

UCSC Data Hubs

Track Hubs and Assembly Hubs

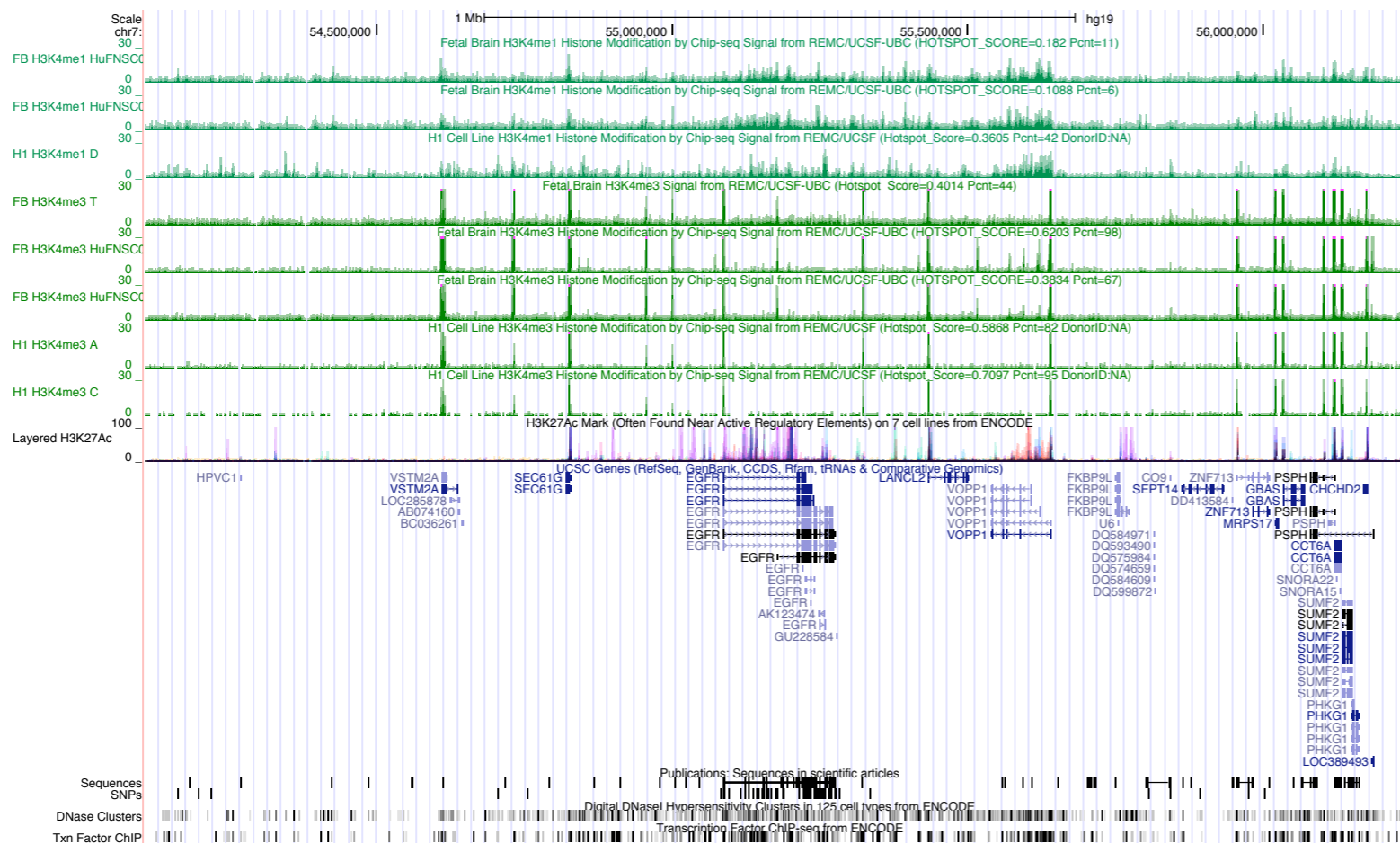


Track Hubs



- Feature developed in August 2011 with Wash U. for display of Epigenome Roadmap data
- Allows large amounts of genomic data to be stored remotely, but treated by the browser as if they were stored locally.
- Track Hubs eliminate the need to transfer large amounts of data over the internet to the browser.

