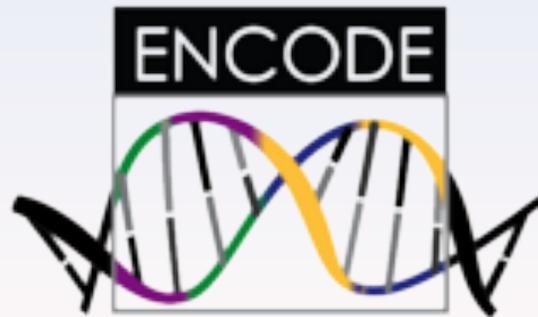


ENCODE Genome-Wide Data on the UCSC Genome Browser

Melissa Cline

ENCODE Data Coordination Center (DCC)

UC Santa Cruz



<http://encodeproject.org>

Slides at <http://genome-preview.ucsc.edu/>

UCSC Genome Browser

Genomes - Blat - Table Browser

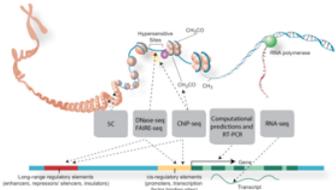
- Genome Browser
- ENCODE**
- Neanderthal
- Blat
- Table Browser
- Gene Sorter
- In Silico PCR
- Genome Graphs
- Galaxy
- VisiGene
- Proteome Browser
- Utilities
- Downloads
- Release Log

ENCODE Encyclopedia of DNA Elements

About ENCODE Data

The [Encyclopedia of DNA Elements](#) (ENCODE) Consortium is an international collaboration of research groups funded by the National Human Genome Research Institute (NHGRI). The goal of ENCODE is to build a comprehensive parts list of functional elements in the human genome, including elements that act at the protein and RNA levels, and regulatory elements that control cells and circumstances in which a gene is active.

WARNING: This is a test site. Data and tools here are under construction, have not been quality reviewed, and are subject to change at any time. For high-quality reviewed annotations on our production server, visit <http://encodeproject.org>.



ENCODE data are now available for the entire human genome. **All ENCODE data are free and available for immediate use via :**

- [Search](#) for displayable tracks and downloadable files
- [Download](#) of data files
- [Visualization](#) in the UCSC Genome Browser (ENCODE data marked with the NHGRI logo)
- [Data mining](#) with the UCSC Table Browser and other [UCSC Genome Bioinformatics tools](#)

To search for ENCODE data related to your area of interest and set up a browser view, use the UCSC [Track Search tool](#) (*Advanced* features). The [Data Summary](#) shows a comprehensive listing of ENCODE data that is released or in preparation. Early access to pre-release ENCODE data is provided at <http://genome-preview.ucsc.edu>. If you would like to receive notifications of ENCODE data releases and related news by email, subscribe to the [encode-announce mailing list](#). For more information about how to access this data, see the free online [OpenHelix ENCODE tutorial](#).

To complement the human ENCODE data, Mouse ENCODE experiments are currently underway. Early access to this data is available on the Mouse mm9/NCBI37 browser at the UCSC preview site. The [Mouse ENCODE Data Summary](#) lists experiments that are planned or in progress.

All ENCODE data is freely available for download and analysis. However, before publishing research that uses ENCODE data, please read the [ENCODE Data Release Policy](#), which places some restrictions on publication use of data for nine months following data release. [Read more](#) about ENCODE data at UCSC.

News

8 July 2011 - **Mouse ENCODE data releases: DNaseI hypersensitivity (UW DNaseI HS) and histone modifications (LICR Histone)**

Two tracks of ENCODE data were released on the mm9 genome browser, from the UCSD/Ludwig Institute for Cancer Research and the University of Washington Mouse ENCODE groups.

[Histone Modifications by ChIP-seq from ENCODE/LICR](#): This track shows a comprehensive survey of cis-regulatory elements in the mouse genome by using ChIP-seq to identify transcription factor binding sites and chromatin modification profiles in many mouse tissues and primary cells, including bone marrow, cerebellum, cortex, heart, kidney, liver, lung, spleen, mouse embryonic fibroblast cells (MEFs) and embryonic stem (ES) cells.

