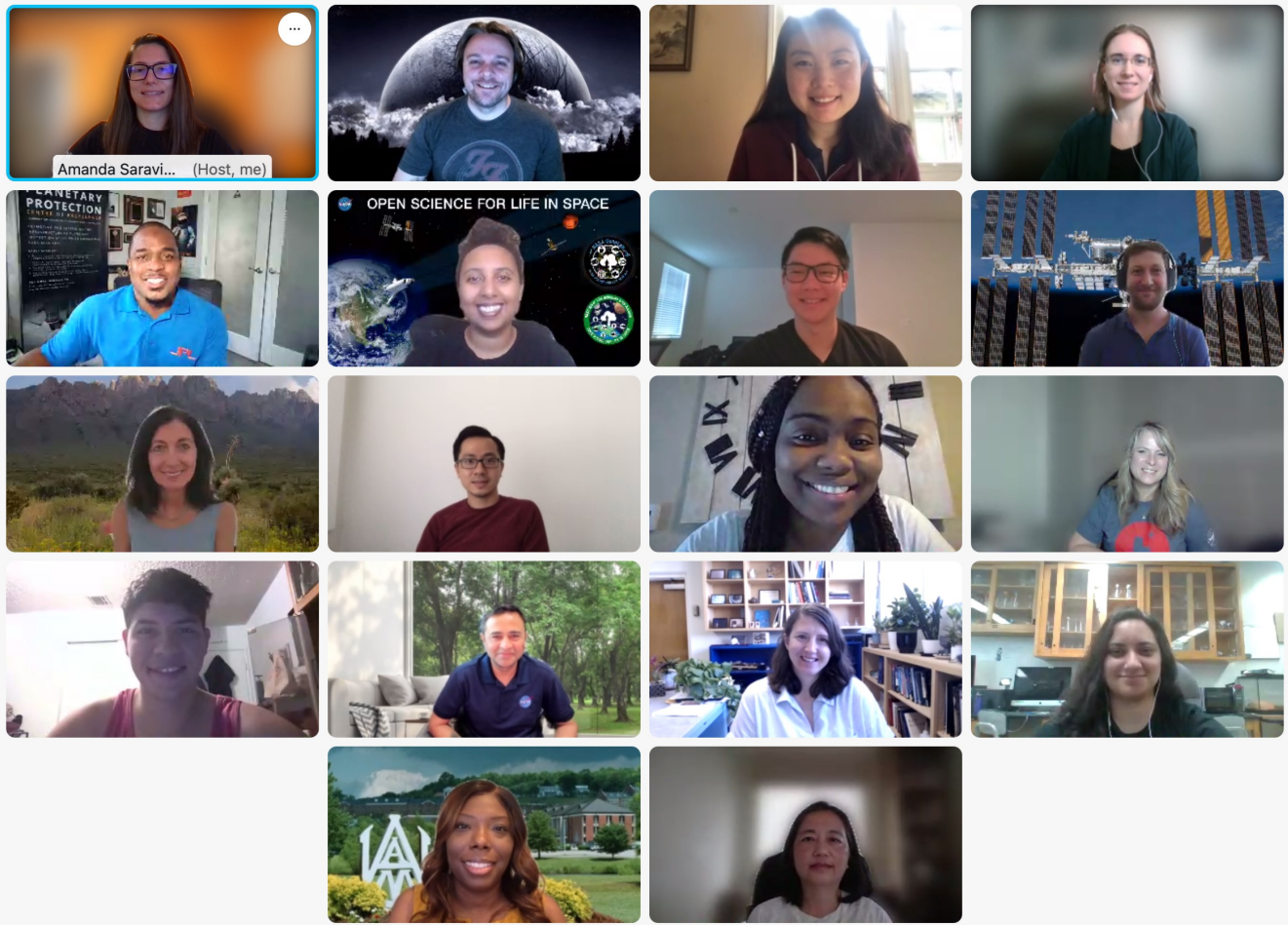


Suzi Arzoumany
Teaching Associate,
Biology
California State University,
Northridge

