

# View your data with ease and privacy in the UCSC Genome Browser

Galt P. Barber<sup>1</sup>, Brian J. Raney<sup>1</sup>, Maximilian Haeussler<sup>1</sup>, Angie S. Hinrichs<sup>1</sup>, Ann S. Zweig<sup>1</sup>, Donna Karolchik<sup>1</sup>, Hiram Clawson<sup>1</sup>, Jonathan Casper<sup>1</sup>, Matthew L. Speir<sup>1</sup>, Brian T. Lee<sup>1</sup>, Timothy R. Dreszler<sup>1</sup>, Pauline A. Fujita<sup>1</sup>, Robert M. Kuhn<sup>1</sup>, David Haussler<sup>1,2</sup>, W. James Kent<sup>1</sup>

<sup>1</sup> University of California Santa Cruz, Center for Biomolecular Science and Engineering, School of Engineering, Santa Cruz, CA, USA

<sup>2</sup> Howard Hughes Medical Institute, University of California Santa Cruz, Center for Biomolecular Science and Engineering, School of Engineering, Santa Cruz, CA, USA

Several recently added features and improvements make viewing your own data on the UCSC Genome Browser easier and safer than ever before.

## Hubs

**Track Hubs** allow you to quickly and easily create custom annotations with sophisticated display settings. Several data formats are currently supported: BigBed, BigWig, BAM, HAL, and VCF/tabix. Users host track data on their own servers. Data formats are compressed and indexed. Data is fetched using random-access via HTTP, HTTPS, or FTP, and then cached locally. BigBed and BigWig have precomputed zoomed-views for fast display of large chromosome ranges.

