



National Institute of
General Medical Sciences



Ming Lei
Director, Division for
Research Capacity
Building

Building the NIGMS Sandbox for Cloud- based Learning

A Step Toward Democratizing Cloud Computing
for Biomedical Research

Cloud-Computing Presents Opportunities for Biomedical Researchers

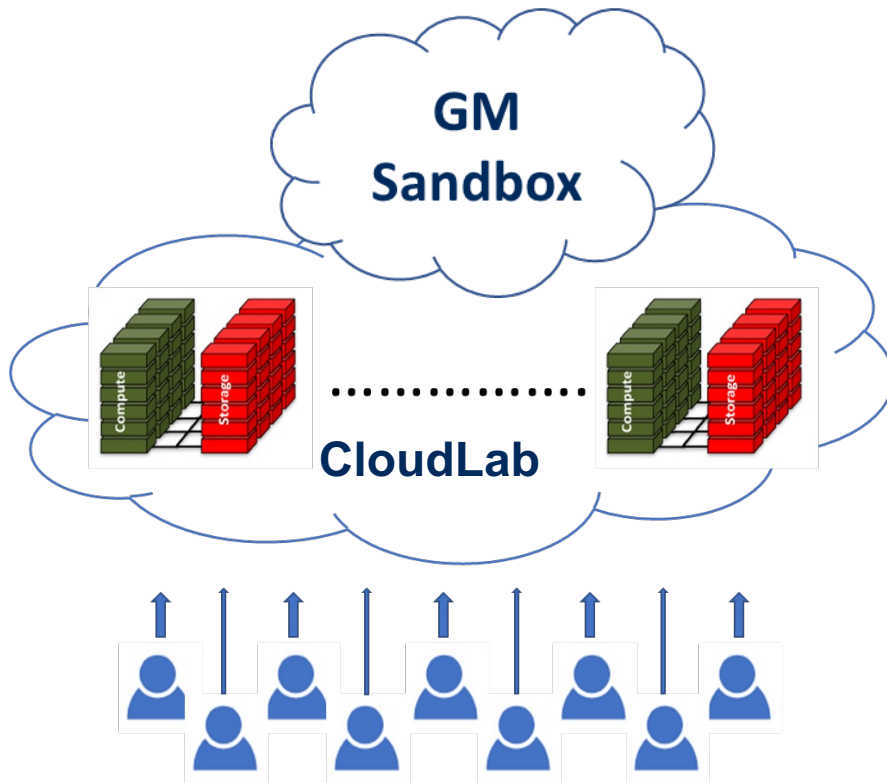
- Biomedical research is becoming increasingly data-driven
 - Need: **access** to big data and data analytical capacity for cutting-edge research
 - Opportunity: building infrastructure to provide such **access** broadly would open research opportunities for investigators and students with limited access to expensive research supplies and instrumentations
- Cloud computing can provide **broad access** to big data and data analytical capacity, breaking the insurmountable barrier of operating HPC facilities at every institution
- We have been working on a project aiming to enable investigators and students from under-resourced institutions to learn and conduct biomedical research through Cloud computing

A Step Toward Democratizing Cloud Computing for Biomedical Research - (January 2021-present)

- A collaboration of NIGMS, NIH ODSS/STRIDES, Google-GCP, Amazon-AWS, and GM- funded IDeA-INBRE and Diversity Enhancing Training programs
- Three sets of activities:
 - Training on Cloud computing fundamentals for investigators and students by GCP/AWS. Trained 500 participants through 60 six-hour sessions (each for 15 participants)
 - Migrating the IDeA National Resource for Quantitative Proteomics' data operation from its mainframe server to the Cloud, which led to a 30-fold increase in service capacity

“This cloud environment has been absolutely transformative for our resource.” *Alan Tackett, PI*
 - Building the NIGMS Sandbox, a cloud-based learning platform

