

Use The UCSC Genome Browser To Visualize And Analyze Your Genomic Data

Matthew L. Speir¹, Angie S. Hinrichs¹, Maximilian Haeussler¹, Brian J. Raney¹, Hiram Clawson¹, Kate R. Rosenbloom¹, Galt P. Barber¹, Jonathan Casper¹, Steve Heitner¹, Luvina Guruvadoo¹, Pauline A. Fujita¹, Brian T. Lee¹, Ann S. Zweig¹, Donna Karolchik¹, Robert M. Kuhn¹, David Haussler^{1,2}, W. James Kent¹

¹ University of California Santa Cruz Genomics Institute, Santa Cruz, CA; ²Howard Hughes Medical Institute, UCSC, Santa Cruz, CA

Introduction

The UCSC Genome Browser (<http://genome.ucsc.edu>) is a free, web-based tool that integrates and displays genomic data from a wide variety of sources, including GenBank, ENCODE, UCSC and many others. We provide several tools to help users upload their own data and view it alongside this genomic information or export the data for analysis with other applications. Large genomic data sets and custom genome assemblies can be uploaded and displayed using the browser's data hub tools. If you need to view private data, such as protected patient data from a clinical trial, Genome Browser in a Box (GBiB) allows you to run your own private copy of the Genome Browser on your own computer. The new Data Integrator tool lets you quickly combine input from up to five genome-wide data sets, including your own data uploaded through custom tracks or track hubs, and then export a customized output set based on intersections with a primary track. We are continually working to extend our toolset to allow users to explore their data in new and unique ways.

Use The Data Integrator To Intersect Multiple Data Sets At Once

The Data integrator allows you to quickly and effectively combine data from multiple sources. For instance, you could use it to find the genes, variants and conserved genomic regions that overlap with peaks from a recent RNA-seq experiment. Here's how you can build a query to get that data:

Start by choosing your assembly from one of the hundreds hosted by UCSC, or upload your own with an assembly hub.

Select Genome Assembly and Region

group	genome	assembly
Mammal	Human	Feb. 2009 (GRCh37/hg19)

region to annotate

position or search term: chr21:33,031,597-33,041,570

Next, choose a region to query. This can be a single or set of regions, genome-wide, or a position found by search term.

Configure Data Sources

- RNA-seq Peaks [View table schema](#)
- UCSC Genes [View table schema](#)
- My Patient Variants [View table schema](#)
- Conservation - 100 Vert. EI [View table schema](#)

Add Data Source

track group: Genes and Gene Predictions

track: UCSC Genes (knownGene)

get more data: track hubs, custom tracks

Choose up to four tracks to intersect with a primary track.

Tracks can be ones UCSC provides or ones uploaded using custom tracks or track hubs.

Output Options

Send output to file name: myPeaks.overlap.txt

Compress with gzip (.gz)

Choose fields...

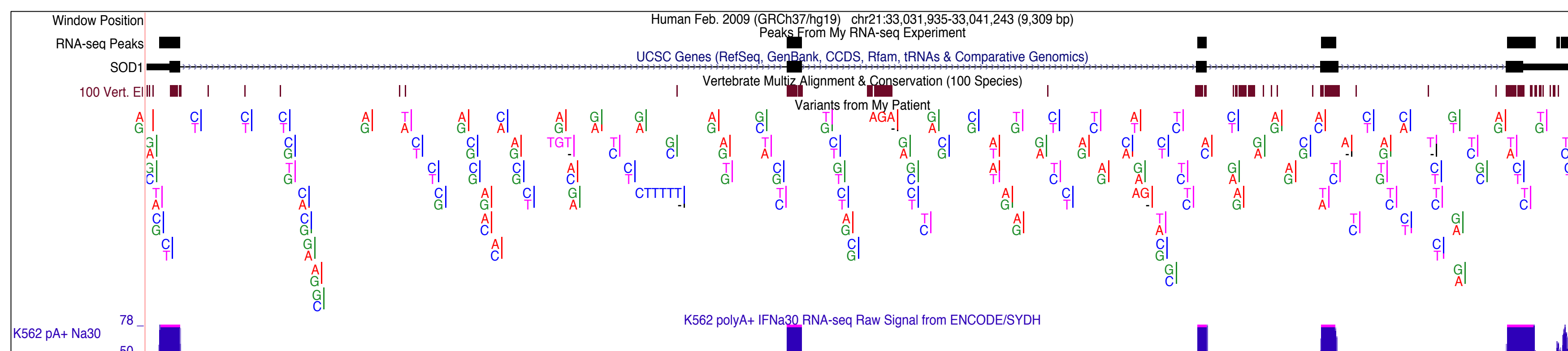
Get output

Lastly, configure your output.

