

The Genomics Education Alliance

Scalable, Sustainable Infrastructure for Undergraduate CUREs

Jason Williams

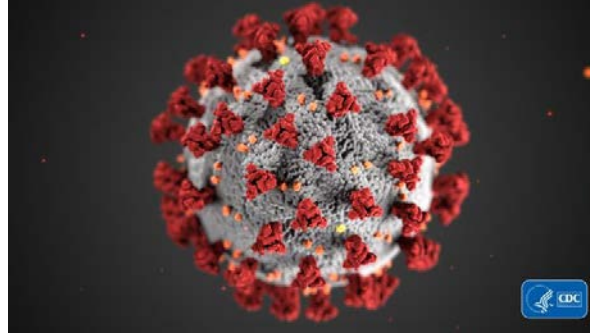
Cold Spring Harbor Laboratory, DNA Learning Center

williams@cshl.edu

 [@JasonWilliamsNY](https://twitter.com/JasonWilliamsNY)



Genomics is essential



Q.E.D.

Hands-on teaching approaches work

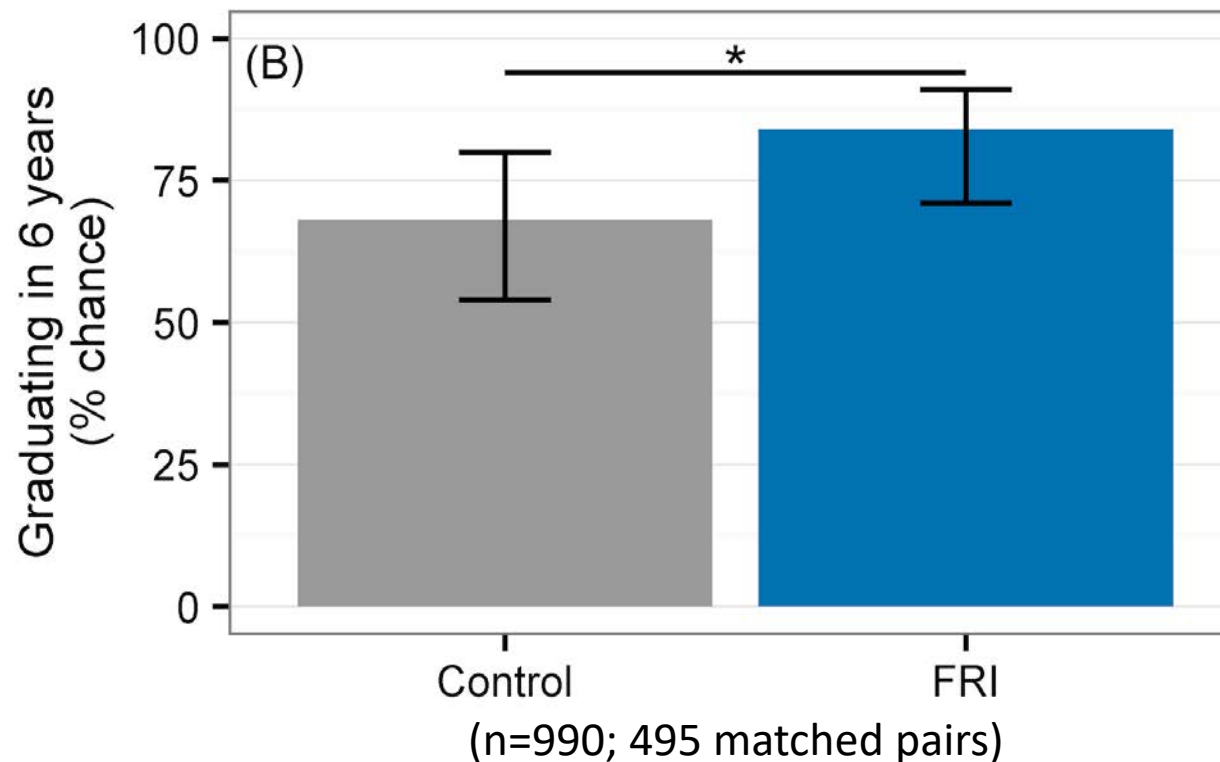
Course-based research is the gold standard