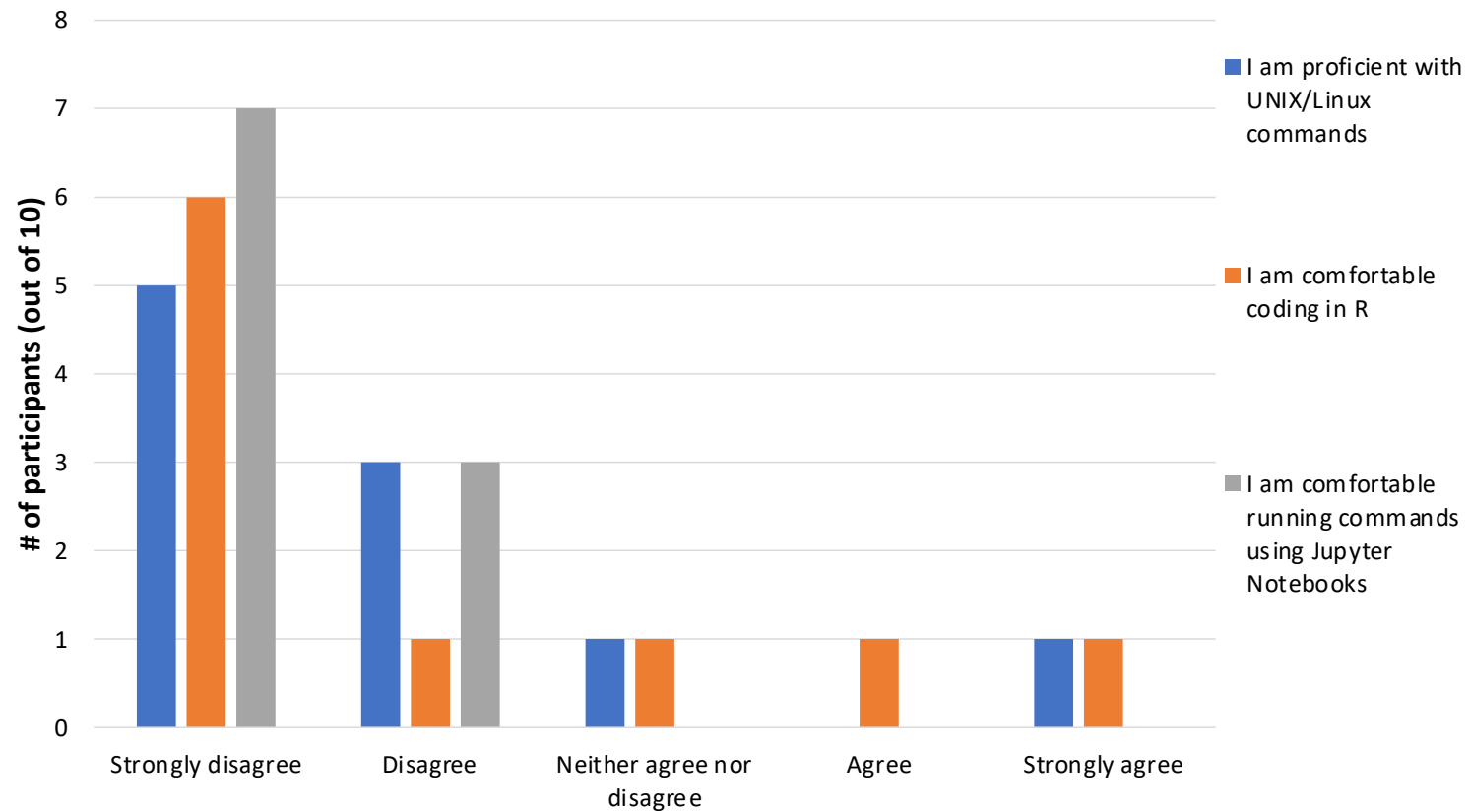


Mark Ortiz
Master's Student, Biology
California State Polytechnic
University

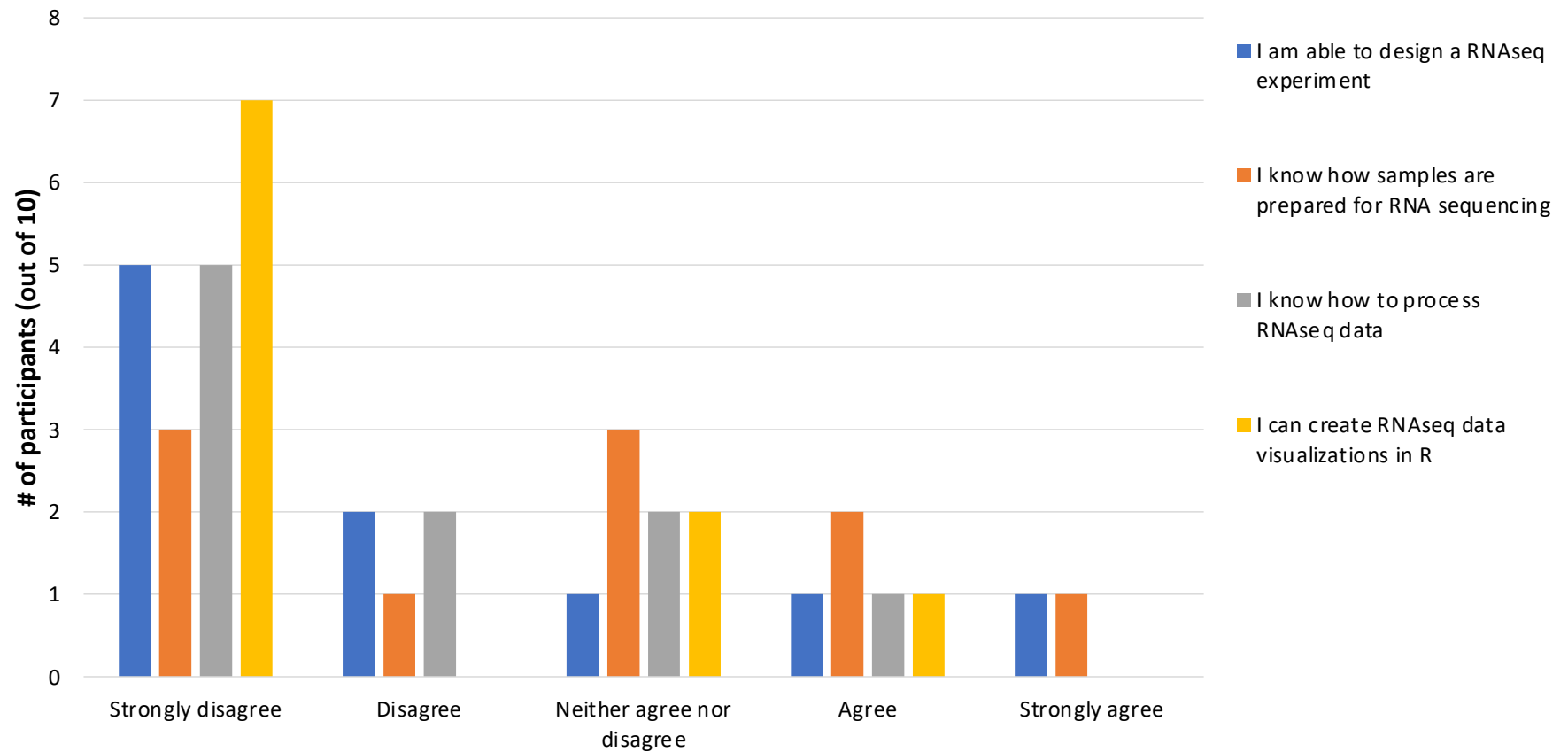
Pilot GL4U: RNAseq Educator Bootcamp Participants



