Assembly Hubs and Genome Browser in a Box Makes Viewing Private Annotations and Custom Sequences Easy

B. J. Raney, M. Haeussler, A. S. Hinrichs, A. S. Zweig, D. Karolchik, H. Clawson, J. Casper, M.L. Speir, B. Lee, G. Barber, P. Fujita, D. Haussler, W. J. Kent

University of California at Santa Cruz, Center for Biomolecular Science and Engineering, School of Engineering, Santa Cruz, CA, USA

UNIVERSITY OF CALIFORNIA
SANTA CRUZ

## Assembly Hubs: View your sequence with the UCSC Genome Browser

Assembly Data Hubs allow researchers to use the UCSC Genome Browser to view their own sequences with associated annotation, without the requirement that UCSC support a browser on that sequence.   An Assembly Data Hub is a set of Internet-accessible data files that define the reference sequence to be used for a browser instance, as well as all the data files that define the annotation for that sequence.   User sequences can be as complex as whole genome assemblies, or just a few scaffolds from a re-sequencing project. The end-user maintains control over the sequence and the annotations and can update them at her convenience.

Assembly Data Hubs are an extension to Track Data Hubs, which allow user-level annotation on existing reference sequences.   Track Data Hubs are flat files that are Internet-accessible.   They need not exist on the same system, but can be distributed widely. Track hub annotations are stored as compressed binary indexed files in BigBed, BigWig, BigGenePred, BAM, HAL, or VCF/tabix format. When a hub track is displayed in the Genome Browser, only the relevant data needed to support the view of the current genomic region is transmitted to UCSC, rather than the entire file. All transmitted data is cached on a UCSC server so future access is local to UCSC servers.
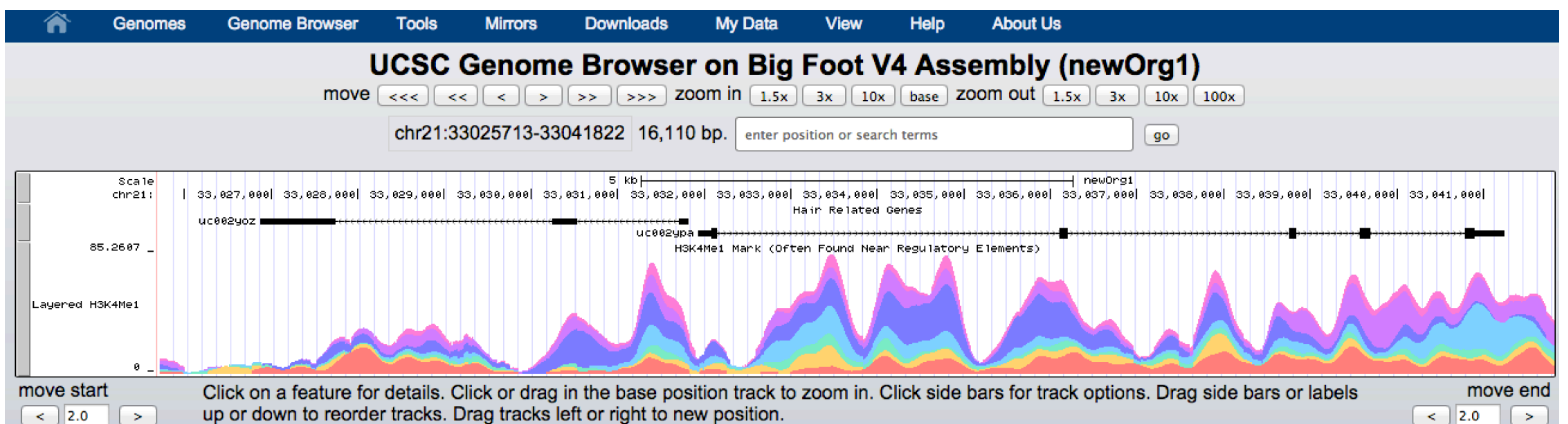
A UCSC Genome Browser session using an imaginary Big Foot Sequence