Example of a Track Hub hosted at <http://yoursite.edu/~you/hub.txt>:

```
hub.txt
hub:RoadmapIntegrative
description: http://www.roadmapgenomics.org
shortLabel: Roadmap Epigenomics Integrative Analysis Hub
longLabel: Roadmap Epigenomics Integrative Analysis Hub at Washington University in St. Louis
genome: hg19
email: hwang@genetics.wustl.edu

genomes.txt
genome: hg19
trackDb: hg19TrackDb.txt

trackDb.txt
track: RoadmapIntegrativeAssay
superTrack: on: show
group: RoadmapIntegrative
shortLabel: By Assay
longLabel: Roadmap Integrative Analysis Hub organized by assay
priority: 1

track: RoadmapIntegrativeAssayAssay27004
compartment: chrMM
parent: RoadmapIntegrativeAssay
longLabel: chrMM tracks from Roadmap
shortLabel: view: Default, View: Default, View: Imputed, segmentation=imputed, segmentation: Auxiliary, segmentation: Auxiliary, segmentation: subGeno2
sampleType: Sample, Type: sample13006+CD14, primary, cells: sample1216+H9, DF, & I, sample13004+Rectal_Smooth, Muscle: sample13007+breast_vMHC
sample13003+Muscle, Sample: Cell: sample1212+H9, DF, I, sample13003+Breast_Muscle: sample13002+Stomach, Smooth: Muscle:
sample13014+Periosteal_Blood_Mononuclear, Primary: cells: sample13015+CD24, Primary: cells: sample13013+Adipose_Nuclei: sample13010+Adip_Liver
sample13016+Bone_Marrow_Deived_Mesenchymal_Stem: cell: sample13014+Colon, Muscle: sample13015+Renal, Muscle: sample13016+Skin_Memory_Primary_Cells
sample13012+CD4_CDS291_CD127+, Tumor: Primary: Cells: sample13015+CD4_CDS291_IL17+, PMA-Induced, stimulated, TH17, Primary: Cells: sample13016+CD4_CDS291_CD127+, Memory_Primary_Cells: sample13015+CD4_CDS291_7%, Primary: Cells: sample13016+Brain_Germinal_Maturation: sample13016+CD4_CDS291_CD127+, Treg_Primary_Cells
sample1222+Falx_Astriae_Gland: sample1223+Falx_Innervated_Large: sample1101+H1, sample1106+HUES48: sample1107+HUES54: sample1104+H9: sample1222+Falx_Muscle_Tumor:
sample1131+H1_BMP4_Deived_Mesoderm_Cultured_Cells: sample1130+H1_BMP4_Deived_Mesoderm_Cultured_Cells: sample1131+H1_BMP4_Deived_Mesoderm_Cultured_Cells
sample1130+Chondrocytes_Spin_Spine_Marrow_Deived_Mesoderm_Cultured_Cells: sample1131+H1_BMP4_Deived_Mesoderm_Cultured_Cells

track: Group3 data: Type: Data_Type: Real: Read: Imputed: Imputed
dimensions: dim: chrMM: view: dimension? = sample: Type: dim: data: Type:
sortOrder: sample: Type: view:
dim: chrMM: sample: Type: view:
display: chrMM: on
visibility: dense
type: bed3

track: RoadmapIntegrativeAssayAssay27004ViewDefault_View
shortLabel: Default_View
view: Default_View
mapHeight: 64:32:16
parent: RoadmapIntegrativeAssayAssay27004
visibility: dense

track: ED17_15_coxMeths_dense
shortLabel: LNSG MRRO
longLabel: ED17 MRRO chromatin state
type: bigBed 9
hgtrack: on
bigBedUrl: http://hgdownload.cse.ucsc.edu/hg19/ED17_15_coxMeths_dense.bigBed
visibility: dense
subGeno2: view: Default_View: sample: Type: sample13010: assay: Type: assay27004: data: Type: Real
roadmap: on
mapHeight: 64:32:16
metadata: "GROUP="MRRO" download_url="http://www.broadinstitute.org/~anahj/projects/roadmap/segmentations/models/coxMeths/parallel/set2final/ED17_15_coxMeths_dense.bed.gz" "NEW
ED17_15" Standardized Epigenome Name="MRRO test lung fibroblasts Cell Line="Epigenome Memoric="LNSG MRRO" TYPE="Cell Line"
parent: RoadmapIntegrativeAssayAssay27004ViewDefault_View

track: ED35_15_coxMeths_dense
shortLabel: BLD CD34 PC
longLabel: ED35 CD34 PC Primary cells chromatin state
type: bigBed 9
hgtrack: on
bigBedUrl: http://hgdownload.cse.ucsc.edu/hg19/ED35_15_coxMeths_dense.bigBed
visibility: dense
subGeno2: view: Default_View: sample: Type: sample13019: assay: Type: assay27004: data: Type: Real
roadmap: on
mapHeight: 64:32:16
metadata: "GROUP="TSC & B cell" download_url="http://www.broadinstitute.org/~anahj/projects/roadmap/segmentations/models/coxMeths/parallel/set2final/ED35_15_coxMeths_dense.bed.gz" "NEW
ED17_15" Standardized Epigenome Name="Primary human gingival stem cells" Epigenome Memoric="BLD CD34 PC" TYPE="Primary Cell"
parent: RoadmapIntegrativeAssayAssay27004ViewDefault_View

Assembly Hubs are a new extension of the hub system that allows you to create and display your own custom genome assemblies and annotation tracks. Users simply provide the genomes in .2bit format hosted on their own servers. The genomes.txt file for an assembly hub has several additional settings for each genome.

genomes.txt
genome: EscherichiaColi042Jid161985
webDirPath: EscherichiaColi042Jid161985/EscherichiaColi042Jid161985.2bit
trackDb: EscherichiaColi042Jid161985/trackDb.txt
groups: groups.txt
formatPath: EscherichiaColi042Jid161985/description.html
organism: 042
orderKey: 4803
scientificName: EscherichiaColi042Jid161985
defaultPos: NC0176261.1:1-1000
[...] (repeat for each genome)

The assembly hub genomes appear on the gateway page for easy selection. Compiled utilities for creating these formats are available here:
http://hgdownload.cse.ucsc.edu/admin/exe/
They may be compiled from source code as well.

Basic Hub Quick Start Guide:
http://genome.ucsc.edu/goldenPath/help/hubQuickStart.html
```

**Assembly Hubs** are a new extension of the hub system that allows you to create and display your own custom genome assemblies and annotation tracks. Users simply provide the genomes in .2bit format hosted on their own servers. The genomes.txt file for an assembly hub has several additional settings for each genome.

```
genomes.txt
genome: EscherichiaColi042Jid161985
webDirPath: EscherichiaColi042Jid161985/EscherichiaColi042Jid161985.2bit
trackDb: EscherichiaColi042Jid161985/trackDb.txt
groups: groups.txt
formatPath: EscherichiaColi042Jid161985/description.html
organism: 042
orderKey: 4803
scientificName: EscherichiaColi042Jid161985
defaultPos: NC0176261.1:1-1000
[...] (repeat for each genome)
```

