

# Collaboration Yields Largest-ever Cloud-based Genomic Analysis

Glenna Picton, Baylor College of Medicine



With their participation in the completion of the largest cloud-based analysis of genome sequence data, researchers from the Baylor College of Medicine (BCM) Human Genome Sequencing Center are helping to usher genomic scientists and clinicians around the world into a new era of high-level data analysis.

“The mission of the Baylor Human Genome Sequencing Center is to drive genomics and genomic analysis to be at the leading edge of everything in the field,” said Dr. Jeffrey Reid, assistant professor in the Human Genome Sequencing Center at BCM, who led the BCM portion of the project. “In terms of analysis, the future of genomic research and genomic medicine is in the cloud. We are very much going towards more computing and not less.”

Together with the Platform-as-a-Service company DNAnexus and Amazon Web Services, the largest provider of cloud computing, BCM sequenced the DNA of more than 14,000 individuals- 3,751 whole genomes and 10,771 whole exomes- using next generation sequencing.

An exome contains all the genes in a genome and are the part of the genome that provides the blueprints for proteins.

The individuals whose genetic material was sequenced are part of the Cohorts for Heart and Aging Research in Genomic Epidemiology Consortium or CHARGE project aimed at advancing understanding of human genetics and the contributions to heart disease and aging.

Reid gave a presentation on the project at the American Society of Human Genetics annual meeting in Boston.

## Collaboration Yields Largest-ever Cloud-based Genomic Analysis

Published on Scientific Computing (<http://www.scientificcomputing.com>)

---

The BCM Human Genome Sequencing Center-developed Mercury pipeline, a semi-automated and modular set of tools for the analysis of next generation sequencing data in both research and clinical contexts, was an integral part of the project. The pipeline identifies mutations from genomic data, setting the stage for determining the significance of these mutations as a cause of serious disease.

Led by Dr. Eric Boerwinkle, professor and director of the Human Genetics Center at The University of Texas Health Science Center at Houston and associate director of the Human Genome Sequencing Center at BCM, the CHARGE project involves more than 300 researchers across five institutions around the world. The cloud-based analysis makes it possible for the large group to have access to an expansive network of data over a server that is HIPAA certified to not compromise patient privacy.

“The collaboration between the CHARGE consortium and the Human Genome Sequencing Center is leading to discovery of those genes contributing to risk of the most important diseases plaguing the U.S. population across all age groups,” said Boerwinkle. “Ultimately, these discoveries forge a path toward novel therapeutics and diagnostics. The use of cloud computing and collaboration with DNAnexus is allowing us to achieve our goals faster and in a more cost-effective manner.”

Boerwinkle will give an updated presentation Nov. 15 at the Cold Spring Harbor Laboratory’s Personal Genomes & Pharmacogenomics Meeting.

“Having access to this much data was unique,” said Reid. “Many institutions do not have the local compute resources and infrastructure to support large scale analysis projects like this one, so we were lucky to come together with DNAnexus and Amazon Web Services to make this project possible.”

The project required approximately 2.4 million core-hours of computational time, generating 440 TB (terabytes) of results and nearly a petabyte of storage that took place over a four-week period.

By comparison, the 1,000 genomes project sequenced 2,535 exomes and required 25 TB of data.

“It is very important for us to create a centralized space where researchers from all over the world can come and collaborate with the data,” said Reid. “This project creates expansive access to this data over a protected network that will advance research.”

**Source URL (retrieved on 12/12/2013 - 10:39pm):**

<http://www.scientificcomputing.com/news/2013/10/collaboration-yields-largest-ever-cloud-based-genomic-analysis>