Results from UT Austin Freshman Research Initiative

Course-based research is the gold standard

Results from UT Austin Freshman Research Initiative

6-year graduation rate



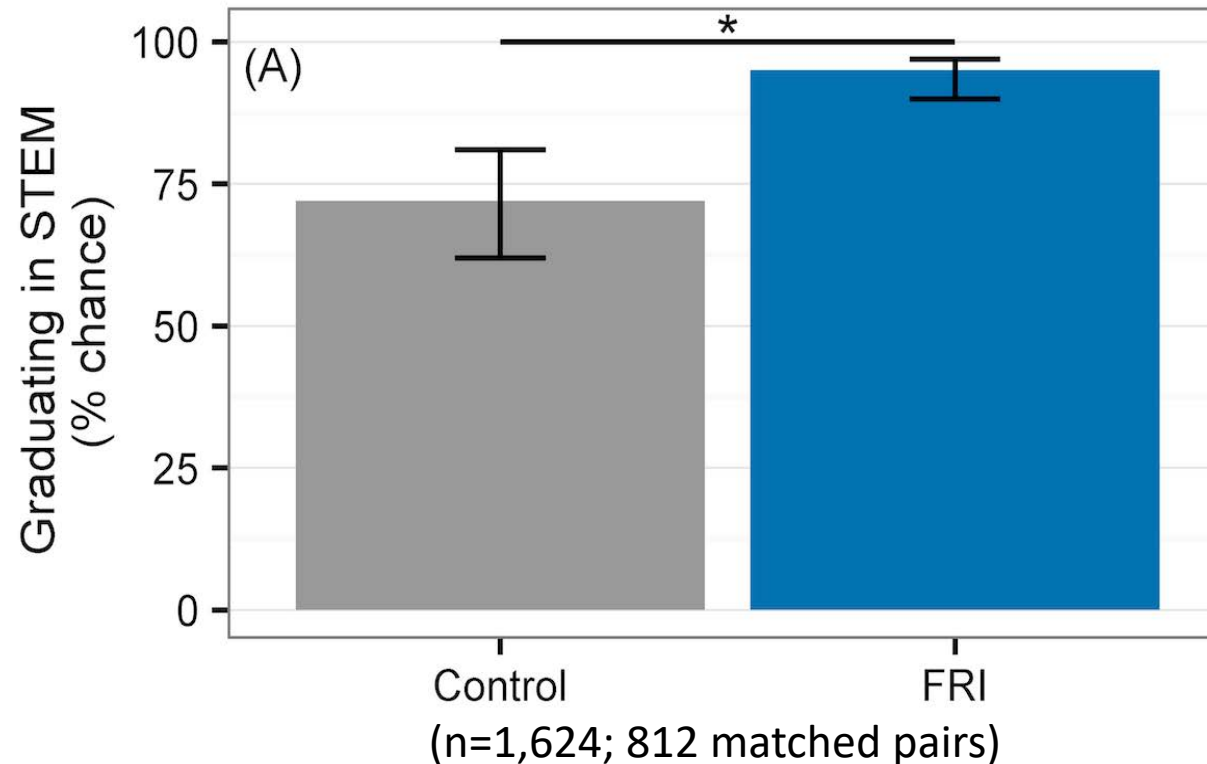
Yes!
17% more likely
83% vs. 66%

Rodenbusch, Hernandez, Simmons, & Dolan,
CBE-Life Sciences Education, June 2016

Course-based research is the gold standard

Results from UT Austin Freshman Research Initiative

Graduation with STEM degree



Yes!
23% more likely
94% vs. 71%

Rodenbusch, Hernandez, Simmons, & Dolan,
CBE-Life Sciences Education, June 2016

Bioinformatics and data science
skills are crucial

Skills gap

95%

of faculty indicate that bioinformatics should be integrated into the life science curriculum;

Skills gap

95%

of faculty indicate that bioinformatics should be integrated into the life science curriculum;

32% of faculty report achieving this



RESEARCH ARTICLE

Barriers to integration of bioinformatics into undergraduate life sciences education: A national study of US life sciences faculty uncover significant barriers to integrating bioinformatics into undergraduate instruction



How can we enable the best
possible genomics education?

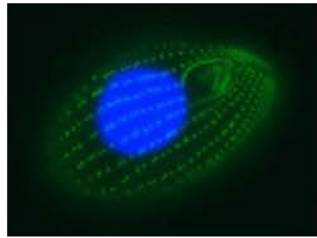
Circle the wagons



An NSF Research Collaboration Network

Founding Members

Ciliate Genomics Consortium



Genome Solver



GCAT-Seek



Network for Integrating Bioinformatics in
Life Sciences Education



Genomics Education Partnership



CyVerse



Address common needs

GEA Challenges

- **How can we enable faculty to teach authentic computational skills and support student research?**

GEA Challenges

- **How can we enable faculty to teach authentic computational skills and support student research?**
- **How can faculty get the training and resources needed to teach genomics?**

GEA Challenges

- **How can we enable faculty to teach authentic computational skills and support student research?**
- **How can faculty get the training and resources needed to teach genomics?**
- **What materials are needed to support learning goals and how can they be made organized and findable?**

GEA Challenges

- **How can we enable faculty to teach authentic computational skills and support student research?**
- **How can faculty get the training and resources needed to teach genomics?**
- **What materials are needed to support learning goals and how can they be made organized and findable?**
- **What are the appropriate assessments and how can we give faculty a framework for developing their own?**

GEA Committees

- Bioinformatics Infrastructure

(Jason Williams, CSHL)

- Develop website
- Identify and host core tools

- Faculty Professional Development

(Vince Buonaccorsi, Juniata College)

- Survey faculty to ID training needs (self directed or with colleagues)
- Curate resources tailored to faculty need

- Student Learning Committee

(Anne Rosenwald, Georgetown U.)