[DNaseI Hypersensitivity by Digital DNaseI from ENCODE/University of Washington](#): This track shows DNaseI sensitivity measured genome-wide in mouse tissues and cell lines using the Digital DNaseI methodology and DNaseI hypersensitive sites.

1 July 2011 - **ENCODE data releases: Broad ChromHMM, Open Chrom Synth, UChicago TFBS, Duke Affy Exon**

Four tracks of ENCODE production data and analysis were released in June, from the Broad Institute (Kellis lab), OpenChromatin (Duke, UNC, UT-A) and University of Chicago (White Lab) ENCODE groups. This is the first data release from the University of Chicago ENCODE group, which joined the Consortium as part of the NIH ARRA stimulus grants.

[Chromatin State Segmentation by HMM from ENCODE/Broad](#): This track, and the companion hg18 track, display chromatin state segmentation of the human genome into fifteen states grouped to predict functional elements.

[DNaseI/FAIRE/ChIP Synthesis from ENCODE/OpenChrom\(Duke/UNC/UTA\)](#): This track displays a synthesis of open chromatin regions and binding of selected regulatory factors, based on three complementary methodologies.

[Transcription Factor Binding Sites by Epitope-Tag ChIP-seq from ENCODE/University of Chicago](#): This track maps human transcription factor binding sites genome-wide using expressed transcription factors as GFP tagged fusion proteins after BAC recombineering.

What is ENCODE?

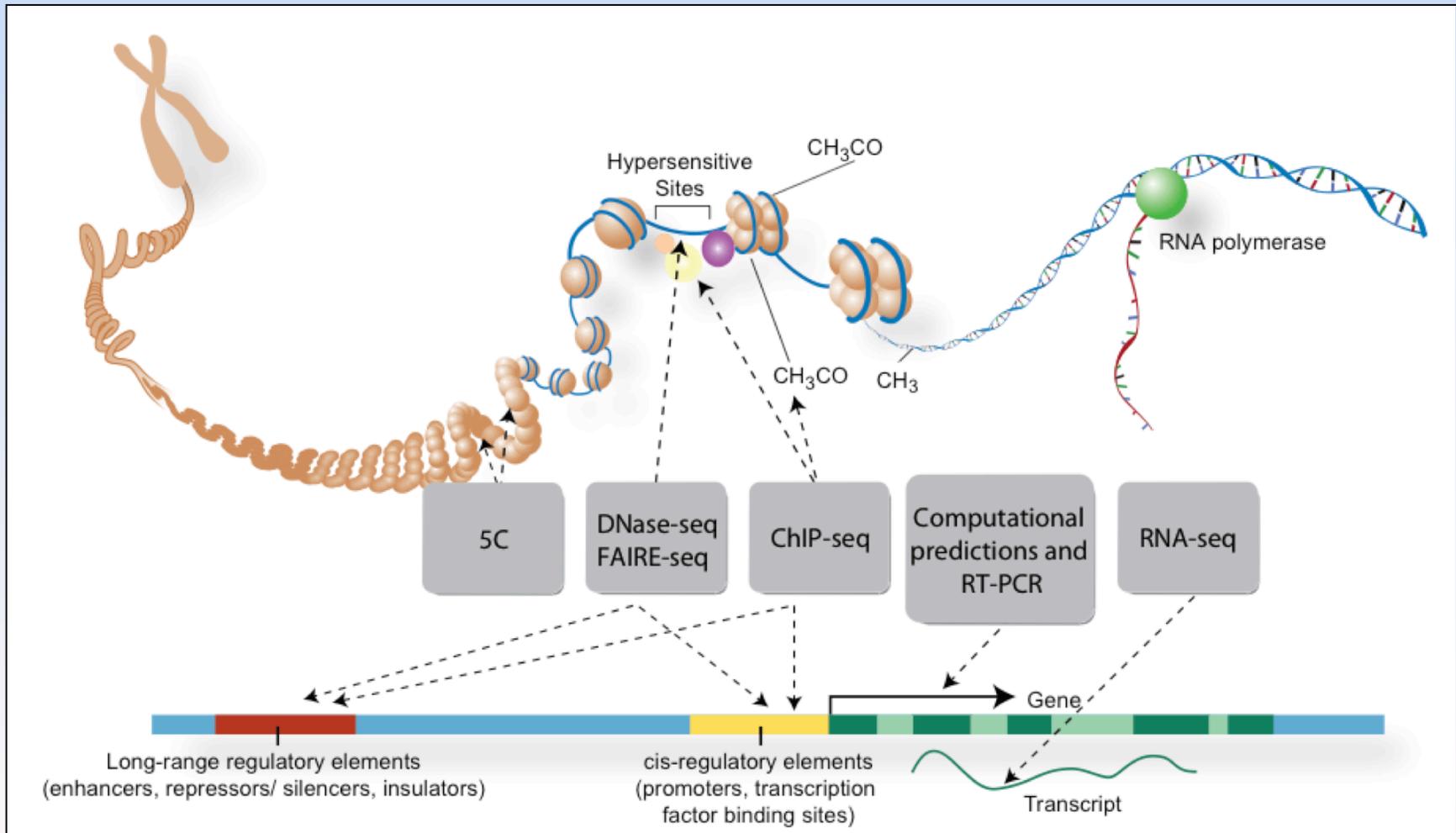
- International consortium project with the goal of cataloguing the functional regions of the human genome

```
GTTTGCCATCTTTTG  
CTGCTCTAGGGAATC  
CAGCAGCTGTCACCA  
TGTAACAAGCCCAG  
GCTAGACCAGTTACC  
CTCATCATCTTAGCT  
GATAGCCAGCCAGCC  
ACCACAGGCATGAGT
```

