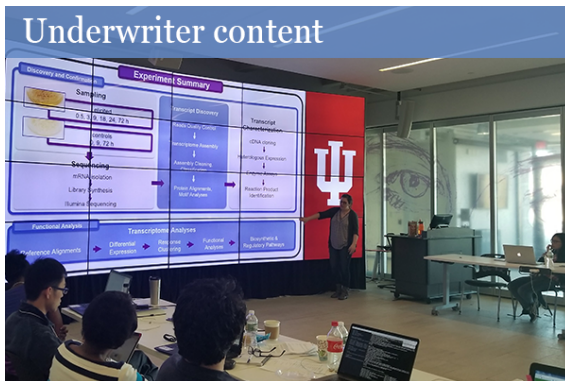


[Subscribe](#)[Login](#)[Home](#) [Archive](#) [Contribute](#) [Sponsor](#)[About](#) [Give Now](#)

# Teaching HPC and genomics side-by-side

The National Center for Genome Analysis Support (NCGAS) bridges the gap.

In late April, the **National Center for Genome Analysis Support (NCGAS)** gave its first national-level workshop in bioinformatics. NCGAS covered how to assemble short RNA sequence reads into expressed gene sequences (called a transcriptome) on large scale computers.

This analysis is very cheap and easily accessible to researchers new to genomics. But the analysis requires large memory machines to complete. For many biologists this is their first experience in larger-scale computing, meaning they are novices not only in genomic biology but are also new to using the Linux-based servers that are required.

**“The cheapest part of sequencing is now sequencing” ~C. Titus Brown**

Genetic sequencing has plummeted in cost for over a decade, giving rise to a boom in biology and a focus on designing

**Posted on 24 JUL, 2018**

**Sheri Sanders**

**Share this story**

---



**in**

**↻ Republish**

**Tags**

---

**National Center for  
Genome Analysis  
Support (NCGAS)**

better experiments to get the most information out of data. However, instruction on how to efficiently complete these tasks on clusters is often neglected, especially when it comes to highlighting skills such as job and data management that can dramatically reduce the time and storage it takes to complete an analysis.

This is critical, as researchers must pay for analyst time, graduate student time, and compute time and storage space. Post-sequencing analysis is now considered the cost- and time-limiting step in genomics projects, and one that depends heavily on properly using compute clusters.

NCGAS strove to fill this instructional gap by splitting the workshop evenly between HPC skills and genomic analysis. We provided the researchers with ready-to-run job scripts that serve as an unintimidating start point for their analysis.

**Bioinformatics**

**Genomics**

**Genetic  
sequencing**

**Workshop**

**High-performance  
computing (HPC)**

**Extreme Science  
and Engineering  
Discovery  
Environment  
(XSEDE)**

**Underwriter  
content**

However, we did not simply provide job scripts and mention their structure—we walked students through an activity where they had to act as a job scheduler while playing “job tetris” with jobs requiring different resources. We didn’t just list what resources were available to them, we toured the IU data center to show them physical machines and had them design data management plans around real research projects and available XSEDE resources.

Developing a deeper understanding of how these systems work helps students more intuitively grasp how to use them, giving them a foundation to build further necessary skills (i.e. optimizing individual jobs and using the right resource) and reduces anxiety in approaching new analyses.

“This was the first workshop of its kind that I have attended and one of the most useful workshops in my doctoral degree.”~participant survey

Giving biologists a framework for analyses and training them to efficiently use HPC systems not only gets researchers work done faster, but also reduces the load on HPC systems. Since many of these analyses require large memory and several days to complete, inefficient use of job requests and repeated failed jobs causes an unnecessary burden on the clusters.

Using resources poorly matched to the job only confounds this issue. Offering workshops in a domain-specific context is necessary to attract researchers to HPC skills.

Common biological workflows with high computational demands, especially workflows that attract new researchers to genomics, are an excellent opportunity to get buy-in from the research community to dedicate time to the basic HPC training they need to get to interesting questions. Targeting these large, common workflows also help to motivate HPC

managers to actively participate in training these communities to increase efficient system use.

Due to high demand from the community, NCGAS will be running this workshop again in early October.

---

**Join the  
conversation**

**Contribute**

Do you have story ideas or  
something to contribute? **Let  
us know!**

---

## OUR UNDERWRITERS

---

Thank to you our underwriters, who have supported us since the transition from International Science Grid This Week (iSGTW) into Science Node in 2015. We are incredibly grateful.

**[View all underwriters](#)**

## CATEGORIES

---

**Advanced computing**  
**Research networks**  
**Big data**  
**Tech trends**  
**Community building**

## CONTACT

---

**Science Node**

Email:

**[editors@sciencenode.org](mailto:editors@sciencenode.org)**

Website:

**[sciencenode.org](https://sciencenode.org)**



Copyright © 2023 Science Node <sup>™</sup> | [Privacy Notice](#) | [Sitemap](#)

Disclaimer: While Science Node <sup>™</sup> does its best to provide complete and up-to-date information, it does not warrant that the information is error-free and disclaims all liability with respect to results from the use of the information.