- Develop lessons that teach the biology using the tools
- Develop tags that allow faculty to find appropriate lessons
- ID genomics core concepts

- Assessment Committee

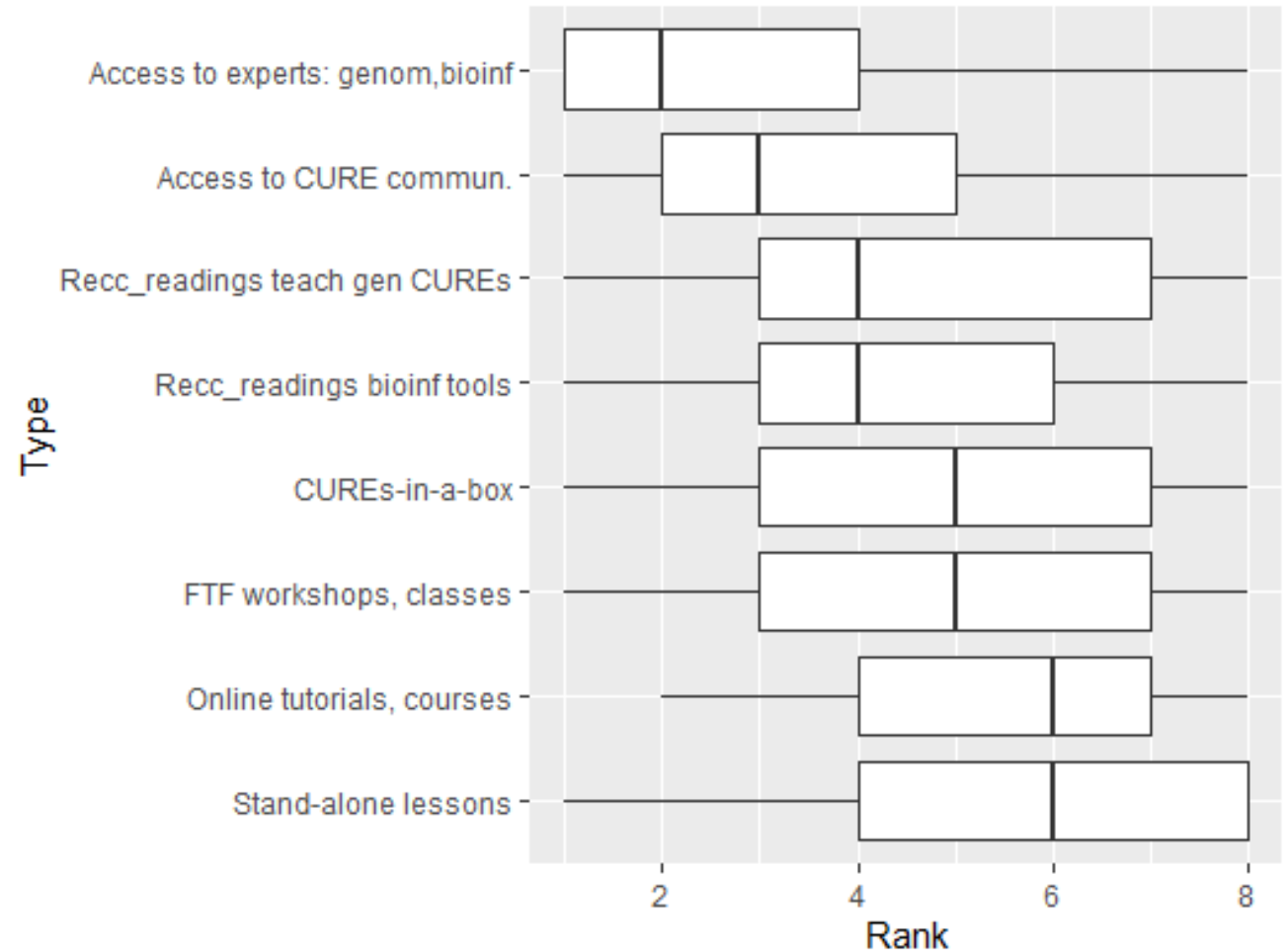
(Rochelle Tractenberg, Georgetown U.)

