Visualize and Analyze Your Genomic Data on the UCSC Genome Browser

Luvina Guruvadoo, Angie S. Hinrichs, Maximilian Haeussler, Brian J. Raney, Matthew L. Speir, Jonathan Casper, Ann S. Zweig, Donna Karolchik, Robert M. Kuhn, W. James Kent
University of California Santa Cruz Genomics Institute, Santa Cruz, CA

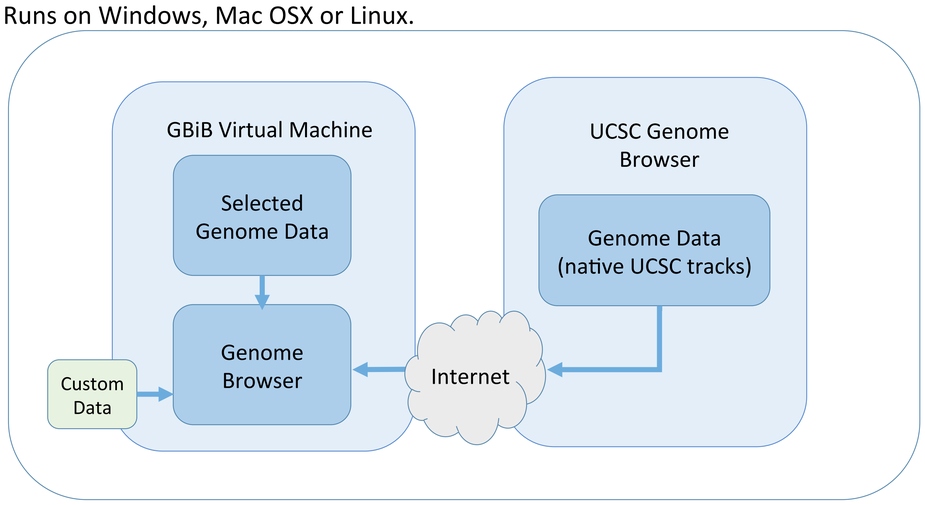
Introduction

Over the last 15 years, the UCSC Genome Browser (http://genome.ucsc.edu) has been one of the leading tools for visualizing and analyzing genomic data. Genomic data from a number of sources, such as GenBank, ENCODE, UCSC and many others, are combined on our free, web-based tool. It allows users to visualize and export this data for analysis with other tools. The Genome Browser in a Box (GBiB) is a complete installation of the Genome Browser in a virtual machine. Without the hassle of configuring a mirror site, the GBiB allows you to view sensitive data, such as protected clinical data, on your own machine without the need to put it on publicly accessible servers. GBiB can be obtained from the Genome Browser store, https://genome-store.ucsc.edu/. We have also developed the Data Integrator tool, which can combine data from the Genome Browser database, custom annotation tracks and track hubs. Data from up to five user-selected tracks, including related data when applicable, can be exported based on intersections with the primary track.

View Private Data on the Genome Browser In A Box (GBiB)

GBiB is a virtual machine of the UCSC Genome Browser that runs on your computer.

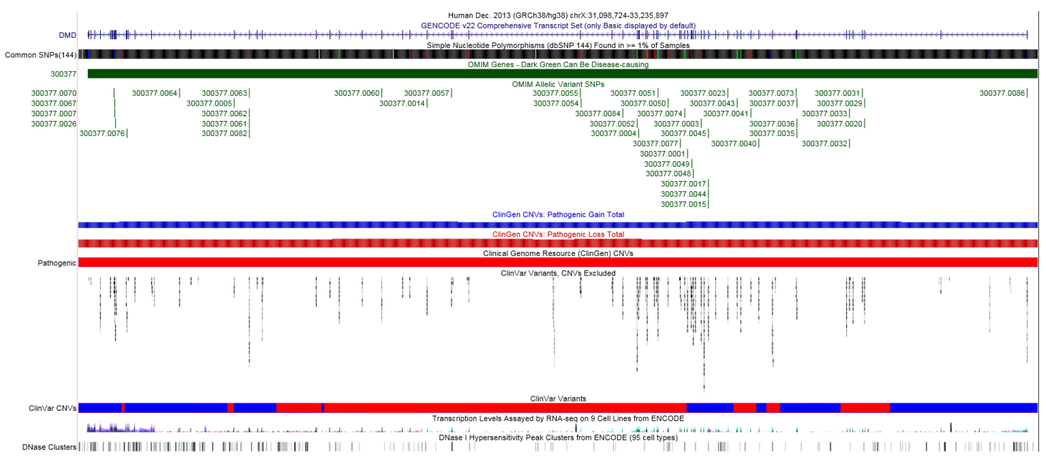
- Use the Genome Browser toolset to view sensitive or protected data.
- View custom tracks and/or data hubs alongside native UCSC tracks.
- Runs on Windows, Mac OSX or Linux.