How Track Hubs work



Public remote data hub

info.txt
bigBed
bigWig
BAM
VCF/tabix

UCSC local data

Data Hub Tech

- Requires random access file types (BAM, VCF/tabix, bigBed, and bigWig), so that only the region of the data currently being viewed in the browser can be quickly accessed and uploaded.
- Relies on URL Data Cache for text files and the large data files
- Browser fetches data from up to 100 hub data tracks in parallel
 - Will fetch data from ~10 tracks per second from Wash U in St. Louis to UCSC
 - For popular data, caching makes performance as good as local (~500 tracks/second)

How *you* can use Track Hubs.

genome.ucsc.edu > Genome Browser

The screenshot shows the UCSC Genome Browser Gateway interface. At the top is a blue navigation bar with links: Home, Genomes, Blat, Tables, Gene Sorter, PCR, Session, FAQ, and Help. Below this is a light blue header for the "Human (*Homo sapiens*) Genome Browser Gateway". A paragraph of text states: "The UCSC Genome Browser was created by the [Genome Bioinformatics Group of UC Santa Cruz](#). Software Copyright (c) The Regents of the University of California. All rights reserved." Below the text is a search form with five fields: "clade" (Mammal), "genome" (Human), "assembly" (Feb. 2009 (GRCh37/hg19)), "position or search term" (chr21:33,031,325-33,041,570), and "gene" (empty). A "submit" button is to the right of the "gene" field. Below the search form is a link: "Click here to reset the browser user interface settings to their defaults." At the bottom of the form are five buttons: "track search", "add custom tracks", "track hubs" (highlighted with a red circle), "configure tracks and display", and "clear position".

Track Data Hubs

Track data hubs are collections of tracks from outside of UCSC that can be imported into the Genome Browser. To import a public hub check the box in the list below. After import the hub will show up as a group of tracks with its own blue bar and label underneath the main browser graphic, and in the configure page. For more information, see the [User's Guide](#).

NOTE: Because Track Hubs are created and maintained by external sources, UCSC is not responsible for their content.

Public Hubs		My Hubs		
Display	Hub Name	Description	Assemblies	URL
<input type="checkbox"/>	SDSU NAT	Sense/antisense gene/exon expression using Affymetrix exon array from South Dakota State University, USA	rn4,mm9,hg19	http://bioinformatics.sdstate.edu/datasets/2012-NAT/hub.txt
<input type="checkbox"/>	DNA Methylation	DNA Methylation	rheMac3,mm9,hg18,hg19	http://smithlab.usc.edu/trackdata/methylation/hub.txt
<input type="checkbox"/>	Translation Initiation Sites (TIS)	Translation Initiation Sites (TIS) track	hg19	http://gengastro.1med.uni-kiel.de/suppl/footprint/Hub/tisHub.txt
<input type="checkbox"/>	ENCODE Analysis Hub	ENCODE Integrative Analysis Data Hub	hg19	http://ftp.ebi.ac.uk/pub/databases/ensembl/encode/integration_data_jan2011/hub.txt
<input type="checkbox"/>	miRcode microRNA sites	Predicted microRNA target sites in GENCODE transcripts	hg19	http://www.mircode.org/ucscHub/hub.txt
<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

Currently have 8 public hubs, Roadmap being the first!

(contact genome@soe.ucsc.edu to add a public hub)

Track Data Hubs

Track data hubs are collections of tracks from outside of UCSC that can be imported into the Genome Browser. To import a public hub check the box in the list below. After import the hub will show up as a group of tracks with its own blue bar and label underneath the main browser graphic, and in the configure page. For more information, see the [User's Guide](#).

NOTE: Because Track Hubs are created and maintained by external sources, UCSC is not responsible for their content.