View your output in your web browser or save it to a file, optionally compressed with gzip.

By default, all fields from selected tracks are output. Click "Choose fields" to customize the fields in your output.

ct_RNAseqPeaks_3524.chrom	ct_RNAseqPeaks_3524.chromStart	ct_RNAseqPeaks_3524.chromEnd	ct_knowGene.name	ct_MyPatientVariants_2846.id	ct_MyPatientVariants_2846.ref	ct_MyPatientVariants_2846.alt
phastConsElements100way.name	phastConsElements100way.score					
chr21	33032015	33032152	myPeak.1	uc002ypa.3	rs142752986	T A
chr21	33032015	33032152	myPeak.1	rs200447364	C G	10d=49 379
chr21	33032015	33032152	myPeak.1	rs199766524	C T	10d=144 485
chr21	33032015	33032152	myPeak.1			10d=21 295
chr21	33032015	33032152	myPeak.1			10d=19 285
chr21	33032015	33032152	myPeak.1			10d=37 351
chr21	33032015	33032152	myPeak.1			10d=132 477



The UCSC Genome Browser allows you to view your annotations alongside those from a variety of sources, including ENCODE, dbSNP, ClinVar, Ensembl and many others. Here, we can see a combination of custom tracks, including a VCF file of patient SNPs, alongside UCSC gene annotations and ENCODE RNA-seq data.

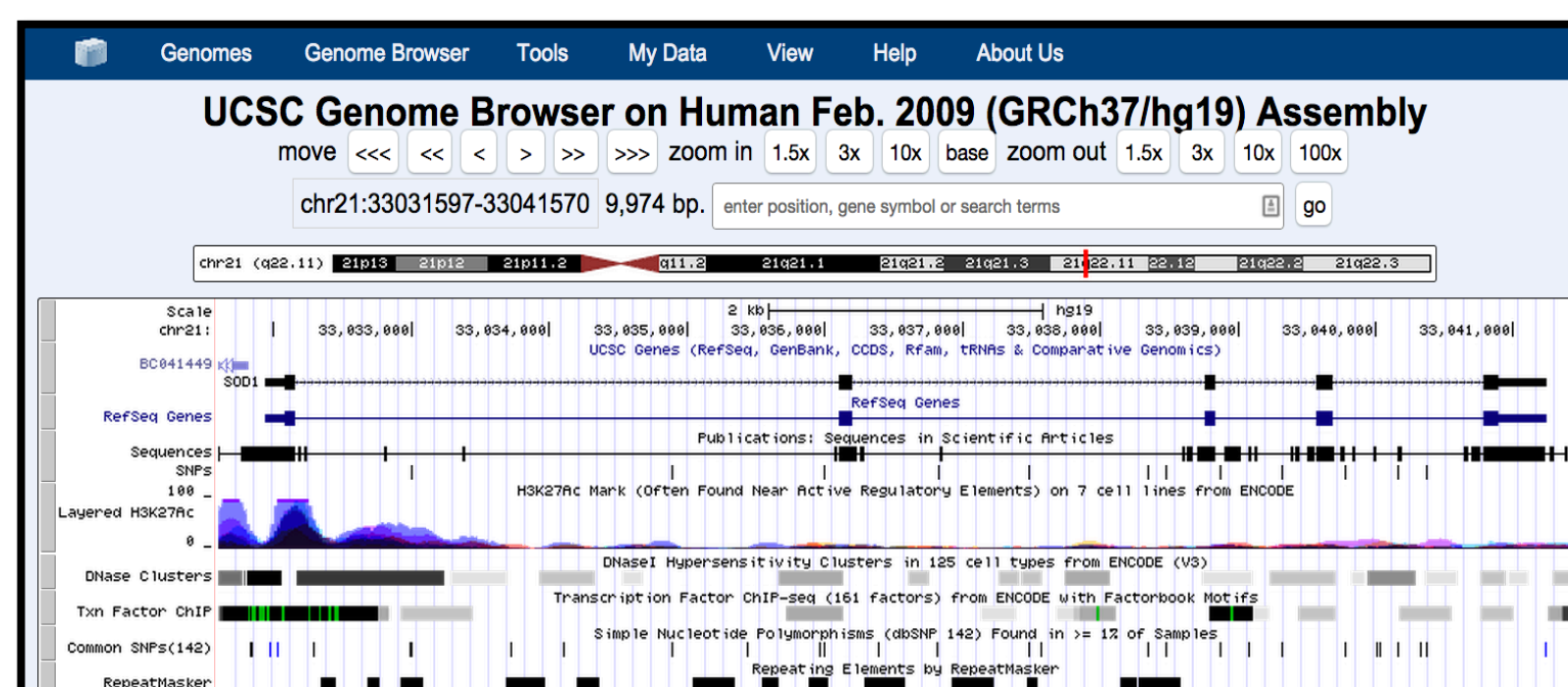
Load Your Data Into The Genome Browser Using Data Hubs

