

The Cancer Genomics Hub

Submitted by: Linda Rosewood, Program Director, rosewood@ucsc.edu. 831-459,4440

David Haussler, Principal Investigator
Linda Rosewood, Program Director
Mark Diekhans, Technical Director
Brian Sanford, Project Manager
Christopher Wilks, Software Engineer
Melissa Cline, Software Engineer
Erich Weiler, System Administrator
Haifang Telc, System Administrator
Paul Tatarsky, System Administrator
John Groboske, Support Technician
Daniel Wilson, Security Compliance Manager
Elizabeth Mitchell, Administrative Specialist
Barbara Jolley, Research Analyst

Project Description

One way to understand CGHub's role in genomics is to look at the US. space program. Genomics is only in its infancy, much like the space program in the late 1950s. Where the Mercury Program had limited goals of getting a manned capsule to orbit, its success was a necessary step toward the final goal of the Apollo program of a moon landing. CGHub is like project Mercury, where in a first and significant step, an atlas of cancer genomics data will be made available to researchers using modern software design (APIs), customer-focused user support, and standards that make collaboration possible.

CGHub, which was designed, built, and operated by UC Santa Cruz, is the largest repository of biologic information in the world. Its adoption of a bittorrent-based method of transferring data leverages a tool long used by other disciplines and which now has additional features needed for this particular research discipline. CGHub demonstrates how a UC campus could quickly respond to an opportunity, establish a new unit, secure the data, and deliver a valuable resource to researchers in only 6 months. It will continue to be the model for other large repositories of genomics data.

CGHub is best described in the videos linked below, as they are the most dynamic and immediately understood method of seeing the role that CGHub plays in current cancer research. CGHub provides a valuable repository, a set of tools--and support-- for genomics research, and helps establish standards and

encourages collaboration.

Press release, May 2012:

[UC Santa Cruz builds national data center for cancer research.](#)

[Cancer Genomics Hub](#), a presentation to Usenix by Professor David Haussler. An invitation to the unix community to contribute to the future of cancer research.

[Cancer Genomics Hub](#), CENIC Innovations in Networking award, presentation by Linda Rosewood. This is a long file. Advance the video to the 1:00 hour mark to see Linda's talk.

[Cancer Genomics Hub, Gene Torrent, presentation by Chris Wilks](#). Presentation starts at 19:35. This talk describes GeneTorrent, open source software maintained by CGHub.

The timeframe of implementation

CGHub began design in August of 2011. Hardware was installed November 2012. Data migration of ~250TB completed in December 2011. CGHub was live or downloaders, including a technical support team and infrastructure in January of 2012. New submissions began in February 2012. The GUI data browser was released in beta in November 2012, and will be formally released May 2013.

Objective customer satisfaction data (not to exceed 2 pages)

Our users are not very responsive to requests for satisfaction surveys, but here are a few unsolicited comments:

"got it. resolved! Thanks, as always. You guys are awesome at helping people out!

Yesterday the submission guide for CGHub made it to my desk, for the first time; after thumbing through it I wanted to congratulate you and your team for exemplar work, as both the design and exposition come across as crisp, clear, and very high in quality. At this point there's no doubt in my mind that if the operation of CGHub is given the same level of care as its thoughtful design portends, then the switch from SRA & dbGAP will be manyfold worth the cost.

The information provided by you was indeed helpful! Thanks a ton! I am actually impressed by the quick response from your support system.

The user from the ticket 1729 Wanted us to know that he thinks we are doing a great job and is very impressed with the way CGHub is working for him as opposed to the old way of downloading TCGA data. This is a nice comment because this user initially came to us with a very complex problem (NIH login issue) and was still happy enough with our service that he went out of his way to tell me on the phone how well CGHub is working.

The user from 1916 said that his IT department said that he had to run GT with the rate-limit option turned on, and could only run it in the evening and over weekends because it was saturating his network bandwidth. An accolade in a roundabout way.