- Develop methods to evaluate CURE assessments

Progress so far

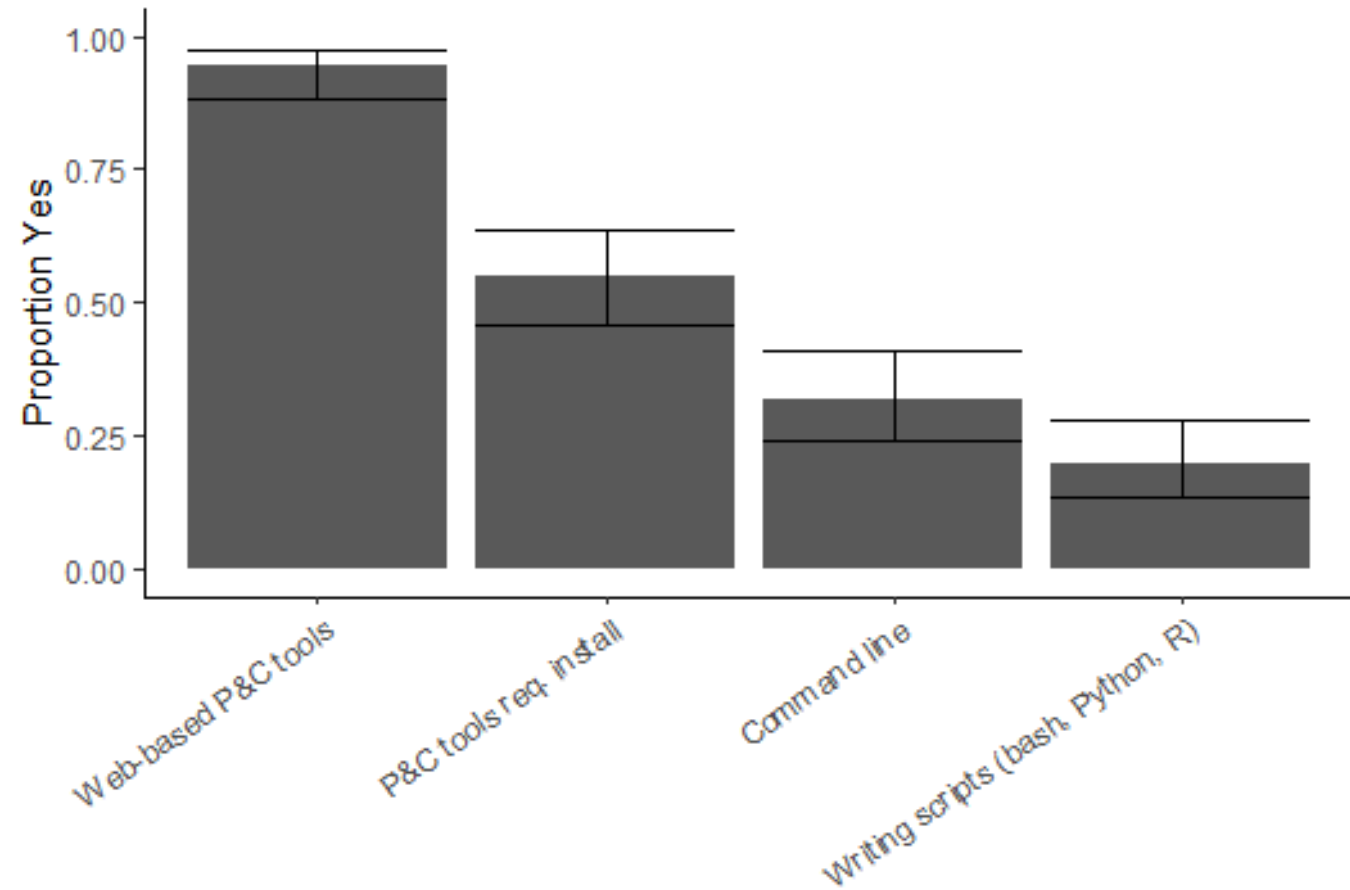
Survey of faculty development needs

Please rank the kinds of training resources you need in the future to implement a bioinformatics or genomics CURE.



Survey of faculty development needs

In the CURE you have or plan to implement, to what extent did/will analyses by students involve the following?



Pilot lessons

Development of pilot lessons

- **BLAST**
 - Introduce students to the use of the Basic Local Alignment Search Tool (BLAST) to identify related sequences and compare similarity between them

Development of pilot lessons

- **BLAST**
 - Introduce students to the use of the Basic Local Alignment Search Tool (BLAST) to identify related sequences and compare similarity between them
- **Understanding Eukaryotic Genes**
 - Familiarize students with a genome browser (UCSC)

Development of pilot lessons

- **BLAST**

- Introduce students to the use of the Basic Local Alignment Search Tool (BLAST) to identify related sequences and compare similarity between them

