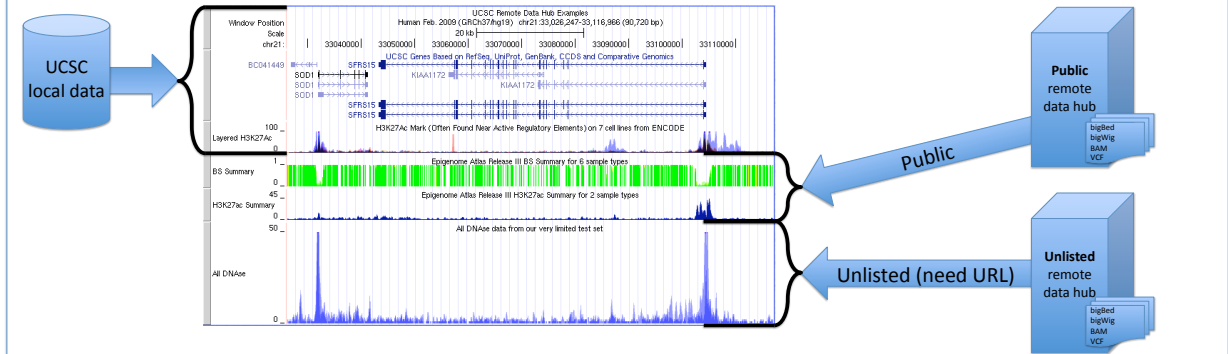


Remote Data Track Storage for Viewing on the UCSC Genome Browser

Brian J. Raney¹, Gail Barber¹, Angie Hintsche¹, Mary Goldman¹, Greg Roe¹, Brooke Rhead¹, Luvinia Gurusvaidoo¹, Pauline Fujita¹, Venkat S. Malladi¹, Ann Zweig¹, Donna Karolichik¹, David Haussler^{1,2}, W. James Kent¹
¹Center for Biomolecular Science and Engineering, University of California at Santa Cruz, California, USA
²Howard Hughes Medical Institute, University of California at Santa Cruz, California, USA

Your Remote Data Displayed In the Context of Other Annotations



Track Data Hubs are collections of annotation tracks that can be stored anywhere on the Internet, and can be accessed by the UCSC Genome Browser as if they were stored locally. The file types supported in Track Data Hubs are BAM, VCF/tabix, bigBed, and bigWig, all of which are random access. Only the data being viewed is uploaded to the UCSC Genome Browser, which caches the data locally; therefore, the upload need only occur once when repeatedly viewing the same area. Track Data Hubs are supported by the Genome Browser graphical interface, as well as by the UCSC Table Browser, which allows for intersections between tables, filtering on table fields, and collating data from several different tables.

UCSC lists public hubs, but private hubs can be added too

Display	Hub Name	Description	Assemblies	URL
<input checked="" type="checkbox"/>	Roadmap Epigenomics Data Complete Collection at Wash U VizHub	Roadmap Epigenomics Data Complete Collection at Wash U VizHub	hg19	http://vizhub.wustl.edu/VizHub/RoadmapReleaseAll.txt
<input type="checkbox"/>	UMassMed ZHub	UMassMed H3K4me3 ChIP-seq data for autistic brains	hg19	http://zlab.umassmed.edu/zlab/publications/UMassMedZHub/hub.txt
<input type="checkbox"/>	Cancer genome polyA site & usage	An in-depth map of polyadenylation sites in cancer (matched-pair tissues and cell lines)	hg19	http://johmlab.org/xpad/Hub/UCSC.txt

Load Selected Hubs | Contact genome@soe.ucsc.edu to add a public hub.

Hub track controls appear with other tracks

The screenshot shows a detailed view of track controls. It includes buttons for 'hide', 'show', 'pack', and 'unpack' for various tracks. The tracks shown include UCSD Histone, UCSC Histone, DNase, Footprinting, RNA, DNA Methylation, Mapping and Sequencing Tracks, Phenotype and Disease Associations, and Genes and Gene Prediction Tracks.

Formatting your Data

Data files must be in one of the following formats: bigBed, bigWig, BAM or VCF indexed by tabix. Database information is stored in text files with the following structure:

- myHub/ - directory for the hub as a whole
- hub.txt - text file containing a short description of the hub
- genomes.txt - text file containing a list of genome assemblies
- hg19/ - directory for a recent human assembly
- trackDb.txt - text file containing track display details including names, colors, data types, etc.
- dnase.html - (example) HTML file describing a DNase track to users
- dnaseSignal.bw - (example) wiggle plot of DNase Signal
- dnaseReads.bam - (example) BAM file of DNase Reads

Coming Soon: Assembly Data Hubs

To address the increasing need for researchers to annotate sequence for which UCSC does not provide an annotation database, we will be introducing Assembly Data hubs. These will allow researchers to include the underlying reference sequence, as well as data tracks that annotate that sequence. Sequence will be stored in the UCSC twoBit format, and the annotation tracks will be stored in the same manner as Track Data Hubs.

Acknowledgements

This work was funded by National Human Genome Research Institute award 5P41 HG002371-12 to UCSC Center for Genomic Science.

See our wiki for more information:
<http://genomewiki.cse.ucsc.edu/index.php/GI2012TrackHubsPoster>