Pre-Survey Results: Basic Coding



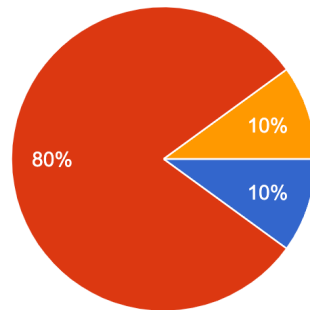
Pre-Survey Results: RNAseq Knowledge



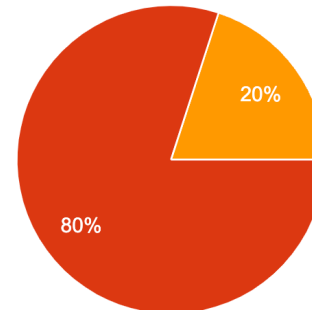
Pre-Survey Results: Home Institution Curriculum



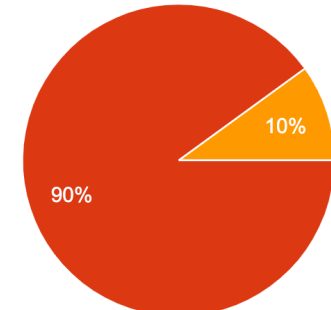
I currently teach UNIX/Linux commands in one (or more) of my courses



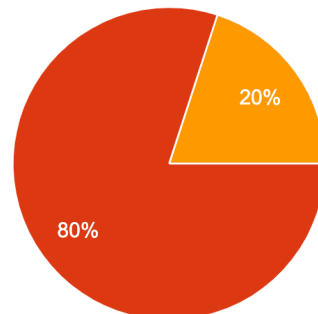
I currently teach R in one (or more) of my courses



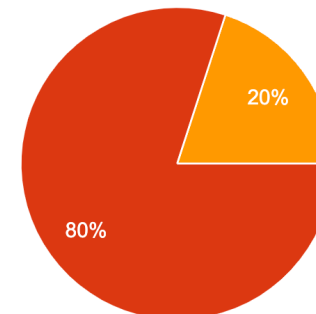
I currently use Jupyter in one (or more) of my courses



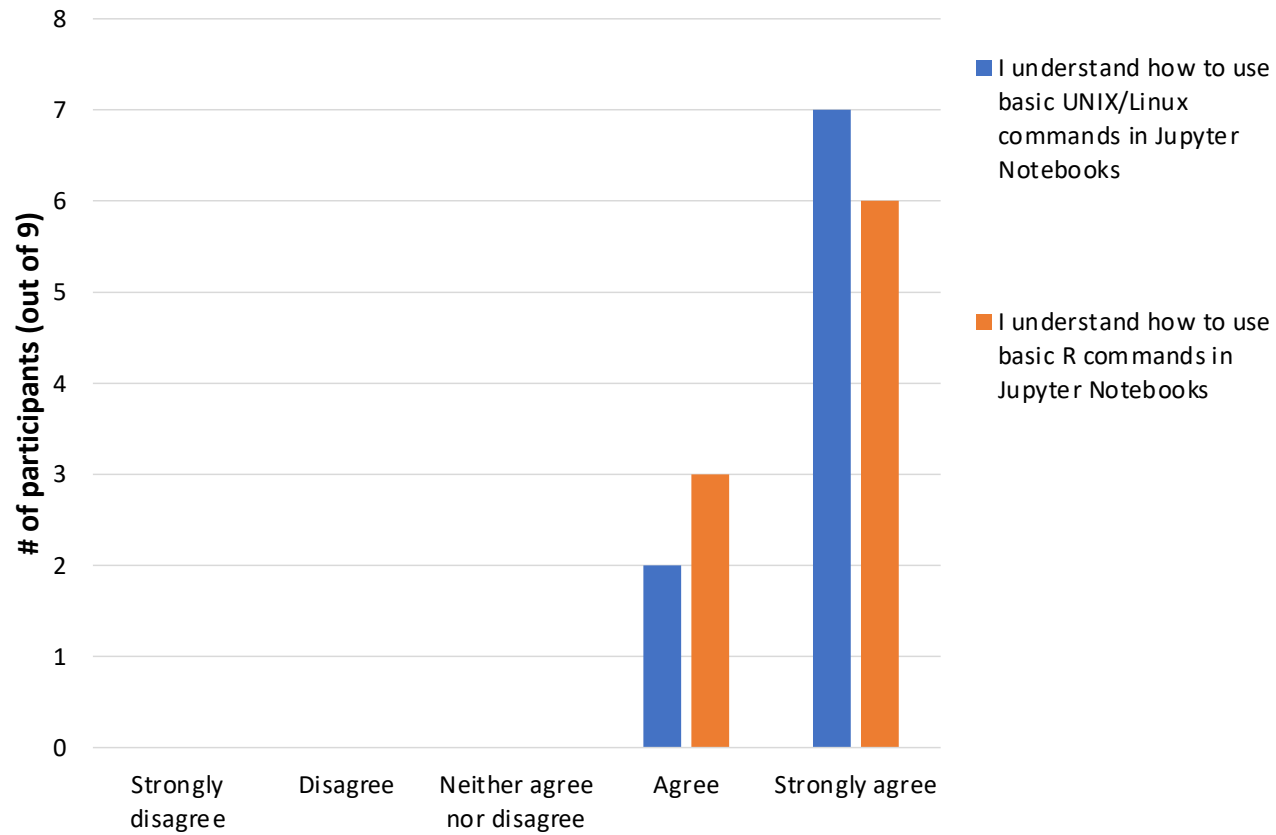
I currently teach a bioinformatics course