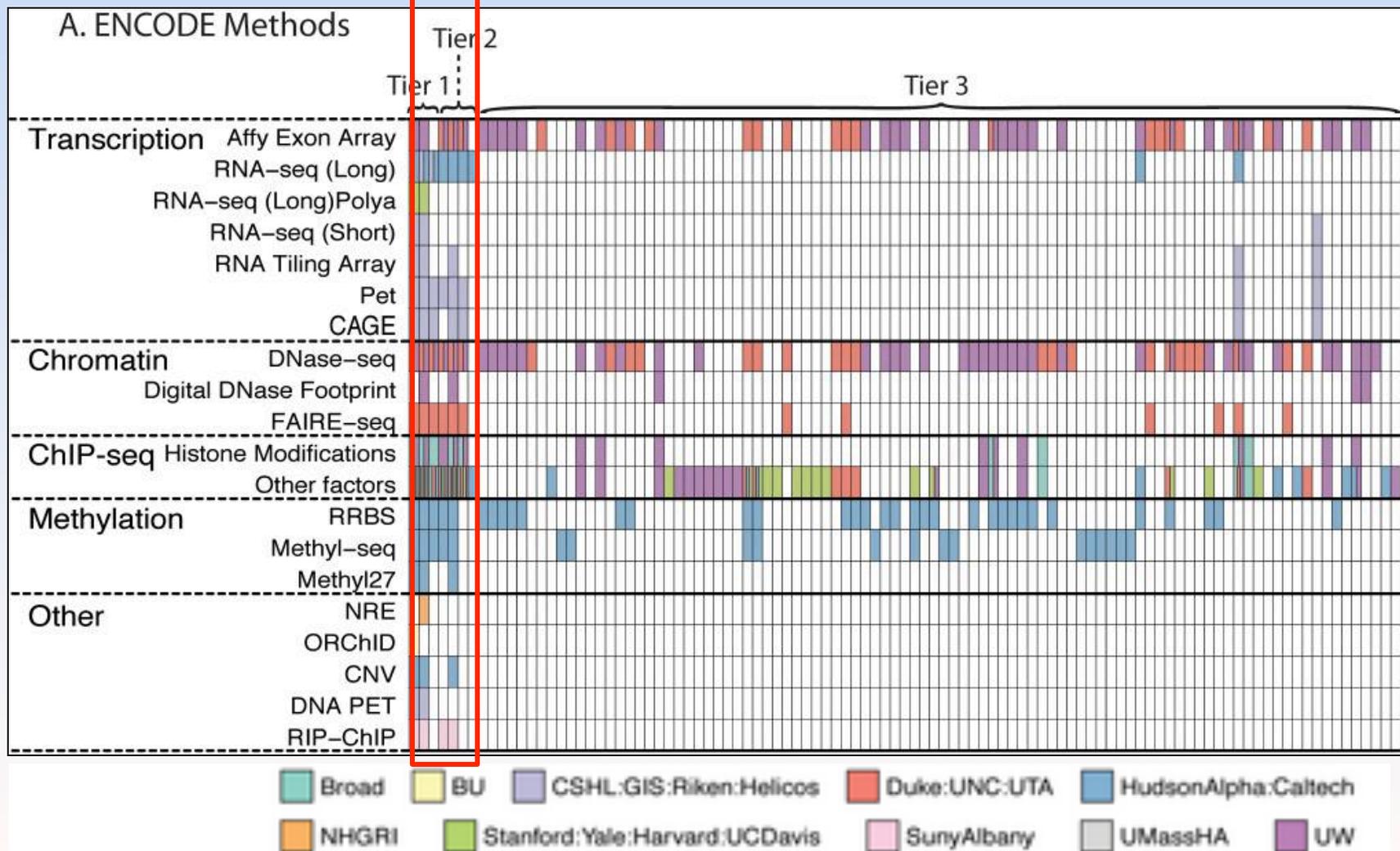


- A gold mine of experimental data for independent researchers with