The NIGMS Sandbox, A Cloud-based Learning Platform



Objectives

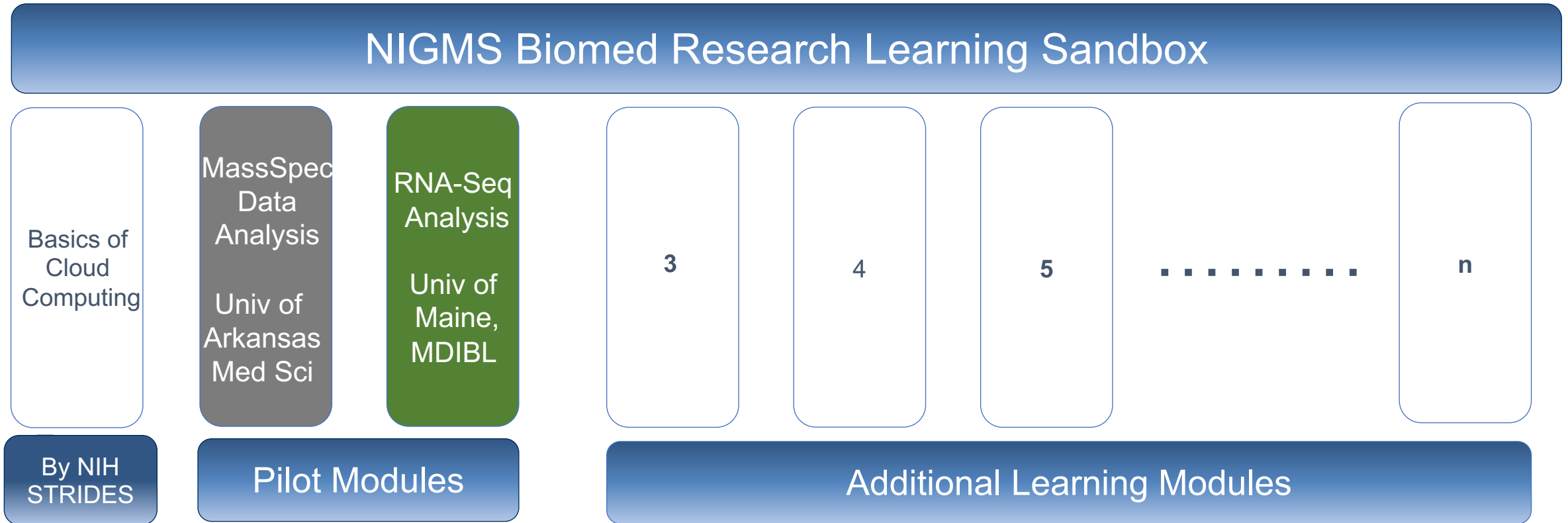
- A collection of learning modules each covering a technology or research method
- Interactive, enabling self-learning and running with users' own research data
- Accessible via Cloud user accounts by large number of users
- Users only pay for Cloud computing time

Pilots: Building Cloud-based Learning Modules



Train-the-Trainer:
GM-supported investigators gain programming skills to develop more cloud-based tools in the future

The NIGMS Sandbox that Can Expand with New Learning Modules

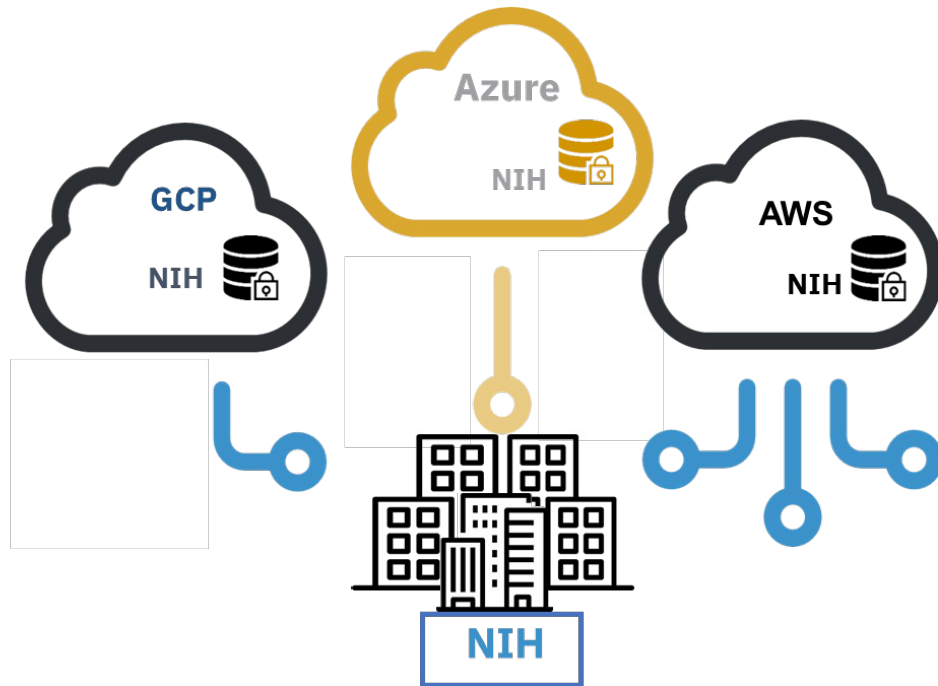


Plan forward: Funding 10 investigator teams to build additional learning modules (2022)