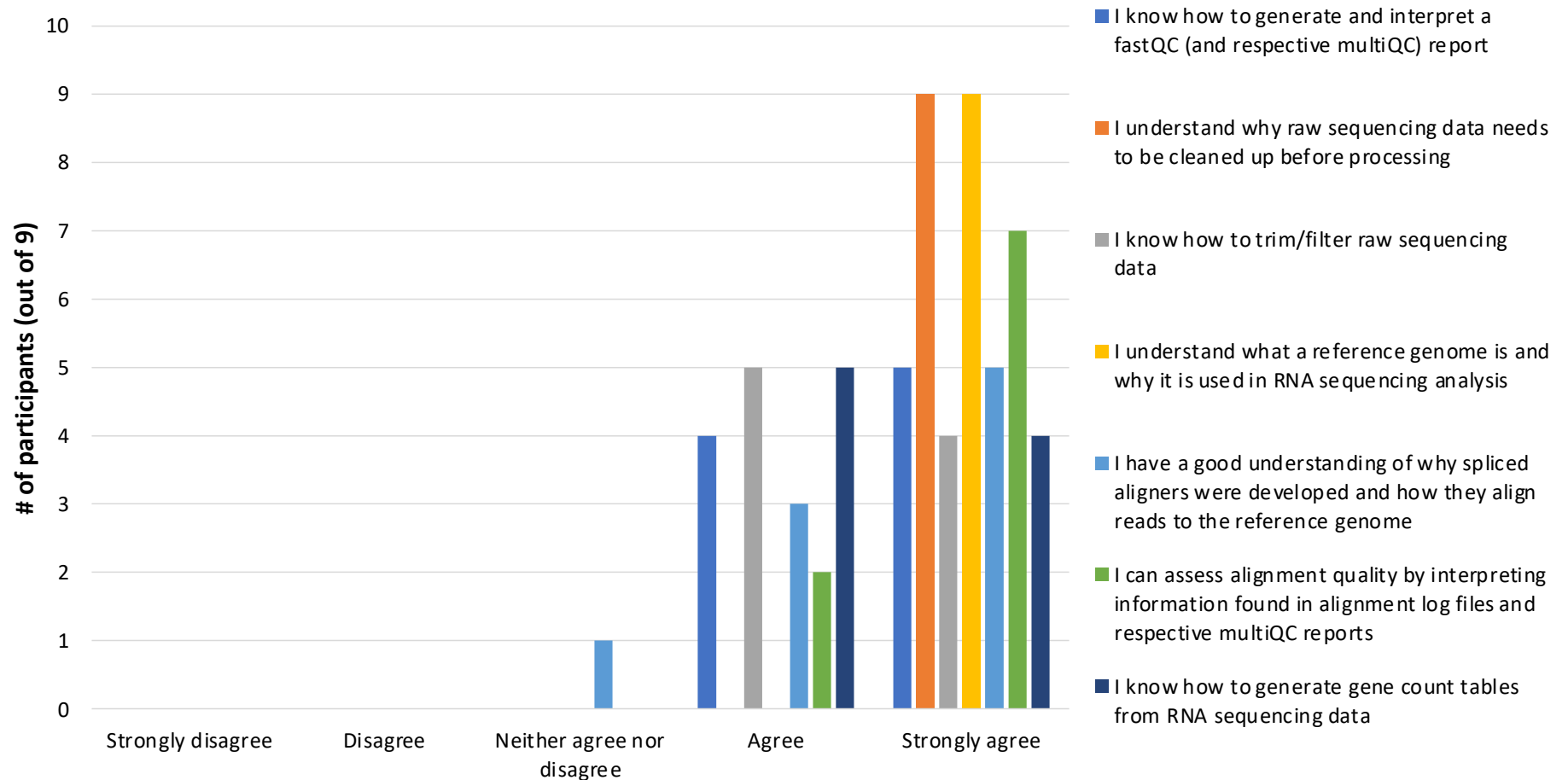
I currently teach RNA sequence analysis in one (or more) of my courses