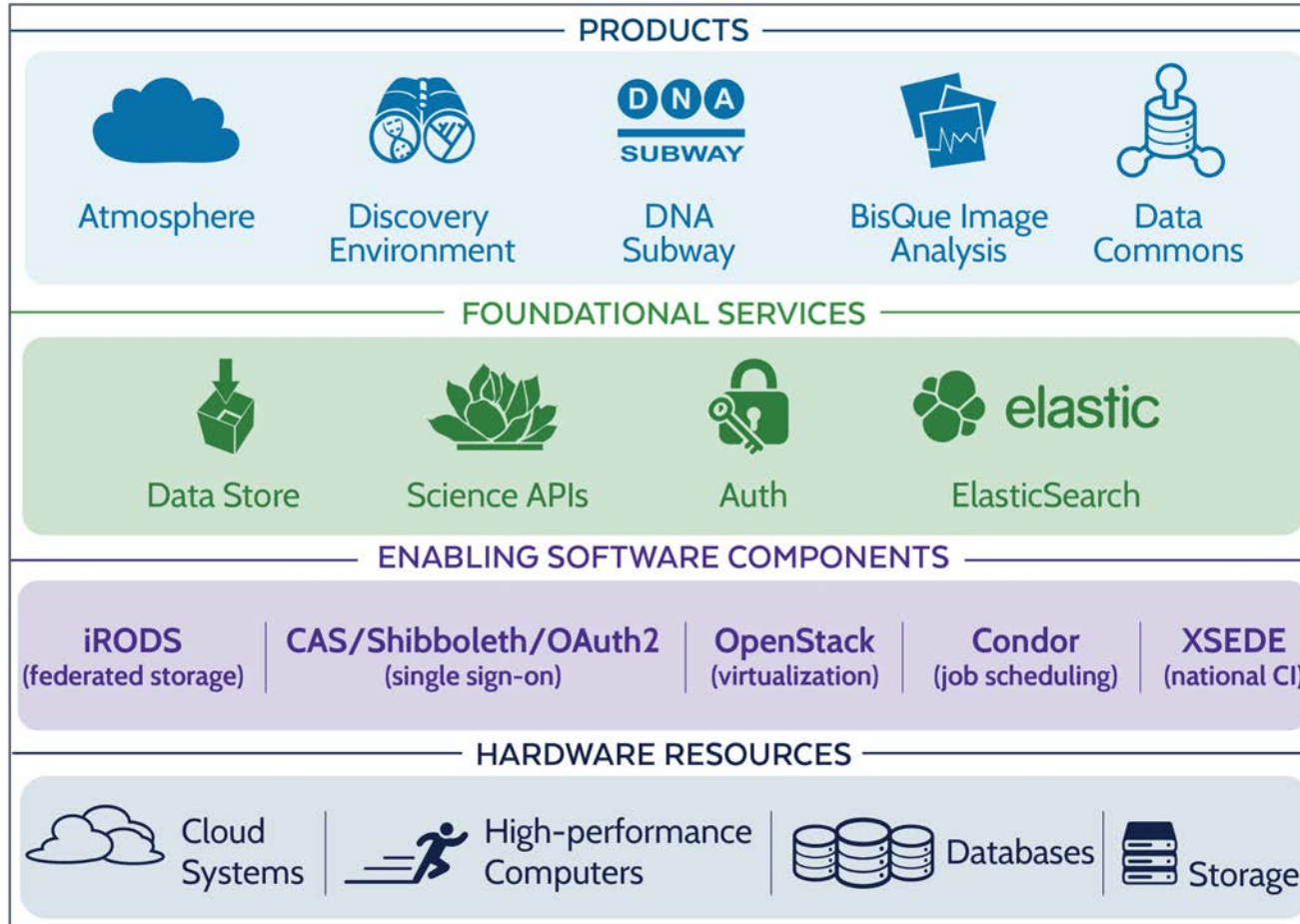
- **Understanding Eukaryotic Genes**

- Familiarize students with a genome browser (UCSC)