Public Hubs		My Hubs		
Display	Hub Name	Description	Assemblies	URL
<input type="checkbox"/>	SDSU NAT	Sense/antisense gene/exon expression using Affymetrix exon array from South Dakota State University, USA	rn4,mm9,hg19	http://bioinformatics.sdstate.edu/datasets/2012-NAT/hub.txt
<input type="checkbox"/>	DNA Methylation	DNA Methylation	rheMac3,mm9,hg18,hg19	http://smithlab.usc.edu/trackdata/methylation/hub.txt

Public Hubs		My Hubs		
URL:	<input type="text"/>	<input type="button" value="Add Hub"/>		
No Track Hubs for this genome assembly				

Currently ~4000 “My Hubs” type hubs as well (~500 hosts)

Track Hubs are displayed beneath the genome browser along with all the other native tracks.

The screenshot displays the 'Roadmap Epigenomics Data Complete Collection at Wash U VizHub' interface. At the top, there is a blue header bar with a minus sign icon on the left, the title 'Roadmap Epigenomics Data Complete Collection at Wash U VizHub', and a 'refresh' button on the right. Below the header, the interface is organized into a grid of track categories. Each category is represented by a blue link and a corresponding control button (e.g., 'hide', 'full', 'show'). The categories include: Broad Histone, UCSD Histone, UCSF-UBC-USC Histone, DNase, Footprinting, RNA, DNA Methylation, By Assay..., By Sample..., Assay Summary..., Sample Summary..., Methylation Summary..., Roadmap ChromHMM, Roadmap ChromHMM Core Marks, and Roadmap Uniformly Signal. Below this grid, there is a vertical list of track hubs, each with a plus sign icon on the left, a central title, and a 'refresh' button on the right. The track hubs are: Mapping and Sequencing Tracks, Phenotype and Disease Associations, Genes and Gene Prediction Tracks, Literature, mRNA and EST Tracks, Expression, Regulation, Comparative Genomics, Neandertal Assembly and Analysis, Denisova Assembly and Analysis, and Variation and Repeats.

Track Category	Control
Broad Histone	hide
UCSD Histone	hide
UCSF-UBC-USC Histone	full
DNase	hide
Footprinting	hide
RNA	hide
DNA Methylation	full
By Assay...	hide
By Sample...	show
Assay Summary...	hide
Sample Summary...	hide
Methylation Summary...	hide
Roadmap ChromHMM	hide
Roadmap ChromHMM Core Marks	hide
Roadmap Uniformly Signal	hide

Track Hub	Control
Mapping and Sequencing Tracks	refresh
Phenotype and Disease Associations	refresh
Genes and Gene Prediction Tracks	refresh
Literature	refresh
mRNA and EST Tracks	refresh
Expression	refresh
Regulation	refresh
Comparative Genomics	refresh
Neandertal Assembly and Analysis	refresh
Denisova Assembly and Analysis	refresh
Variation and Repeats	refresh

Track Hubs are designed for large data sets with many genome tracks, so the data can be displayed in a matrix selection format.

UCSF-UBC-USC Histone Track Settings [Subtracks↓](#) [Description↓](#)