available disk space

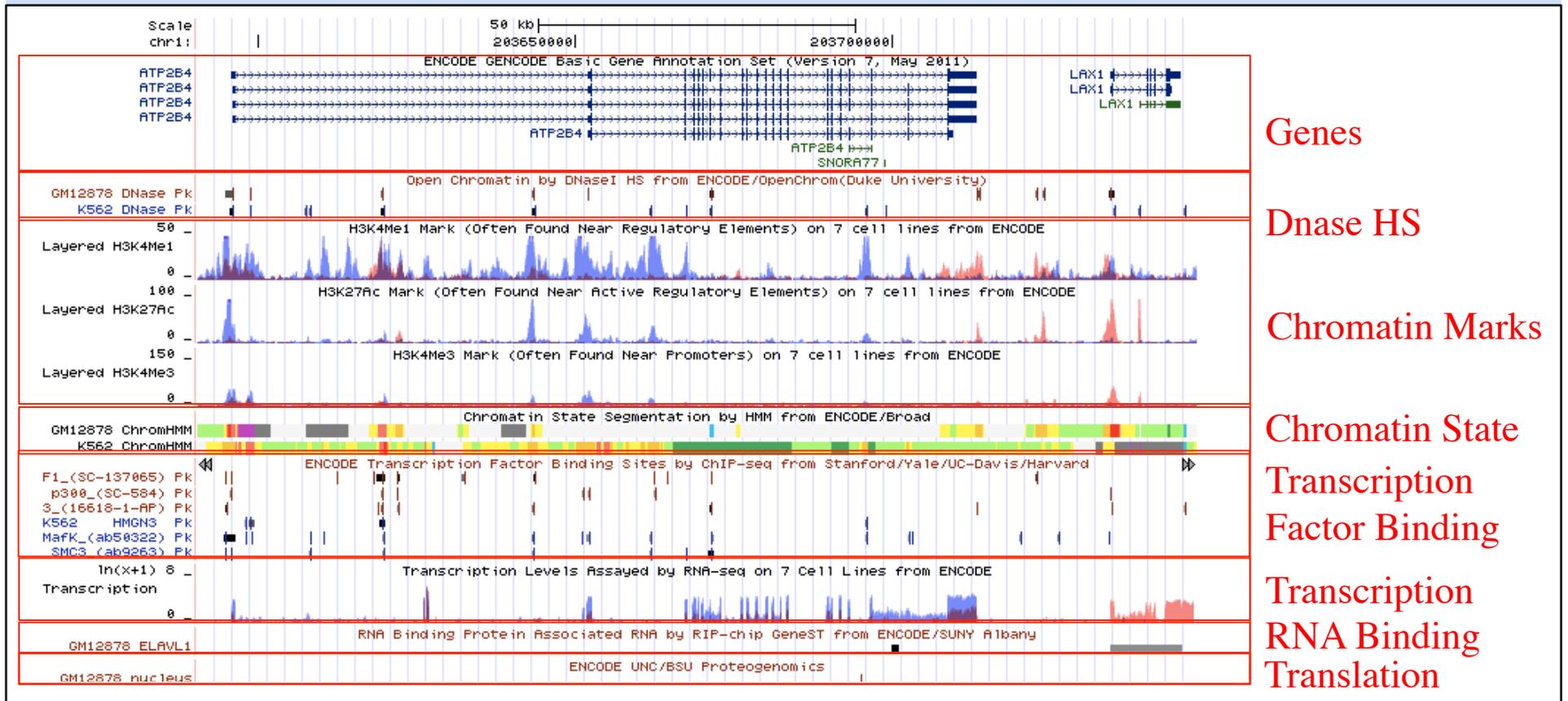
ENCODE covers diverse regulatory processes