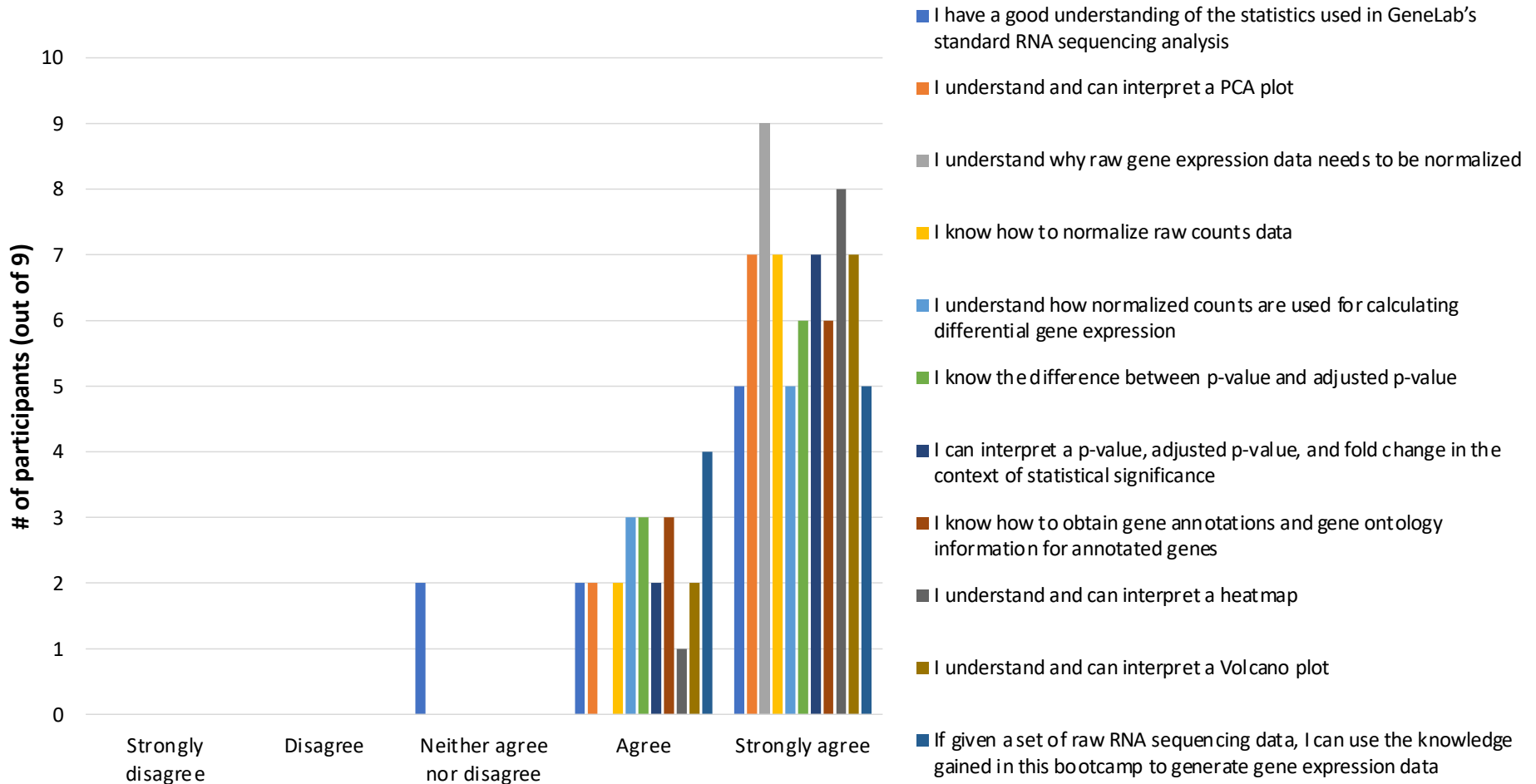
Post-Survey Results: Basic Coding



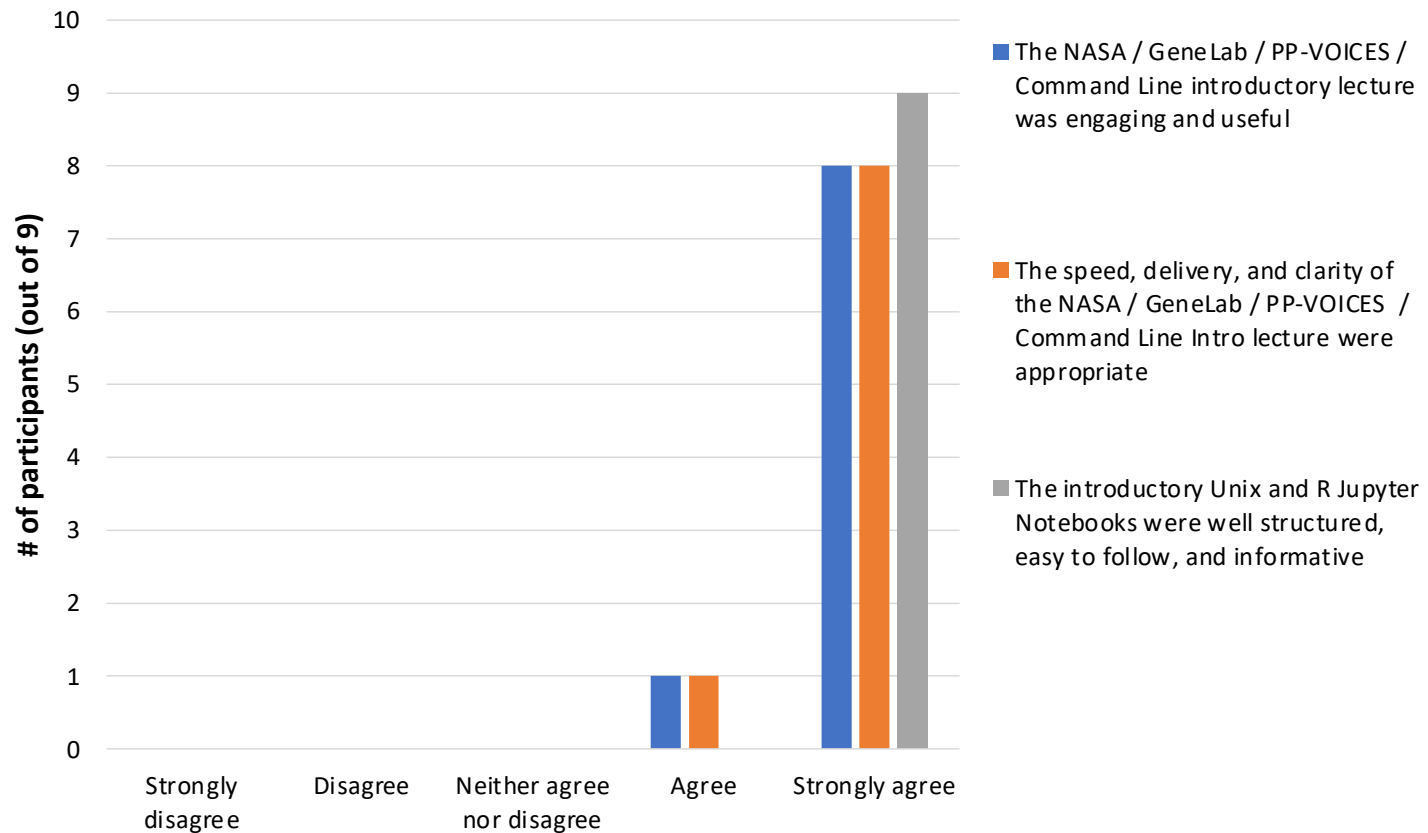