Histone Modification by Chip-seq Signal from REMC/UCSF-UBC-USC

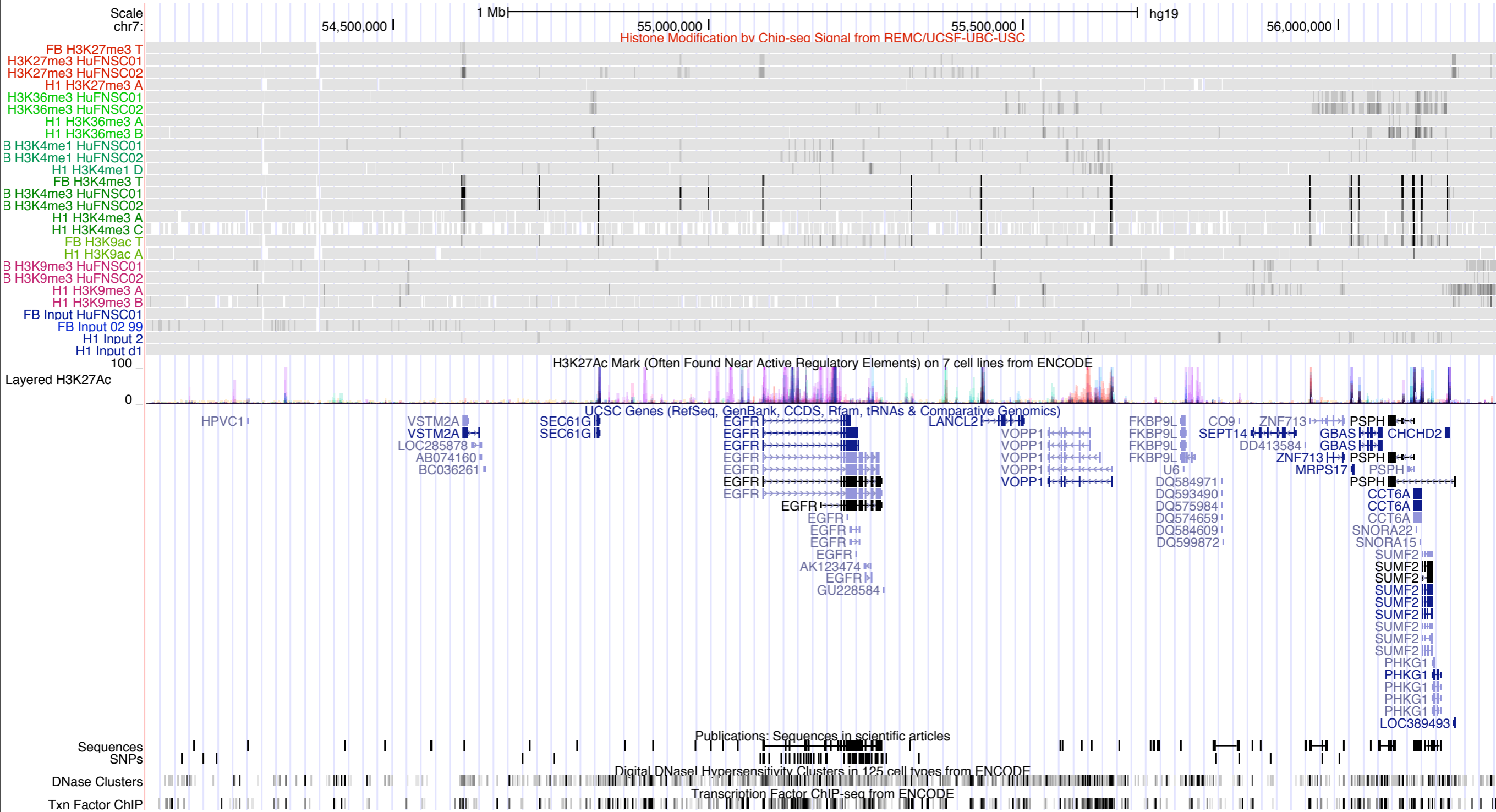
Maximum display mode: [Reset to defaults](#)

Select views ([help](#)):
[Coverage](#)

Select subtracks by assay type and sample type:

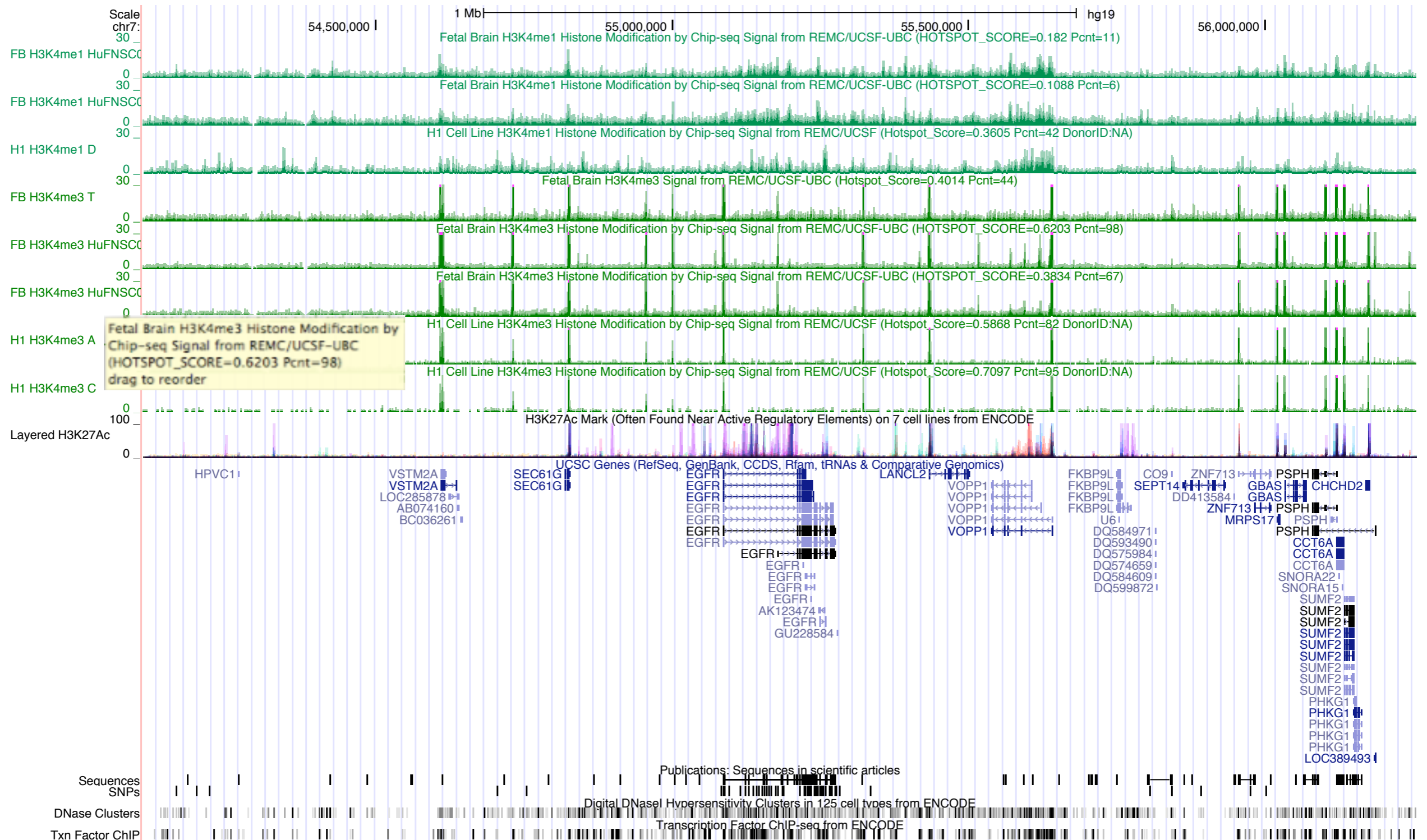
<input type="checkbox"/> <input type="checkbox"/> All	Assay Type	H3K27ac	H3K27me3	H3K36me3	H3K4me1	H3K4me3	H3K9ac	H3K9me3	Input
Sample Type		<input type="checkbox"/> <input type="checkbox"/>	<input type="checkbox"/> <input type="checkbox"/>	<input type="checkbox"/> <input type="checkbox"/>	<input type="checkbox"/> <input type="checkbox"/>	<input type="checkbox"/> <input type="checkbox"/>	<input type="checkbox"/> <input type="checkbox"/>	<input type="checkbox"/> <input type="checkbox"/>	<input type="checkbox"/> <input type="checkbox"/>
Brain Germinal Matrix	<input type="checkbox"/> <input type="checkbox"/>		<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>		<input type="checkbox"/>	<input type="checkbox"/>
Breast Luminal Epithelial Cells	<input type="checkbox"/> <input type="checkbox"/>		<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>			<input type="checkbox"/>	<input type="checkbox"/>
Breast Myoepithelial Cells	<input type="checkbox"/> <input type="checkbox"/>		<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Breast Stem Cells	<input type="checkbox"/> <input type="checkbox"/>								<input type="checkbox"/>
Breast vHMEC	<input type="checkbox"/> <input type="checkbox"/>		<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>		<input type="checkbox"/>	<input type="checkbox"/>
CD4 Naive Primary Cells	<input type="checkbox"/> <input type="checkbox"/>			<input type="checkbox"/>			<input type="checkbox"/>		
CD8 Naive Primary Cells	<input type="checkbox"/> <input type="checkbox"/>		<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Fetal Brain	<input type="checkbox"/> <input type="checkbox"/>		<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
H1Es	<input type="checkbox"/> <input type="checkbox"/>		<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
Neurosphere Cultured Cells Cortex Derived	<input type="checkbox"/> <input type="checkbox"/>		<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>		<input type="checkbox"/>	<input type="checkbox"/>
Neurosphere Cultured Cells Ganglionic Eminence Derived	<input type="checkbox"/> <input type="checkbox"/>		<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>		<input type="checkbox"/>	<input type="checkbox"/>
Peripheral Blood Mononuclear Primary Cells	<input type="checkbox"/> <input type="checkbox"/>		<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>		<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Penis Foreskin Keratinocyte Primary Cells	<input type="checkbox"/> <input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Penis Foreskin Fibroblast Primary Cells	<input type="checkbox"/> <input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>		<input type="checkbox"/>	<input type="checkbox"/>
PFM	<input type="checkbox"/> <input type="checkbox"/>		<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>		<input type="checkbox"/>	<input type="checkbox"/>
PFK	<input type="checkbox"/> <input type="checkbox"/>		<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>		<input type="checkbox"/>	<input type="checkbox"/>
PFF	<input type="checkbox"/> <input type="checkbox"/>				<input type="checkbox"/>	<input type="checkbox"/>			<input type="checkbox"/>
Penis Foreskin Melanocyte Primary Cells	<input type="checkbox"/> <input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>		<input type="checkbox"/>	<input type="checkbox"/>
UCSF-4star	<input type="checkbox"/> <input type="checkbox"/>		<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>		<input type="checkbox"/>	<input type="checkbox"/>