How to install GBiB:

1. Visit the Genome Browser Store to download GBiB: https://genome-store.ucsc.edu
2. Download and install VirtualBox: https://www.virtualbox.org
3. Unzip the GBiB file and load it into VirtualBox

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



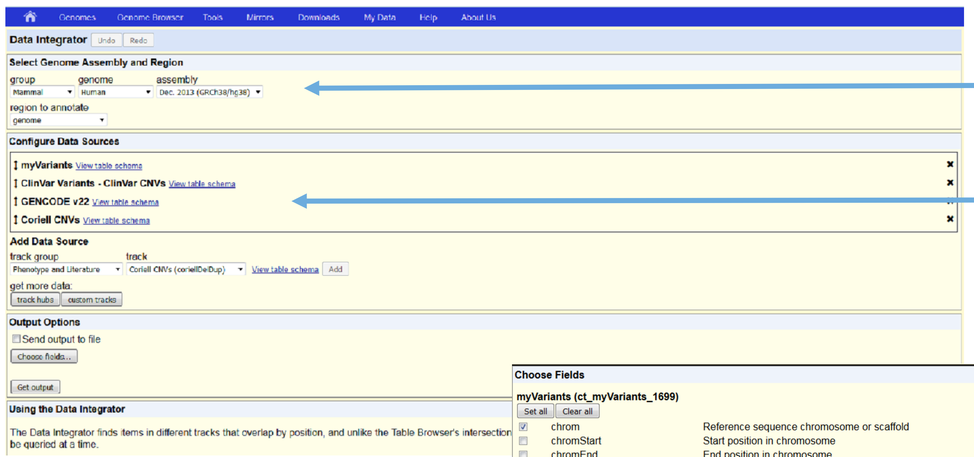
The UCSC Genome Browser allows you to view multiple data sets simultaneously. In the above snapshot, we have opened the GENCODE Genes track (dystrophin gene) on the Human hg38 assembly, alongside data from SNP(144), OMIM, ClinGen, ClinVar, and ENCODE.

Use the Data Integrator to Intersect Your Variants with Multiple Data Sets

The Data Integrator can combine data from the Genome Browser database, custom tracks and track hubs. If you have a custom track of variant data, for example, you can use it to find which genes, ClinVar variants, and other variant data overlap. Here's how to build your query:

Step 1. Create a custom track of your variant data

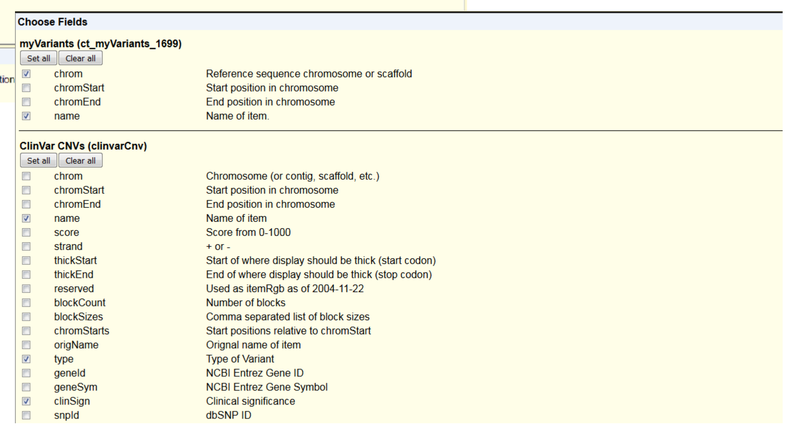
```
Track name=myVariants
chr21           33031821        33031822        rs17883296
chr21           33031926        33031927        rs17878855
chr21           33031973        33031974        rs7277748
chr21           33031992        33031993        rs17881581
chr21           33031995        33031996        rs139202139
chr21           33032034        33032035        rs142752986
chr21           33032037        33032038        rs200447364
chr21           33032096        33032097        rs199766524
chr21           33032286        33032287        rs17881180
```



Step 2. Select a genome assembly and region of interest

Step 3. Configure data sources: Choose up to four tracks to intersect with your custom track

Step 4. Configure your output: Select the fields to include in your output




```
# hgIntegrator: database=hg38 region=genome Thu Feb 25 13:10:25 2016
#ct_myVariants_1699.chrom       ct_myVariants_1699.name          clinvarCnv.name clinvarCnv.type clinvarCnv.clinSign
                knownGene.name  coriellDelDup.name
chr21           rs17883296      21p11.2-q22.3x3 copy number gain Pathogenic       uc061zpb.1      GM08134
chr21           rs17883296      21q11.2-22.3x3  copy number gain Pathogenic                       GM03090
chr21           rs17883296      21q11.2-22.3    copy number gain Pathogenic                       GM13783
chr21           rs17883296      21q11.2-22.3x3  copy number gain Pathogenic                       GM01921
chr21           rs17883296      21q21.3-22.3x3  copy number gain Uncertain significance                      GM04592
chr21           rs17883296      21q11.2-22.3x3  copy number gain Pathogenic
chr21           rs17883296      21q11.2-22.3x3  copy number gain Pathogenic
chr21           rs17883296      21p11.2-q22.3x3 copy number gain conflicting data from submitters
chr21           rs17883296      21q11.2-q22.3x3 copy number gain Pathogenic
chr21           rs17883296      21p11.2-q22.13x1 copy number loss Pathogenic
```

Step 5. Get output: Save your output to a file or view it in the web browser

For variants in VCF or pgSnp format, the **Variant Annotation Integrator** can add functional effect predictions (e.g. coding synonymous, frameshift, splice site, etc).

More information on the Data Integrator:
http://genome.ucsc.edu/goldenPath/help/hgIntegratorHelp.html

In addition to custom tracks, you can upload and intersect Track Hubs. Hubs are a useful tool for visualizing a large number of genome-wide data sets. Read more about Track Hubs here:
http://genome.ucsc.edu/goldenPath/help/hgTrackHubHelp.html

Future Work

Some features we intend to add to the Data Integrator:

- Reorder output columns
- Filter inputs by value
- Configure overlap rules
- Send output to Galaxy
- Paste in identifiers

More Information

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

View Genome Browser tutorials and videos:
http://genome.ucsc.edu/training/index.html

Read our blog:
http://genome.ucsc.edu/blog/

Follow us on:
GenomeBrowser
@GenomeBrowser
UCSC Genome Browser

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.

References

The UCSC Genome Browser database: 2016 update. Speir ML, Zweig AS, Rosenbloom KR, Raney BJ, Paten B, Nejad P, Lee BT, Learned K, Karolchik D, Hinrichs AS, Heitner S, Harte RA, Haeussler M, Guruvadoo L, Fujita PA4, Eisenhart C, Diekhans M, Clawson H, Casper J, Barber GP, Haussler D5, Kuhn RM, Kent WJ. Nucleic Acids Res. 2016 Jan 4;44(D1):D717-25. Epub 2015 Nov 20.

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.

ACMG Tampa, FL March 8-10, 2016