Post-Survey Results: RNAseq (Fastq to Counts)



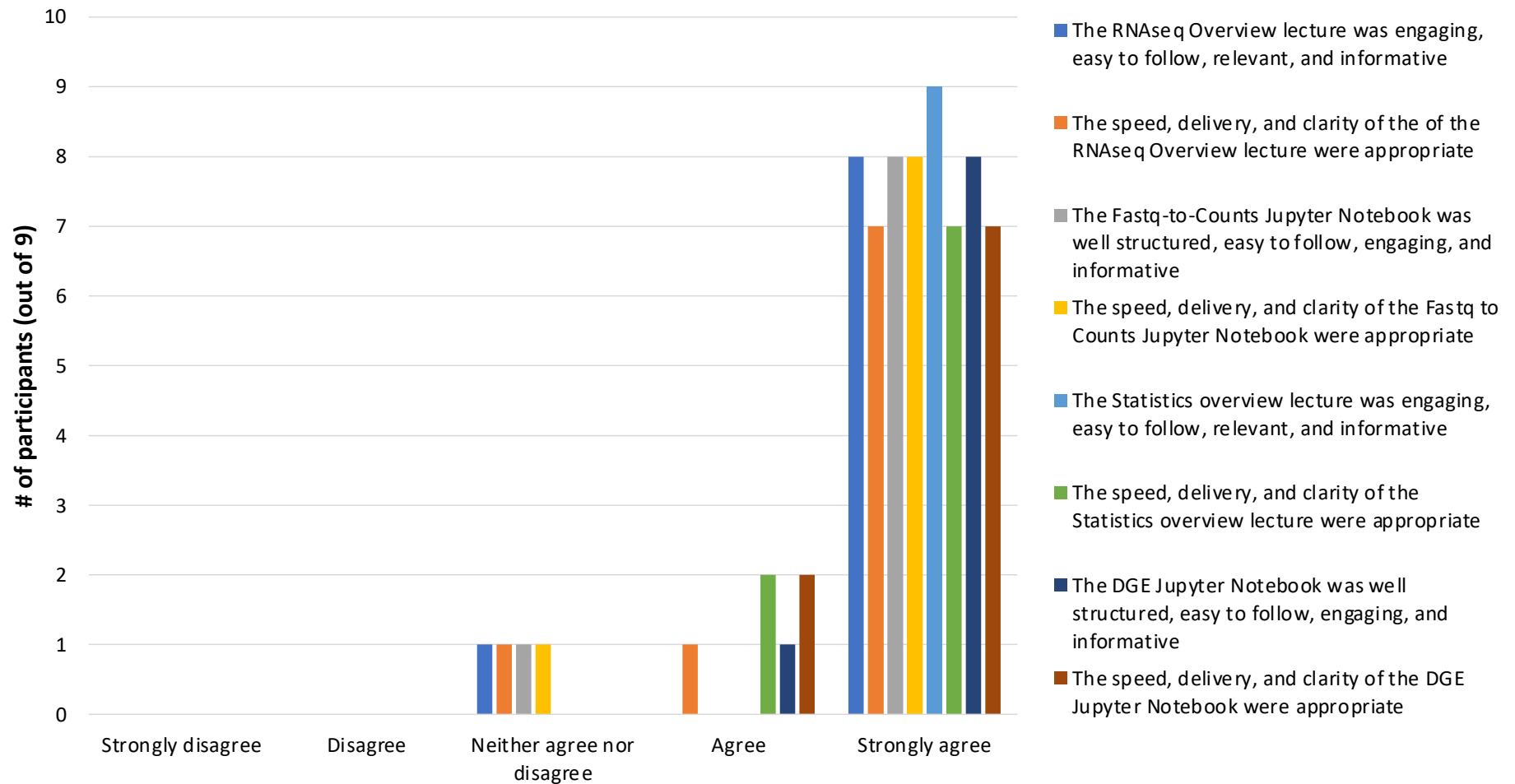
Post-Survey Results: RNAseq (DGE and Visualizations)