- **RNA-Seq**

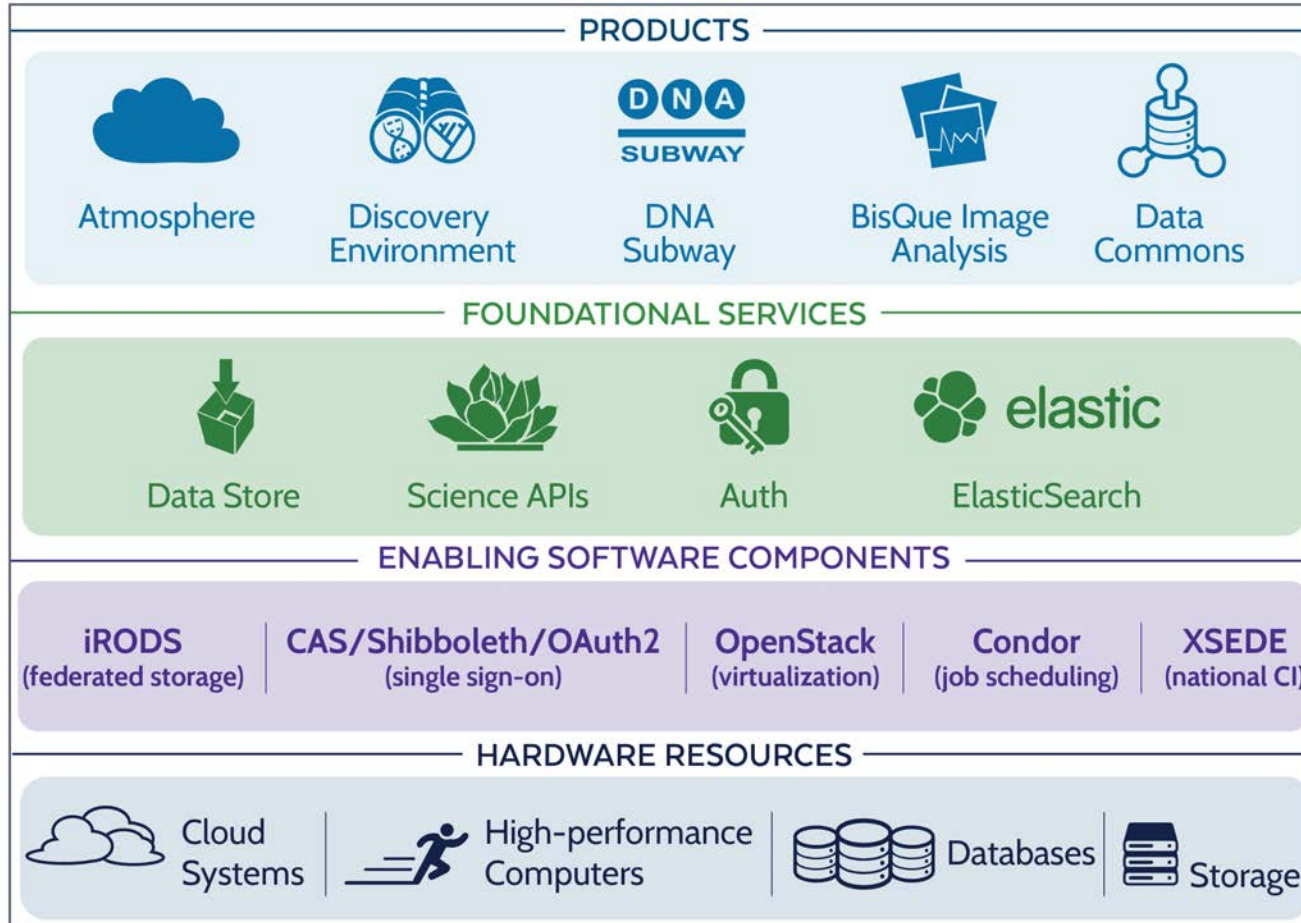
- 3 modules: Use RNA-Seq as a bridge to more advanced bioinformatics and data science (QC, quantification (Kallisto), visualization (UCSC/IGV))

