

Broad Institute, Google to analyze genomics data

26 June 2015 | News | By BioSpectrum Bureau

Broad Institute, Google to analyze genomics data



Broad Institute of MIT and Harvard are teaming up with Google Genomics to develop a computing infrastructure that will help store and process enormous datasets, as well as create tools to analyze such data in biomedical research.

As a first step, Broad Institute's Genome Analysis Toolkit, or GATK, will be offered as a service on the Google Cloud Platform, as part of Google Genomics. The goal is to enable any genomic researcher to upload, store, and analyze data in a cloud-based environment that combines the Broad Institute's best-in-class genomic analysis tools with the scale and computing power of Google.

GATK is a software package developed at the Broad Institute to analyze high-throughput genomic sequencing data. GATK offers a wide variety of analysis tools, with a primary focus on genetic variant discovery and genotyping as well as a strong emphasis on data quality assurance. Its robust architecture, powerful processing engine, and high-performance computing features make it capable of taking on projects of any size.

GATK is already available for download at no cost to academic and non-profit users. In addition, business users can license GATK from the Broad. To date, more than 20,000 users have processed genomic data using GATK.

The Google Genomics service will provide researchers with a powerful, additional way to use GATK. Researchers will be able to upload genetic data and run GATK-powered analyses on Google Cloud Platform, and may use GATK to analyze genetic data already available for research via Google Genomics. GATK as a service will make best-practice genomic analysis readily available to researchers who don't have access to the dedicated compute infrastructure and engineering teams required for analyzing genomic data at scale. An initial alpha release of the GATK service will be made available to a limited set of users.

"Large-scale genomic information is accelerating scientific progress in cancer, diabetes, psychiatric disorders, and many other diseases," said Mr Eric Lander, president and director of Broad Institute. He added, "Storing, analyzing, and managing these data is becoming a critical challenge for biomedical researchers. We are excited to work with Google's talented and experienced engineers to develop ways to empower researchers around the world by making it easier to access and use genomic information."

"Broad and Google share a culture of collaboration and open access to data. Google Genomics is helping scientists make genomic information more accessible and useful. By making Broad's GATK available through the Google Cloud Platform, we hope to accelerate great science," said Mr David Glazer, director, Google Genomics.

In keeping with the Broad's mission to foster openness and innovation, this collaboration will be non-exclusive. Broad and Google will each continue to engage with other community members on genomic projects to empower research worldwide.