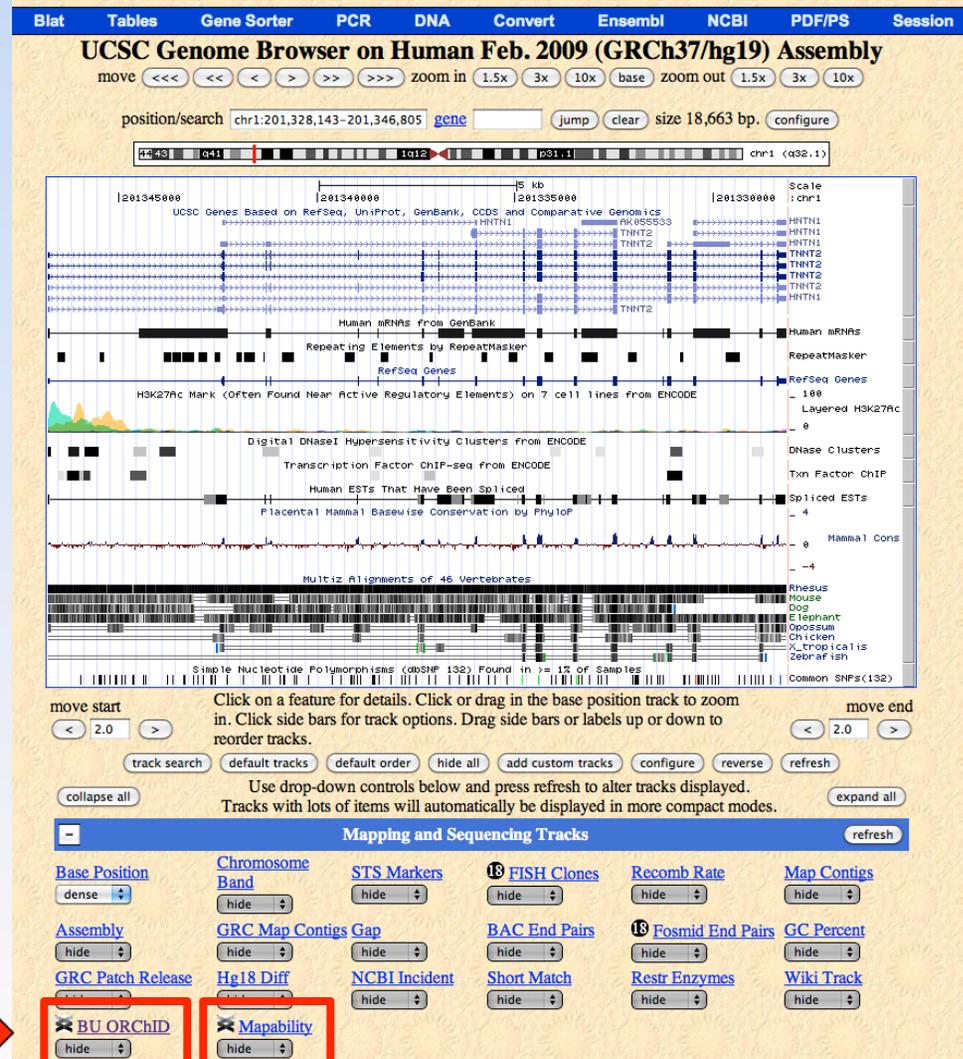
ENCODE experiments are planned for integrative analysis