See Track Hub data displayed on the browser alongside any other native browser tracks.



Here we see H3K4me3 marks in H1 and Fetal Brain line up with some H3K27Ac peaks from ENCODE cell lines.

Full display mode gives a more detailed view of the data.



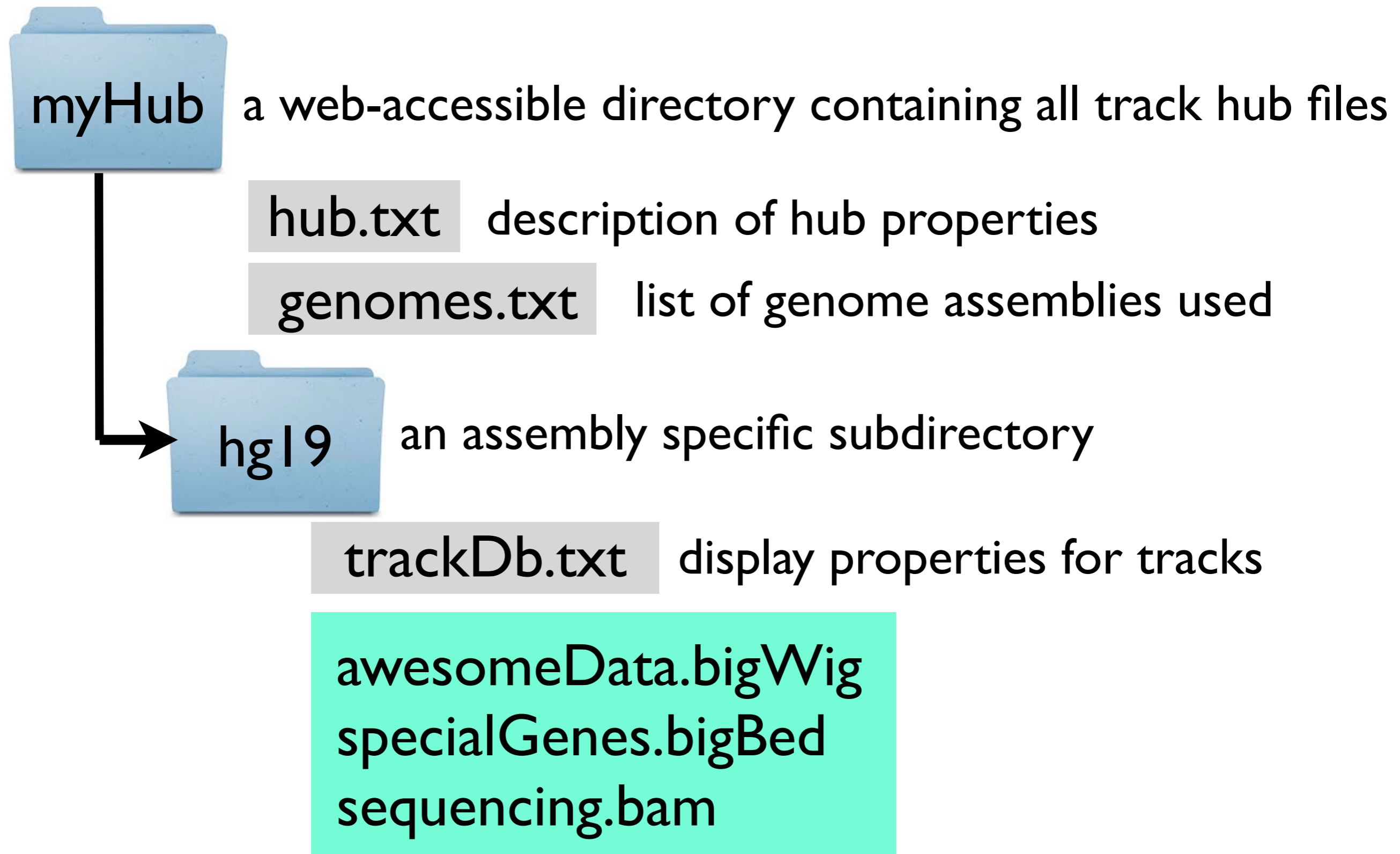
How to create your own Track Hub

You will need

- data sets formatted in one of the compressed binary index formats supported by the Genome Browser: bigBed, bigWig, BAM or VCF
- a set of text files that specify properties for the track hub and for each of the data tracks within it
- an Internet-enabled web server or ftp server

genome.ucsc.edu/goldenPath/help/hgTrackHubHelp.html

Example Hub directory



Example Hub directory

Example 1: Directory hierarchy for a hub containing DNase and RNAseq data for the hg18 and hg19 human genome assemblies. The hg18/ and hg19/ subdirectories contain the assembly-specific data files.

myHub/ - directory containing track hub files

hub.txt - a short description of hub properties

genomes.txt - list of genome assemblies included in the hub data

hg19/ - directory of data for the hg19 (GRCh37) human assembly

trackDb.txt - display properties for tracks in this directory

dnase.html - description text for a DNase track

dnaseLiver.bigWig - wiggle plot of DNase in liver

dnaseLiver.bigBed - regions of active DNase

dnaseLung.bigWig - wiggle plot of DNase in lung

dnaseLung.bigWig - regions of active DNase

...

rnaSeq.html - description text for an RNAseq track

rnaSeqLiver.bigWig - wiggle plot of RNAseq data in liver

rnaSeqLiver.bigBed - intron/exon lists for liver

rnaSeqLung.bigWig - wiggle plot of RNAseq data in lung

