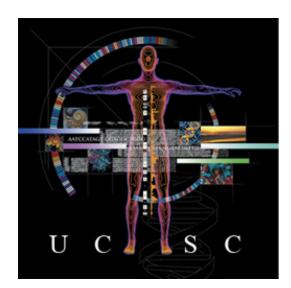


# Visualize and Analyze Your Genomic Data on the UCSC Genome Browser



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.

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

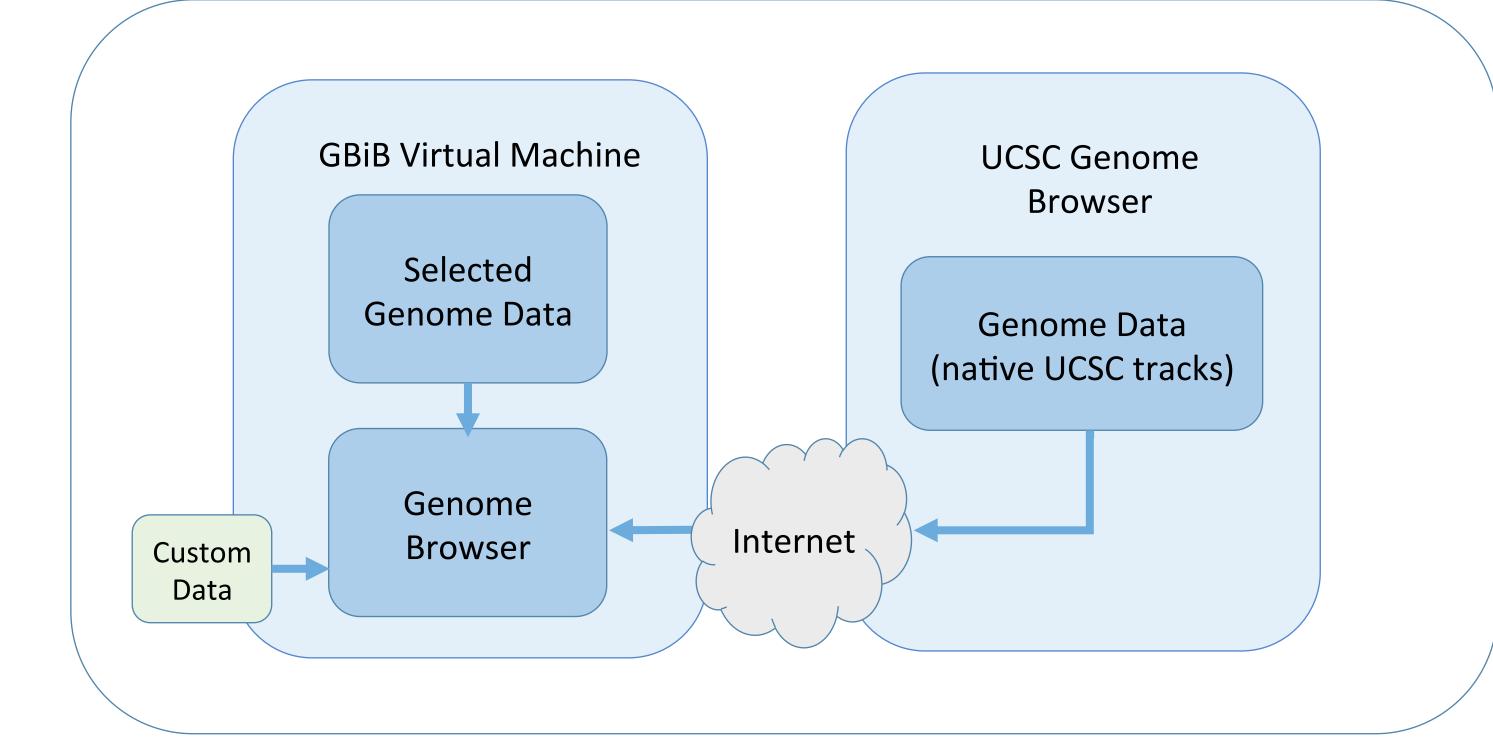
The Data Integrator can combine data from the Genome Browser database, custom tracks and track hubs. If you have a custom track of variant data, for example, you can use it to find which genes, ClinVar variants, and other variant data overlap. Here's how to build your query:

	Track name=	=myVariants		
	chr21	33031821	33031822	rs17883296
	chr21	33031926	33031927	rs17878855
	chr21	33031973	33031974	rs7277748
Step 1. Create a custom track	chr21	33031992	33031993	rs17881581
	chr21	33031995	33031996	rs139202139
of your variant data	chr21	33032034	33032035	rs142752986

# 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.



or your variant data	chr21	33032037	33032038	rs200447364	
	chr21	33032096	33032097	rs199766524	
	chr21	33032286	33032287	rs17881180	