NIGMS Sandbox Version 1

Category	Module	Developer
Introduction of Bioinformatics	Fundamentals of Bioinformatics	Dartmouth College
Single-omics	RNAseq	University of Maine*
	Proteomics	U. of Arkansas for Medical Sciences*
	DNA Methylation Analysis	University of Hawaii
	Transcriptome Assembly	MDI Biological Laboratory
	Genomic Annotation & Pathway Analysis	University of Nevada Reno
Multi-omics and Metagenomics	ATACseq and Sc-ATACseq	U. of Nebraska Medical Center
	Metagenomics of Biofilm	University of South Dakota
	Multi-omic Analysis	University of North Dakota
Machine Learning	Biomedical Image Analysis	University of Arkansas
	Biomarker Discovery with ML/AI	University of Rhode Island
Data Science Methodology	Biomedical Data Science	University of San Francisco

NIH Cloud Lab: Providing Access to the Sandbox to Users

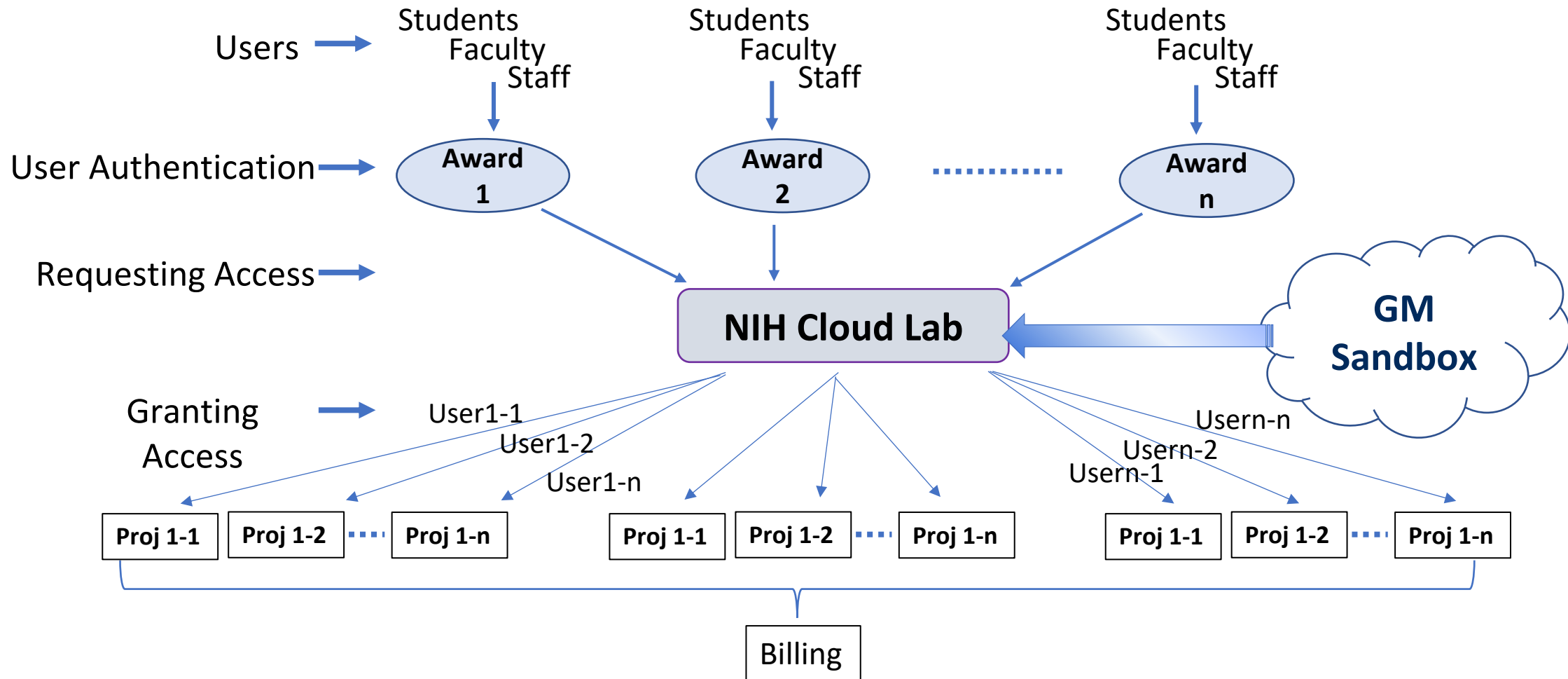


Full Access to the Cloud Console

- Deploy a full range of resources
- Advanced AI/ML capabilities
- Access to compute clusters
- Support from STRIDES Team
- On-demand training

**Advance Biomedical Research by Delivering Secure Access
to Industry Leading Cloud Service Providers**

Access the NIGMS Sandbox via NIH Cloud Lab



ATAC-Seq module Demo Video

NIGMS Sandbox Public Rollout

**June 23, 2023
1:30-3:30 PM EST**

<https://nih.zoomgov.com/j/1611976859?pwd=dS9KMk1KV1NZbWVIVIVxZGsrYzM0UT09>

Questions?