March 8, 2018

Dear GNomEx users,

I am pleased to announce a major evolution in genomics data storage, organization and analysis at Huntsman Cancer Institute. HCI has recently purchased a license for a cloud-based genomics data storage and bioinformatics analysis platform called Seven Bridges Genomics (https://docs.sevenbridges.com/). The decision to move to cloud-based genomics data storage and analysis platform was made after extensive evaluation and vetting by members of the faculty advisory committee for the High-Throughput Genomics and Bioinformatics Shared Resource in close consultation with the Directors of those shared resources, and Research Informatics – my sincere thanks to them for their work on this initiative, especially K-T Varley, David Nix and Tim Parnell.

This Seven Bridges platform will provide GNomEx users with exciting new capabilities including the ability to:
1. Store and manage their genomics data in the Amazon cloud.
2. Share data with collaborators around the world.
3. Run pre-built bioinformatics analysis apps and pipelines through a point-and-click interface.
4. Access public datasets, such as TCGA, TARGET, and CPTAC.

All of these capabilities will be easy to access without programming skills by logging into a website created by Seven Bridges Genomics. The Bioinformatics Core and Seven Bridges will provide on-site training sessions to facilitate adoption of the platform.

PI's, please choose a representative from your group to attend a training session and send their email address to Tim Parnell (Timothy.Parnell@hci.utah.edu), Associate Director of the Bioinformatics Shared Resource. Faculty are welcome to attend a training session as well.

GNomEx will still be used to submit sequencing orders to the High-Throughput Genomics core facility and record information about experiment details, but it will no longer be a long-term storage repository for genomics data (see below).

Transfer of existing datasets:
The Bioinformatics core facility will transfer existing genomics data in GNomEx to the Amazon cloud Glacier archive storage (NOTE: this does not affect the UGP, B2B, and clinical instances of GNomEx). HCI will pay for the storage at Amazon Glacier of all datasets currently in GNomEx for 6 months, starting May 1, 2018. After 6 months, unclaimed datasets will be deleted.
PI's will need to take ownership/claim their datasets by October 1, 2018, by either:

1. Claiming their Seven Bridges account under the University of Utah license (recommended). To enable this, PI lab accounts have been created and grouped by their primary department affiliation.
2. Transferring their data to another cloud data storage provider (e.g. DNAnexus, Amazon S3/Glacier, Google Cloud, Microsoft Azure Cloud).

3. Downloading their data to a local compute facility (e.g. CHPC, departmental IT).
4. Archiving data on single, consumer-based hard drives and computers is generally discouraged due to their propensity for failure, loss, etc.

If a PI does not want their data automatically transferred to the Amazon cloud through the Seven Bridges Platform, please immediately contact Tim Parnell (Timothy.Parnell@hci.utah.edu).

New datasets will only be retained in GNomEx for 6 months:
Beginning May 1st, newly generated data files will only be retained for 6 months in GNomEx. Results from analyses performed by the Bioinformatics Shared Resource Core facility will still be distributed through GNomEx as a convenience (same as before), unless otherwise directed. After 6 months of inactivity, projects will be archived automatically to Amazon Glacier through Seven Bridges if the PI has an active Seven Bridges account; otherwise users will need to make their own arrangements to store their data elsewhere (see the alternative options above).

PI's will be responsible for storage and compute costs:
Once a PI takes ownership of the data in the Amazon cloud they will be responsible for paying for the long-term storage of their data and any bioinformatics costs incurred from analyses performed in the Seven Bridges platform. HCI's purchase of the Seven Bridges license enables users to take advantage of significant cost savings for cloud-based storage and computing. For example, if you pay $1,100 for one lane of paired end 125bp sequencing through the High-Throughput Genomics core facility, you generate approximately 50 GB of data. If you choose to have the data automatically uploaded to the Amazon cloud through the Seven Bridges platform it will cost approximately $7.00 to store the 50GB in Amazon S3 for 6 months while you perform bioinformatics analysis. If you use the Seven Bridges app for exome analysis it will cost approximately $2.00 per sample. When you have completed your analyses, you can use the Seven Bridges platform to archive your data in Amazon Glacier storage and it will cost approximately $3.00 per year to store the 50GB of data. This example serves to illustrate that the cost of data storage and compute is marginal compared to the cost of generating the data, and this system enables you to manage and analyze your data efficiently.

Note that the Bioinformatics Shared Resource will continue to maintain extensive local compute infrastructure through HCI and the CHPC that is made available to the larger community, especially for those who do not wish to use or otherwise are incapable of using external resources due to, for example, IRB restrictions. The Bioinformatics Core is actively investigating making their best-practices workflows executable on both local CHPC resources and the Seven Bridges platform.

If you have any questions or concerns about this new system please contact Tim Parnell (Timothy.Parnell@hci.utah.edu) or David Nix (David.Nix@hci.utah.edu), and/or attend one of the upcoming training sessions. We will be in touch shortly about them.

I look forward to sharing these improved bioinformatics capabilities and efficiencies with you, and the ways they can accelerate discovery and translation.

Best Regards,

Bradley R. Cairns, PhD
Investigator and Senior Director of Basic Science, Huntsman Cancer Institute
Chair, Department of Oncological Sciences, University of Utah
Investigator, Howard Hughes Medical Institute