🏫 Genomes Genome Browser Tools Mirrors Downloads My Data Help About Us			
Data Integrator Undo Redo			
Select Genome Assembly and Region			
group genome assembly Mammal  V Human  V Dec. 2013 (GRCh38/hg38)  V			Step 2. Select a genome
region to annotate			
genome T			assembly and region of
Configure Data Sources			interest
1 myVariants <u>View table schema</u> 1 ClinVar Variants - ClinVar CNVs <u>View table schema</u>			
1 GENCODE v22 View table schema			Stop 2 Configuro data courca
1 Coriell CNVs View table schema			Step 3. Configure data source
Add Data Source			Choose up to four tracks to
track group     track       Phenotype and Literature <ul> <li>Coriell CNVs (coriellDelDup)</li> <li>View table schema</li> <li>Add</li> </ul>			-
get more data:			intersect with your custom
track hubs custom tracks			track
Output Options			
Send output to file			
Choose fields	Chasses Fields		
Get output	Choose Fields		
Using the Data Integrator	myVariants (ct_myVariants_1699) Set all Clear all		
The Data Integrator finds items in different tracks that overlap by position, and unlike the Table Browser's intersection	Chrom	Reference sequence chromosome or scaffold	
be queried at a time.	<ul> <li>chromStart</li> <li>chromEnd</li> </ul>	Start position in chromosome End position in chromosome	
	✓ name	Name of item.	
	ClinVar CNVs (clinvarCnv)		
	Set all Clear all		
	chrom chromStart	Chromosome (or contig, scaffold, etc.) Start position in chromosome	
Step 4. Configure your output:	chromEnd	End position in chromosome	
Select the fields to include in	<ul><li>name</li><li>score</li></ul>	Name of item Score from 0-1000	
Select the neids to include in	strand	+ or -	
your output	<ul> <li>thickStart</li> <li>thickEnd</li> </ul>	Start of where display should be thick (start codon) End of where display should be thick (stop codon)	
	reserved	Used as itemRgb as of 2004-11-22	
	<ul> <li>blockCount</li> <li>blockSizes</li> </ul>	Number of blocks Comma separated list of block sizes	
	chromStarts	Start positions relative to chromStart	
	origName	Orignal name of item	
	<ul><li>✓ type</li><li>□ geneld</li></ul>	Type of Variant NCBI Entrez Gene ID	
	geneSym	NCBI Entrez Gene Symbol	
	<ul> <li>clinSign</li> <li>snpld</li> </ul>	Clinical significance dbSNP ID	

#### How to install GBiB:

1. Visit the Genome Browser Store to download GBiB: <u>https://genome-store.ucsc.edu</u>

- 2. Download and install VirtualBox: <a href="https://www.virtualbox.org">https://www.virtualbox.org</a>
- 3. Unzip the GBiB file and load it into VirtualBox

## For more information on GBiB, see: <a href="http://genome.ucsc.edu/goldenPath/help/gbib.html">http://genome.ucsc.edu/goldenPath/help/gbib.html</a>

DMD Common SNPs(144) 300377	<u>₩₩₩&lt;₩₩₩₩₩₩₩₩₩₩₩₩₩₩₩₩₩₩₩₩₩₩₩₩₩₩₩₩</u>	Human Dec. 2013 (GRCh38/hg38) chrX:31,098,724-33,235,897 GENCODE v22 Comprehensive Transcript Set (only Basic displayed by default) 	<del></del>		
300377.0070 300377.0067 300377.0007 300377.0026	OMIM Allelic Variant SNPs           300377.0064         300377.0063         300377.0060         300377.0057         300377.0051         300377.0023         300377.0073         300377.0031         300377.008           300377.0055         300377.0055         300377.0055         300377.0051         300377.0073         300377.0031         300377.008           300377.0054         300377.0054         300377.0054         300377.0043         300377.0037         300377.0029           300377.0084         300377.0074         300377.0041         300377.0033				
		ClinGen CNVs: Pathogenic Loss Total			
Pathogenic		Clinical Genome Resource (ClinGen) CNVs			
		ClinVar Variants, CNVs Excluded			

#CC_myvarian	nts_1699.chrom knownGene.name	<pre>ct_myVariants_16 coriellDelDup.na</pre>		CIINVarChv.nan	ne clinvarCnv.type	CIINVarChv.CIINS	rgu
chr21	rs17883296			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 sigr	nificance		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 da	ata from submitters		
chr21	rs17883296	21p11.2-q22.3x3	copy number	gain Pathogenic			
chr21	rs17883296	21p11.2-q22.13x1	Lcopy 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

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

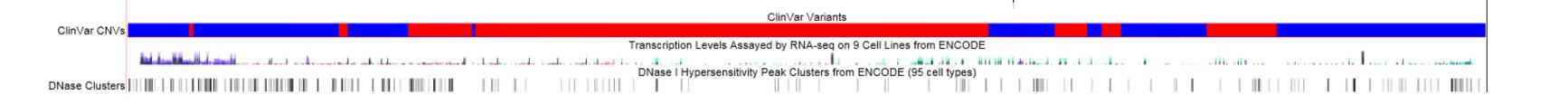
### Future Work

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

#### 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 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.

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

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.

