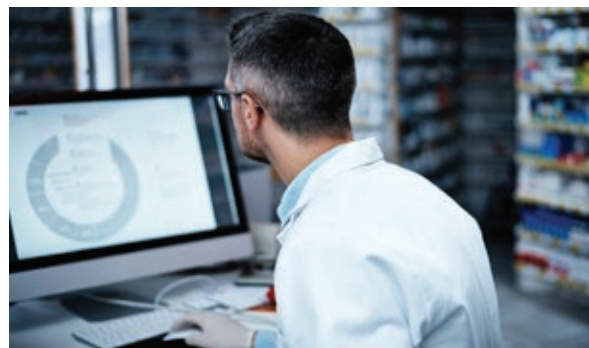


# THE STRIDES INITIATIVE SERIES: University of Washington's TOPMed Program Boosts Collaboration With the Help of the Cloud

## Why TOPMed Started Using Cloud Services

The [Trans-Omics for Precision Medicine \(TOPMed\)](#) program's data coordinating center at the University of Washington's [Genetic Analysis Center \(GAC\)](#) processes whole genome sequence (WGS) data on its high-performance computing cluster. In 2014, the GAC began using cloud services to better analyze data and to share the data it processed with other scientific collaborators—within the TOPMed program and beyond. Using the cloud helped GAC leverage their expertise in statistical genetics and high-performance computing with TOPMed principal investigators and the broader research community. As David Levine, principal research scientist in the Biostatistics Department at the University of Washington said, “Our biggest benefit in working with cloud services – and now doing so through the STRIDES Initiative – is that we get to share the wealth. By uploading our data to the cloud and making it accessible to researchers, we enable others to use our software, particularly other investigators who don't operate at the scale we do, and who wouldn't have access to this data otherwise.”

By the numbers, the WGS project collaborates with TOPMed's 32 working groups, more than 80 studies, and over 1,000 affiliated investigators. Using the STRIDES Initiative as an avenue to access cloud services allows for more effective collaboration between these groups, “bringing [the project] closer to predicting an individual's risk for diseases” noted Levine. “No one has done this before at this scale. To make these data work together and turn them into knowledge is incredibly challenging, but we're uniquely equipped to do the job.”



## Making the Switch to the STRIDES Initiative

In 2018, the GAC team became one of the first users of the Google Cloud Platform through the STRIDES Initiative. The STRIDES Initiative not only offered the GAC team favorable pricing on cloud services and tools, but also a hands-on customer support model with regular meetings and a responsive team of experienced cloud service experts. “[The STRIDES Initiative team] made sure we didn't feel we were working in isolation and that we really had the support we needed. Just having access to that support is helpful,” said Levine.

## Challenges

- » Scaling resources
- » Getting data back to researchers easily
- » Achieving results quickly

## Key Results

- » Democratizes data access
- » Catalyzes meaningful research by larger community
- » Maximizes the productivity and value of the TOPMed program

## Impact of Cloud on TOPMed Research

Data generated by TOPMed is available through NHLBI's [BioData Catalyst](#), a developing cloud-based platform providing researchers with tools, applications, and workflows in secure workspaces. At this time, only a small group of early users can access the system, but it will be more widely available in the future. By sharing its data through the BioData Catalyst, the TOPMed project joined the cloud-based data science ecosystem, a community of users who work collaboratively to solve technical and scientific challenges.

Leveraging the STRIDES Initiative to access cloud services led the University of Washington to:

- » Disseminate software tools to be used by other researchers
- » Share genetics data with other researchers
- » Access unlimited computing resources to run large analyses
- » Avoid purchasing additional equipment by using the cloud services as supplementary storage for the project's existing computing infrastructure



---

I feel very positive about the whole [STRIDES Initiative] effort. Having tight integration and great support from the IT group within the STRIDES Initiative was very important, and that was part of the project that we didn't realize we'd need when we started. The regular communication is the key that made this work so well.

—**David Levine**, principal research scientist, University of Washington



## Future Plans For Using Cloud

As they move forward using cloud services, the team at the University of Washington expects to continue shifting away from local computing and its associated expenses as well as continue supplementing their existing computational infrastructure with cloud services. Eventually, the GAC team will focus on what will become TOPMed 2.0—collecting other -omics data and integrating it with the DNA data already in the cloud. Having already successfully worked in the cloud, the GAC team is well positioned for its future TOPMed 2.0 research.

### About TOPMed

The [National Heart, Lung, and Blood Institute's \(NHLBI's\) Trans-Omics for Precision Medicine \(TOPMed\) program](#) aims to generate scientific resources to improve our understanding of heart, lung, blood, and sleep disorders through advancing precision medicine by collecting whole-genome sequencing and other -omics (e.g., RNA transcripts, proteins, metabolites, etc.) data. NHLBI leverages the STRIDES Initiative for several of its TOPMed projects, including the Whole Genome Sequencing (WGS) project based out of the University of Washington. The WGS project studies the genetic architecture of these diseases in order to improve scientific understanding of the fundamental biological processes and to improve their diagnosis, treatment, and prevention.