Using CyVerse to deliver at scale



CyVerse.org

Using CyVerse to deliver at scale

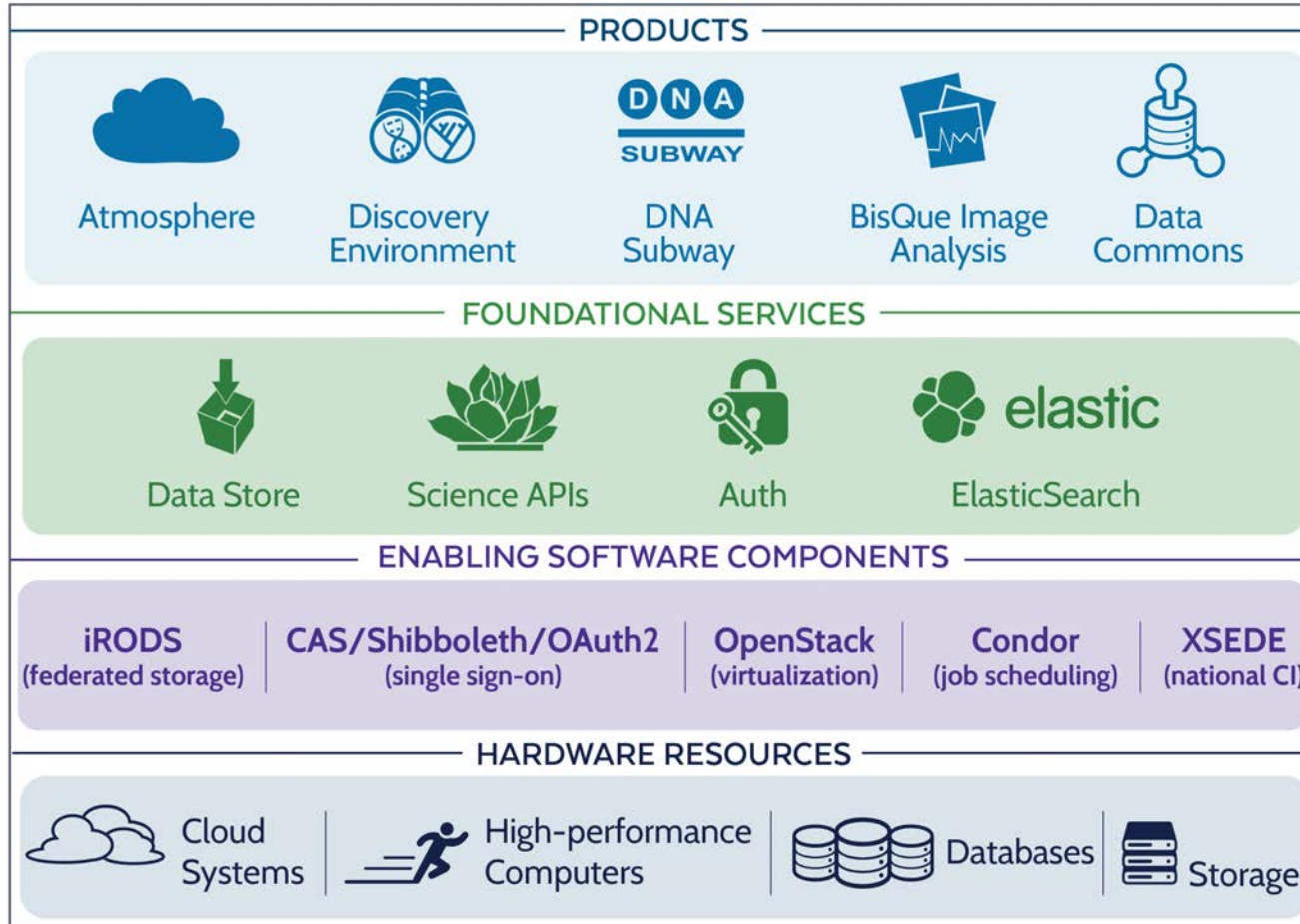


Data storage and computational power

Ability to store and work with terabytes of data on mid-scale to HPC compute

CyVerse.org

Using CyVerse to deliver at scale



Data storage and computational power

Ability to store and work with terabytes of data on mid-scale to HPC compute

Deliver a flexible computational environment

Applications are dockerized and deployed on demand

CyVerse.org

Development of pilot lessons

- **BLAST**

Innovation: stable scalable interface (Sequence Server)

- **Understanding Eukaryotic Genes**

Innovation: data streaming to UCSC browser

- **RNA-Seq**

Innovation: deployment of lesson on Jupyterlab within CyVerse VICE

Infrastructure goals

- **Updating and versioning**

Reduce faculty burden in keeping tools updated; predictable interfaces and release schedule

Infrastructure goals

- **Updating and versioning**

Reduce faculty burden in keeping tools updated; predictable interfaces and release schedule

- **Work with real tools and datasets**

Classroom friendly tools streamlined into GUI and notebooks

Infrastructure goals

- **Updating and versioning**

Reduce faculty burden in keeping tools updated; predictable interfaces and release schedule