GBiB: The UCSC Genome Browser running on your laptop with your private data
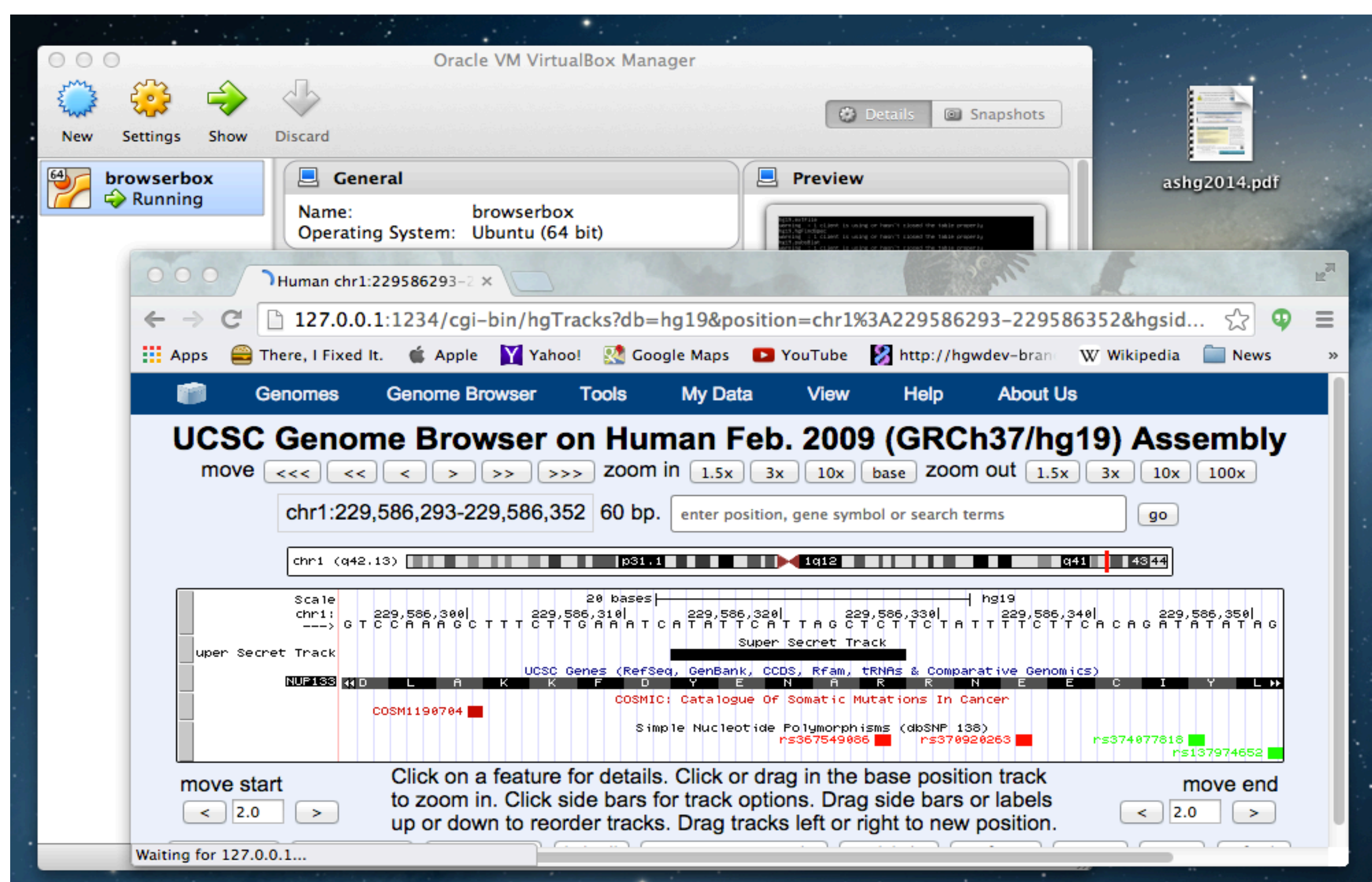
Genome Browser in a Box running on an OSX laptop with private data loaded as a custom track.

Genome Browser in a Box (GBiB) uses virtual machine technology to allow users to easily run the Genome Browser on a protected network, or in situations when annotation cannot be uploaded to UCSC for privacy reasons. To install GBiB one need only download VirtualBox, which is freely available for all major operating systems, and the GBiB virtual image which is a pre-installed copy of the UCSC Genome Browser. The Genome Browser is open source and free for non-commercial users.

UCSC tracks can be mirrored locally for offline use, or the UCSC Genome Browser public mysql server can be used to serve non-local data, or tracks can be supported by local or Internet-accessible track hubs.

See our wiki for more information:

http://genomewiki.cse.ucsc.edu/index.php/G10K2015

Acknowledgments

CENTER for BIOMOLECULAR SCIENCE & ENGINEERING
promoting discovery and invention for human health and well-being