Tracks hubs - package and upload any number of tracks for existing assemblies.

Assembly hubs - upload a custom assembly and related annotations to view in the Genome Browser.

Both support bigBed, bigWig, bigGenePred, BAM, VCF, and HAL formats.

Find more information and examples at: <http://genome.ucsc.edu/goldenPath/help/hgTrackHubHelp.html>



UCSC genome browser track download tool

Group: Mammal | Genome: Human | Assembly: Feb. 2009 (GRCh37/hg19) | Select

Locally mirrored tracks are faster to browse than tracks that are accessed through the internet. Select any number of tracks from the list below and click "Download" when finished. The data will be downloaded from the UCSC servers with rsync and copied to the local mysql database and /data/gbdt.

Total size at UCSC: mysql tables 382.7 GB, gbdt files 6.0 TB
Total size of tables and gbdt on local disk: 0 bytes

Download [show subtracks](#)

Track groups

- Predetermined tracks sets
- Mapping and Sequencing
- Genes and Gene Predictions
- Phenotype and Literature
- mRNA and EST
- Expression
- Regulation
- Comparative Genomics
- Neanderthal Assembly and Analysis
- Drosophila Assembly and Analysis
- Variation
- Repeats

Use The Genome Browser In A Box To View Your Private Data

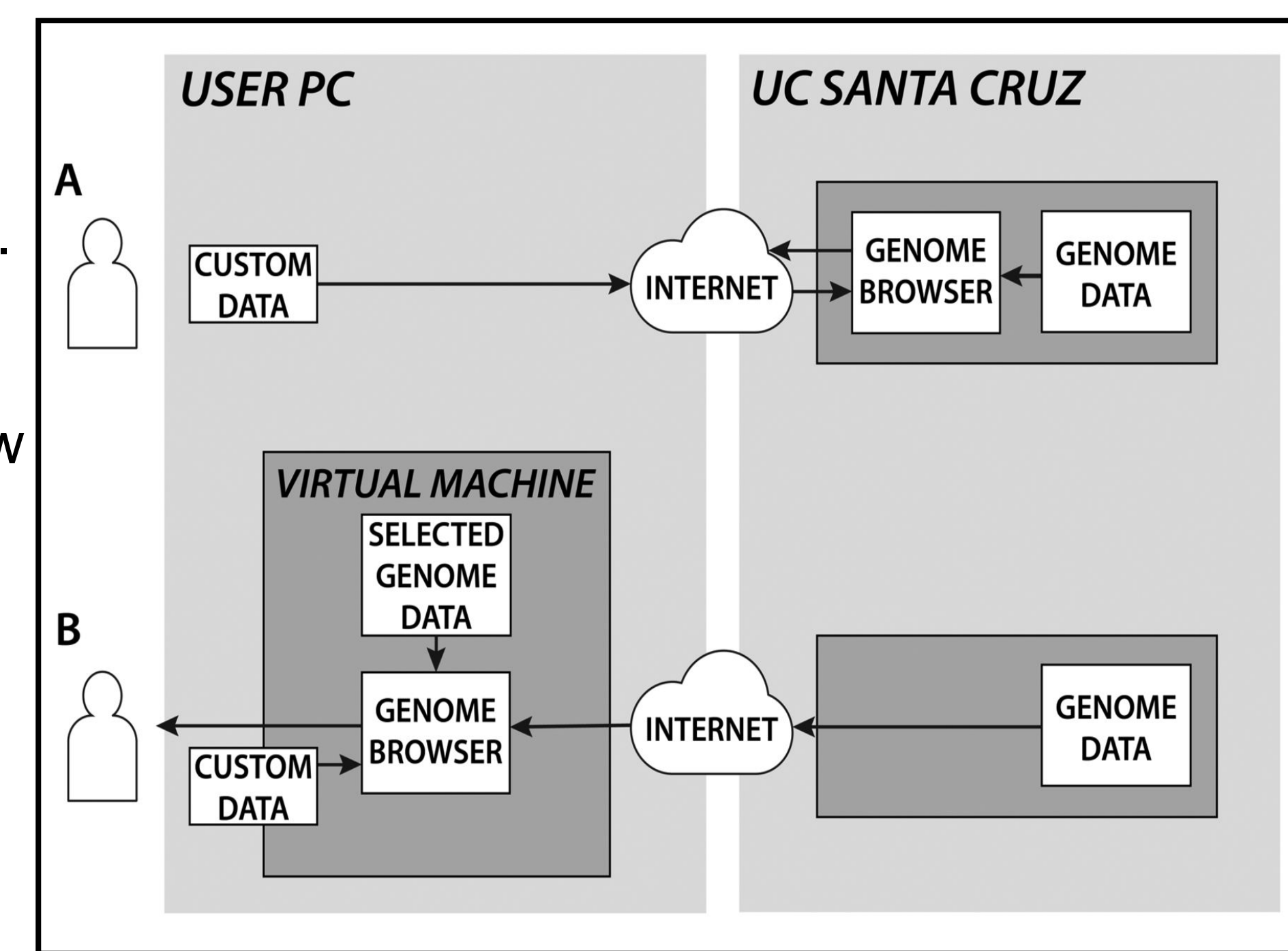


Image Credit: Casey Callow

GBiB is a virtual machine that runs on your computer. It supports custom tracks and data hubs. Any personal data loaded into GBiB never leaves your computer.

The hgMirror tool allows for easy downloading of genomic data for quicker access.

Runs on Windows, Mac, or Linux. Download is 7GB, which expands to 20GB.

Get it today!

- Download GBiB from Genome Browser Store: <https://genome-store.ucsc.edu/>