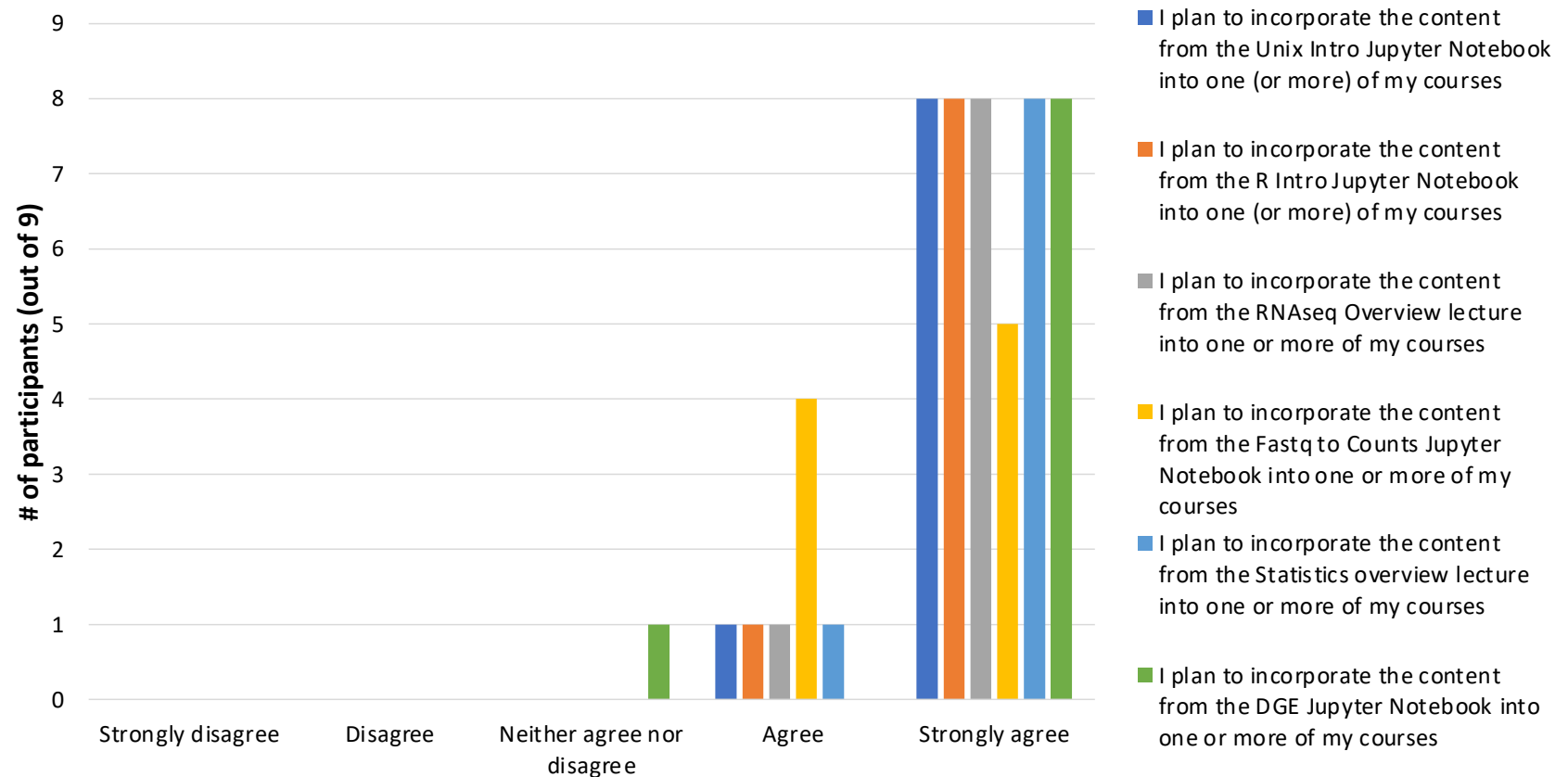
Post-Survey Results: Intro/Coding Content Feedback