rnaSeqLung.bigBed - intron/exon lists for lung

hg18/ - directory of data for the hg18 (Build 36) human assembly

trackDb.txt - display properties for tracks in this directory

dnase.html - description text for a DNase track

dnaseLiver.bigWig - wiggle plot of DNase data in liver

dnaseLiver.bigBed - regions of active DNase

dnaseLung.bigWig - wiggle plot of DNase data in lung

dnaseLung.bigWig - regions of active DNase

...

rnaSeq.html - description text for an RNAseq track

rnaSeqLiver.bigWig - wiggle plot of RNAseq data in liver

rnaSeqLiver.bigBed - intron/exon lists for liver

rnaSeqLung.bigWig - wiggle plot of RNAseq data in lung

rnaSeqLung.bigBed - intron/exon lists for lung

Example text files

hub.txt - defines track hub properties

```
hub UCSCHub
shortLabel UCSC Hub
longLabel UCSC Genome Informatics Hub for human DNase and RNAseq data
genomesFile genomes.txt
email genome@soe.ucsc.edu
```

genomes.txt - lists the genome assemblies used by the hub

```
genome hg18
trackDb hg18/trackDb.txt

genome hg19
trackDb hg19/trackDb.txt
```

trackDb.txt - defines display and configuration properties for each track. This can get very long and complicated.

```
track dnaseSignal
bigDataUrl dnaseSignal.bigWig
shortLabel DNase Signal
longLabel Depth of alignments of DNase reads
type bigWig

track dnaseReads
bigDataUrl dnaseReads.bam
shortLabel DNase Reads
longLabel DNase reads mapped with MAQ
type bam
```

Feature in progress: Assembly Hubs

- Add genomic sequence (in 2bit format) to a track hub.
- Add a little additional information to a “genomes.ra” or “groups.ra” file
- Allow users to attach genome browser to a genome that is not in UCSC database.