- Download and install VirtualBox: <https://www.virtualbox.org/>
- Unzip GBiB file and load it into VirtualBox

For more information on using GBiB, see: <http://genome.ucsc.edu/goldenPath/help/gbib.html>

More information:

Send us a message on our public mailing list: genome@soe.ucsc.edu

Find us on:

GenomeBrowser

@GenomeBrowser

UCSC Genome Browser

View training info, videos, and user's guides:

<http://genome.ucsc.edu/training/>

Acknowledgements:

This work is funded by the National Human Genome Research Institute (5 U41 HG002371) to UCSC Center for Genomic Science). We would like to acknowledge the work of the UCSC Genome Bioinformatics technical staff (<http://genome.ucsc.edu/staff.html>), our many collaborators and our users for their feedback and support.

Reference:

The UCSC Genome Browser database: 2015 update. Rosenbloom KR, Armstrong J, Barber GP, Casper J, Clawson H, Diekhans M, Dreszer TR, Fujita PA, Guruvadoo L, Haeussler M, Harte RA, Heitner S, Hickey G, Hinrichs AS, Hubley R, Karolchik D, Learned K, Lee BT, Li CH, Miga KH, Nguyen N, Paten B, Raney BJ, Smit AF, Speir ML, Zweig AS, Haussler D, Kuhn RM, Kent WJ. Nucleic Acids Res. 2015 Jan;43(Database issue):D670-81.

Navigating protected genomics data with UCSC Genome Browser in a Box. Haeussler M, Raney BJ, Hinrichs AS, Clawson H, Zweig AS, Karolchik D, Casper J, Speir ML, Haussler D, Kent WJ. Bioinformatics. 2015 Mar 1;31(5):764-6.

For more information on using the Data Integrator, see: <http://genome.ucsc.edu/goldenPath/help/hgIntegratorHelp.html>

View online by scanning the code or going to <http://genomewiki.ucsc.edu/index.php/BioengSymp2015Poster>