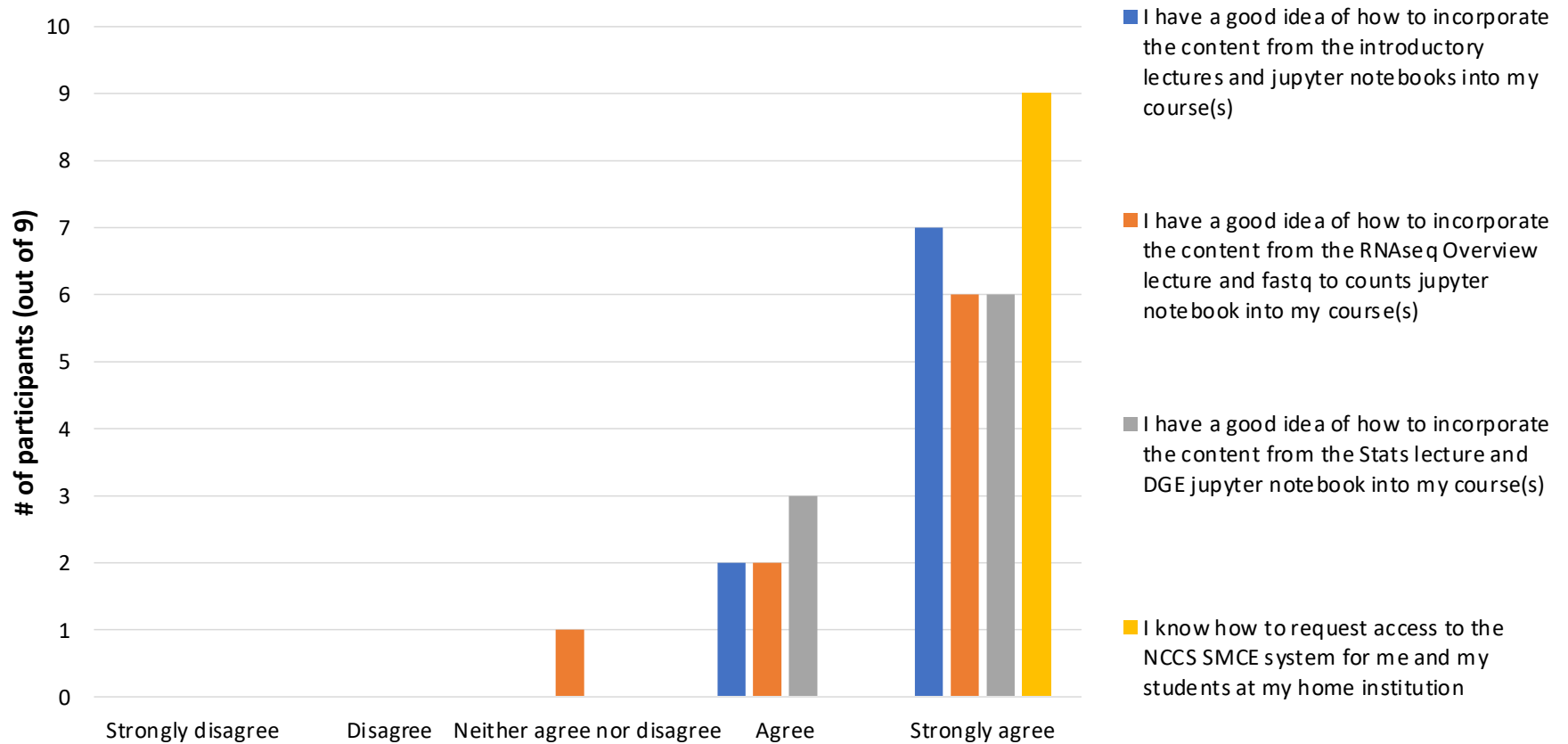
Post-Survey Results: RNAseq Content Feedback



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

nasa / GeneLab-Training Public

Edit Pins Watch 1

<> Code Issues Pull requests Actions Projects Wiki Security Insights

main GeneLab-Training / GL4U / RNAseq / June_2022_Educator_Pilot / Go to file Add file ...

asaravia-butler Update README.md b6e454e 16 days ago History

Intro_JNs	Add files via upload	17 days ago
Lectures	Update README.md	16 days ago
NCCS_SMCE_Bootcamp_Setup	Update README.md	21 days ago
RNAseq_DGE_JN	Add files via upload	17 days ago
RNAseq_fastq_to_counts_JN	Add files via upload	17 days ago
Bootcamp_Schedule.md	Update Bootcamp_Schedule.md	21 days ago
README.md	Update README.md	last month

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
 - Created 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 an e-mail to GL4U-join@lists.nasa.gov
with the Subject: **subscribe**