The assembly hub genomes appear on the gateway page for easy selection. Compiled utilities for creating these formats are available here: <http://hgdownload.cse.ucsc.edu/admin/exe/>

They may be compiled from source code as well. Basic Hub Quick Start Guide: <http://genome.ucsc.edu/goldenPath/help/hubQuickStart.html>

## Security

Hubs can be publicly shared, or restricted to private use. For increased data privacy, you can use **HTTPS** with login for all your URLs. We have also added a number of recent improvements to the Genome Browser database and software to enhance the security of the site and user data.

Cryptographically-secure web IDs are used throughout the site. Database configuration minimizes privileges and prevents access to passwords and sensitive databases. New functions construct queries securely preventing SQL-injection.

## Mirrors

For even greater security and customization you can create a mirror site of the UCSC Genome Browser behind an organizational firewall or other secure environment.

Keep sensitive data on your own machine without needing to put it on publicly accessible servers. Load any data or tracks that you like, creating your own custom Genome Browser.

<https://genome.ucsc.edu/admin/mirror.html>

This approach has worked well for many individuals and institutions, but requires significantly more time and effort than hubs to install and configure the software and dependencies.

## GBiB

UCSC has recently released **Genome Browser in a Box (GBiB)**, a complete installation of the Genome Browser in a virtual machine. This takes the hassle out of configuring your own mirror, and also lets you keep sensitive data on your own machine without needing to put it on publicly accessible servers. GBiB allows you to easily create a mirror on your own laptop or computer by simply downloading Virtual Box, which is free and supported on many operating systems, and then downloading and running a GBiB image for it from UCSC servers.

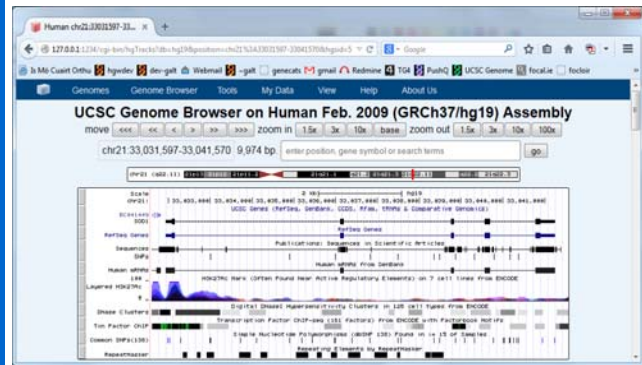
Information for downloading and installing GBiB: <http://genome.ucsc.edu/goldenPath/help/gbib.html>

VirtualBox downloads: [www.virtualbox.org](http://www.virtualbox.org)

Our new license registration page makes it easy for academic and non-commercial users to obtain a free GBiB license and for other users to purchase a commercial license: <https://genome-store.ucsc.edu>

The gbib.zip download is about 7GB. Unzip and double-click browserbox.vbox to launch the virtual machine. Open a web browser and enter this URL: <http://127.0.0.1:1234>

It is that easy!



Requires 20 GB of free space on your hard disk. Optimized for hg19 by default. Other assemblies and tracks may be mirrored as well for speed at the cost of additional space and download bandwidth.

- Go to Tools/Mirror Tracks on the main menu and choose the desired tracks to mirror.
- Can share directories with the host machine, so that large files may be loaded without uploading.
- Can only be accessed from the host machine by default, protecting your data.
- Can also be configured to be publicly accessible if desired.
- Updates automatically.
- Contains the handy UCSC utilities.

This work was funded by National Human Genome Research Institute award 5U4HG002371 to UCSC Center for Genomic Science and 1U41HG006992 subcontract 60141508-106846-A to UCSC ENCODE Data Coordination Center. We would like to acknowledge our many collaborators, data providers, and users for their feedback and support. The UCSC Genome Browser source code and software is free for non-commercial use. Corporate use requires a license.

Contact us at [genome@soe.ucsc.edu](mailto:genome@soe.ucsc.edu) and follow us on Twitter @GenomeBrowser.