# New HIV browser gives researchers access to valuable data from vaccine trials

May 29 2008

A new HIV data browser developed by the University of California, Santa Cruz, and the nonprofit organization Global Solutions for Infectious Diseases (GSID) will give researchers access to a wealth of data collected during clinical trials of an AIDS vaccine. Although the vaccine did not succeed in preventing infections, the clinical trial generated a huge amount of valuable data for researchers studying how the virus evolves and causes new infections.

Modeled on the UCSC Genome Browser, the GSID HIV Data Browser is the brainchild of Phillip Berman, professor and chair of biomolecular engineering in UCSC's Baskin School of Engineering. Berman helped oversee the clinical trials, which ended in 2003, when he was senior vice president for research and development at VaxGen, the company that developed the vaccine and conducted Phase III clinical trials in North America, Europe, and Thailand.

"After the trials concluded, I spent a couple of years trying to think what was the most important thing I could do for HIV research," Berman said. "I concluded it was using new technology to preserve the data from these clinical trials and present it in a form useful to the scientific community."

In 2004, Berman cofounded GSID, based in South San Francisco and dedicated to combining knowledge and expertise from the biotechnology industry and the public health sector to address infectious disease problems in the developing world. He joined the UCSC faculty in 2006.

"Despite the fact that the vaccine trial didn't work, a huge amount of useful information was obtained," Berman said. The "North American" trial included about 60 different clinical sites in North America and one site in the Netherlands. Of particular value to researchers are the genetic sequences of the viruses that infected participants during the trial.

"The trial represented the only up-to-date broad survey of virus sequences from new infections that had ever been carried out," Berman said. "Every time there was a new infection in the vaccine or placebo group, the virus was sequenced. The sequence information provides the best picture we have about what the immune system sees when there is a new infection."

This is important, Berman said, because other major repositories of HIV sequence data are not annotated for the time after infection, the clinical status of the patient, or the histories of the specimens sequenced. That limits their usefulness for studying such a rapidly evolving virus.

HIV is highly mutable and evolves in response to attacks by the immune system. As a result, HIV isolated from a patient years after the initial infection is genetically different from the virus that caused the infection in the first place. A vaccine should target the most infectious form of the virus, Berman said. Yet all the vaccines tested so far have been based on viruses isolated from patients with longstanding infections.

"A current hypothesis in HIV vaccine research is that the antigenic structures of HIV viruses that mediate new infections differ from those recovered from people long after infection," Berman explained. "The specimens in this set represent the largest group from new infections that has ever been collected."

Besides viral genome-sequence data, the database links to a repository of preserved specimens (blood samples and cells) that researchers can

access from GSID and the National Insitutes of Health (NIH) for further study.

"This is the first time that an HIV sequence database has been linked to a specimen repository and a database of clinical information," Berman said. "These clinical specimens are longitudinal, collected from the same person during a two-year follow-up period. This will allow investigators to study the evolution of the virus and the evolution of the immune response and clinical outcomes."

At UCSC, Berman teamed up with the Genome Browser group to develop a browser for the sensitive clinical data collected during the vaccine trial. Jim Kent, associate research scientist for the UCSC Genome Browser and principal investigator on the project, said it was the first time his group had worked with data from participants in a clinical trial.

"This data must be handled differently and great care taken with confidentiality," Kent said. "We learned from this project how to build the infrastructure to cope with that. This will be useful for other medical projects, such as cancer genomics, in the future."

Fan Hsu, director of proteomics for the UCSC Genome Browser, said the emphasis on security was very different from past projects. "Before, everything we have worked on is totally open, totally public. With the GSID project, only authorized users can access the data, so we needed to set up special controls," Hsu said.

How to display the very large number of HIV sequences on the browser was another challenge. "Our original genome browser has only one reference genome. For this HIV database, we have about 350 infected people and more than 1,000 sequences," he said.

Hsu and software developer Galt Barber adapted the genome browser software to accommodate the large number of HIV sequences and the data security along with interactive selection criteria for viewing the data. As the project evolved, Hsu also coordinated the transfer of the software to GSID. The UCSC team, which also included Erich Weiler, Robert Kuhn, and Ann Zweig, worked nights and weekends to bring the new browser online.

The resulting GSID HIV Data Browser is a customized version of the UCSC Genome Browser. It provides researchers with searchable demographic and clinical data from volunteers who became HIV infected during the VaxGen clinical trial. The browser allows users to align viral sequences with one another and with reference or consensus sequences.

"This is something where the university can make a difference, because the private sector is not so interested in vaccines; they're not so profitable," Kent said. "There is very little economic incentive to develop an AIDS vaccine, but there is a tremendous humanitarian incentive."

Kent hopes that just as the UCSC Genome Browser has continued to build the collaborative nature of the genomics research community, this HIV data browser will help motivate the AIDS research community to work together and pool their data.

Vaccine development efforts have been repeatedly frustrated. An HIV vaccine candidate developed by the pharmaceutical company Merck recently failed in clinical trials cosponsored by NIH. "The recent failure of the Merck HIV vaccine has thrown the field into turmoil," Berman said. "All the best ideas for an HIV vaccine in the past 20 years have failed. The information in this database is now more critical than anyone could have imagined. It tells us what's being transmitted."

The next phase of the HIV browser project involves releasing the sequence data from infected participants in the Phase III clinical trial that VaxGen conducted in Thailand.

"In the future, the database will be expanded to allow associations between virus sequences, clinical data, immune response data, and host genetics," Berman said. "We hope to eventually include data from other HIV vaccine trials sponsored by the NIH, private companies, and other HIV vaccine research organizations."

For information on accessing the GSID HIV Data Browser and background on the clinical trials, visit www.gsid.org/index02.html

Source: University of California - Santa Cruz