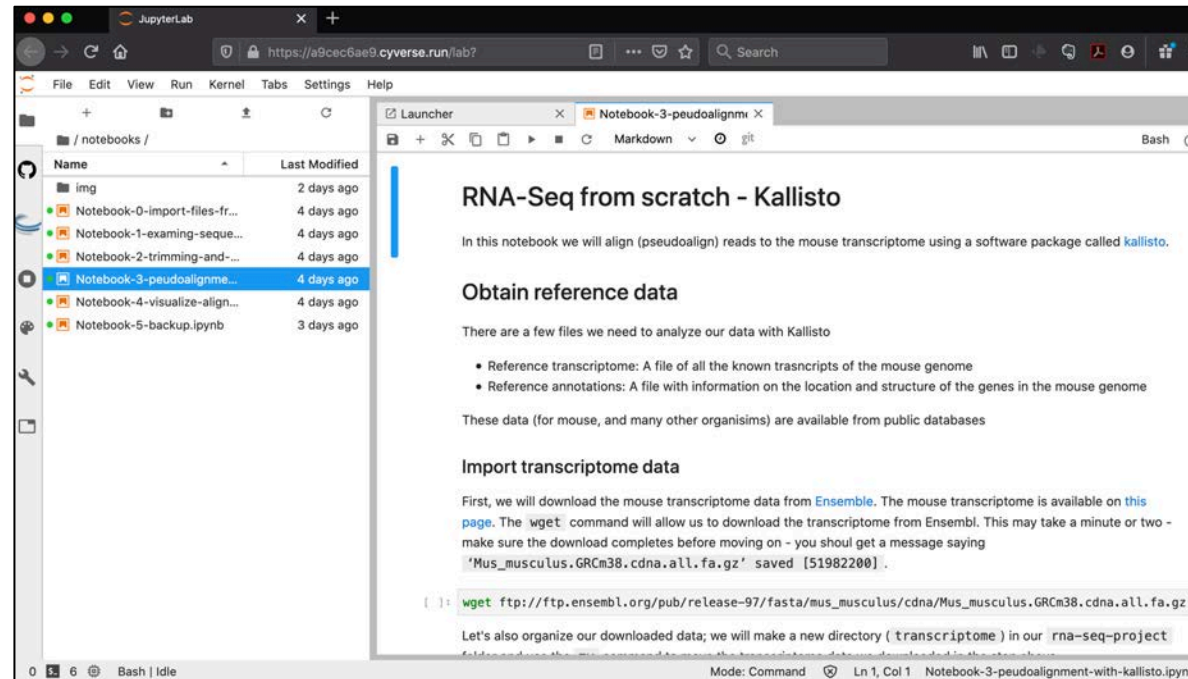
- **Work with real tools and datasets**

Classroom friendly tools streamlined into GUI and notebooks

- **Enable distributed projects/citizen science**

“Massively parallel undergraduates”

Example RNA-Seq Lesson



0 - Import from
SRA



1 - Sequence
QC



2 - Sequencing
Trimming



3 - Read
Pseudoalignment



4 - Read
Visualization



5 - Data
Backup



RNA-Seq analysis of Mouse Leptin Gene



Genomics Education Alliance

latest

Search docs

Lesson home

Launch Lesson on CyVerse

Jupyter Primer

Command Line Primer

Intro to RNA-Seq

Getting Data from NCBI

Assessing Data Quality

Trimming and Filtering Data

Pseudoalignment of Data Using Kallisto

Read the Docs

v: latest

Docs » Introduction to RNA-Seq: Leptin expression in mouse

Edit on GitHub

Introduction to RNA-Seq: Leptin expression in mouse

Submission Details

Submission Date	December, 2019
Version	1.0
Authors	- Jason Williams, Cold Spring Harbor Laboratory - Judy Brusslan, California State University Long Beach - Ray Enke, Jame Madison University - Matthew Escobar, California State University San Marcos - Vince Buonaccorsi, Juniata College
Corresponding author	Jason Williams (williams@cshl.edu)

Student learning objectives

- **Review, development, and testing of lesson content**

Adapting, modifying, creating lesson content, including from existing resources from the consortium; lesson pilot

Student learning objectives

- **Review, development, and testing of lesson content**

Adapting, modifying, creating lesson content, including from existing resources from the consortium; lesson pilot

- **Adaption and alignment of concepts/competencies**

Working to align undergraduate competencies (some may be underspecified or mutually incompatible)

Student learning objectives

- **Review, development, and testing of lesson content**

Adapting, modifying, creating lesson content, including from existing resources from the consortium; lesson pilot

- **Adaption and alignment of concepts/competencies**

Working to align undergraduate competencies (some may be underspecified or mutually incompatible)

- **Lesson tagging**

Developing sets of tags to enhance lesson discovery/adoption by faculty

Assessment – As per Rochelle Tractenberg's talk

The Assessment Evaluation Rubric: Promoting learnING and learnER -
centered teaching in higher education.

<https://doi.org/10.31235/osf.io/bvwhn>

Summary and take-home points

- **We have developed proof-of-concept computer and pedagogical infrastructure to make genomics easier to teach in the classroom**

Summary and take-home points

- **We have developed proof-of-concept computer and pedagogical infrastructure to make genomics easier to teach in the classroom**
- **CUREs are still difficult work - we want to target faculty that might not otherwise adopt genomics curriculum**
 - Faculty at less-resourced institutions (esp. minority-serving, rural, etc.)
 - Community college/freshman and sophomore

Summary and take-home points

- **We have developed proof-of-concept computer and pedagogical infrastructure to make genomics easier to teach in the classroom**
- **CUREs are still difficult work - we want to target faculty that might not otherwise adopt genomics curriculum**
 - Faculty at less-resourced institutions (esp. minority-serving, rural, etc.)
 - Community college/freshman and sophomore
- **We want to advance CURE-in-a-box concept**
 - Engaging researchers who want to make their work into CUREs

GEA Leadership

Doug Chalker*, Washington University in St. Louis
Vince Buonaccorsi*, Juniata College
Sally Elgin, Washing University in St. Louis, Emerita
Arthur Hunt, University of Kentucky
Wilson Leung, Washington University in St. Louis
Laura Reed, University of Alabama

Sandesh Subramanya, Austin Community College
Anne Rosenwald*, Georgetown University
Rochelle Tractenberg*, Georgetown University
Emily Wiley, Claremont Colleges
Jason Williams*, Cold Spring Harbor Laboratory

* PIs

Try the pilot lessons/join us!

<https://gea.qubeshub.org/lessons>



RCN-UBE #1827130