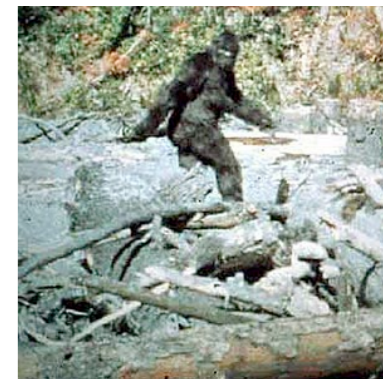
Ex: Brian Raney's test hub

Display	Hub Name	Description	Assemblies	URL	Disconnect
<input checked="" type="checkbox"/>	Braney Test Hub	Braney Test Hub Number One	hg18,hg19,mm9,newOrg1,supMou1,supMou2,megCow1	http://hgwdev.cse.ucsc.edu/~braney/assemblyHub/hub.txt	X

Supports mysterious superMouse, megaCow, and newOrg1 assemblies



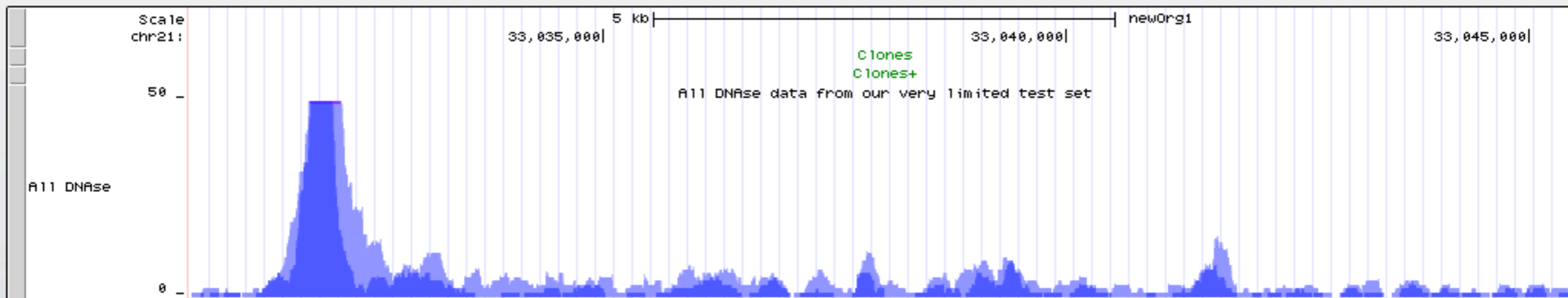
Ex: newOrg1 on test hub



UCSC TEST Genome Browser on BigFoot Big Foot V4 Assembly (newOrg1)

move <<< << < > >> >>> zoom in 1.5x 3x 10x base zoom out 1.5x 3x 10x

chr21:33,030,557-33,045,523 14,967 bp.



move start < 2.0 > move end < 2.0 >

Use drop-down controls below and press refresh to alter tracks displayed.
Tracks with lots of items will automatically be displayed in more compact modes.

Mapping

[Base Position](#) [Short Match](#) [Restr Enzymes](#)

dense hide

mRNA

[Clones](#)

pack

Regulation

Acknowledgements



Browser Staff

- *Jim Kent
- *Brian Raney
- *Galt Barber
- Ann Zweig
- Hiram Clawson
- Angie Hinrichs
- Mary Goldman
- Brooke Rhead
- Luvina Guruvadoo
- Pauline Fujita
- Donna Karolchik



Washington University in St. Louis

SCHOOL OF MEDICINE

Ting Wang

Xin Zhou