Example of ENCODE data



ENCODE tracks on the UCSC Genome Browser



*ENCODE tracks
marked with the
NHGRI helix*



There are currently 2061 ENCODE experiments at the ENCODE DCC

How to find the data you want



Finding ENCODE tracks the hard way

Home Genomes Genome Browser Blat Tables Gene Sorter PCR Session FAQ Help

RIKEN CAGE Loc Track Settings [Downloads](#) [Subtracks](#) [Description](#)

RNA CAGE Subcellular Localization from ENCODE/RIKEN (▲ENC RNA-seq)

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

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

Select subtracks by cell line and localization:

RNA Extract: PolyA- PolyA+ Total RNA
 Rep: 1 2

Cell Line	GM12878 (Tier 1)	H1hESC (Tier 1)	K562 (Tier 1)	HeLa-S3	HepG2	HUVEC	MCF7	NHEK	Prostate	Cell Line
Whole Cell	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	Whole Cell					
Cytosol	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	Cytosol					
Nucleus	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	Nucleus					
Polysome	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	Polysome					
Nucleoplasm	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	Nucleoplasm					
Chromatin	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	Chromatin					
Nucleolus	<input checked="" 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"/>	Nucleolus

List subtracks: only selected/visible all (38 of 153 selected)

Views ¹	Cell Line ^{1,2}	Localization ^{1,3}	RNA Extract ^{1,4}	Rep ^{1,5}	Track Name	Restri
<input checked="" type="checkbox"/> Plus Signal	GM12878	Cytosol	PolyA-	1	GM12878 cytosol polyA- CAGE Plus strand start sites Generation 0 from ENCODE/RIKEN ...	schema 2009
<input checked="" type="checkbox"/> Plus Signal	GM12878	Nucleolus	Total RNA	1	GM12878 nucleolus total CAGE Plus strand start sites Generation 0 from ENCODE/RIKEN ...	schema 2010
<input checked="" type="checkbox"/> Plus Signal	GM12878	Nucleus	PolyA-	1	GM12878 nucleus polyA- CAGE Plus strand start sites Rep 1 from ENCODE/RIKEN ...	schema 2011
<input checked="" type="checkbox"/> Plus Signal	H1hESC	Whole Cell	PolyA-	1	H1-hESC whole cell polyA- CAGE Plus strand start sites Generation 0 from ENCODE/RIKEN ...	schema 2010
<input checked="" type="checkbox"/> Plus Signal	H1hESC	Whole Cell	PolyA+	1	H1-hESC whole cell polyA+ CAGE Plus strand start sites Rep 1 from ENCODE/RIKEN ...	schema 2011
<input checked="" type="checkbox"/> Plus Signal	H1hESC	Whole Cell	PolyA+	2	H1-hESC whole cell polyA+ CAGE Plus strand start sites Rep 2 from ENCODE/RIKEN ...	schema 2011
<input checked="" type="checkbox"/> Plus Signal	H1hESC	Cytosol	PolyA+	2	H1-hESC cytosol polyA+ CAGE Plus strand start sites Rep 2 from ENCODE/RIKEN ...	schema 2011
<input checked="" type="checkbox"/> Plus Signal	H1hESC	Nucleus	PolyA+	2	H1-hESC nucleus polyA+ CAGE Plus strand start sites Rep 2 from ENCODE/RIKEN ...	schema 2011
<input checked="" type="checkbox"/> Plus Signal	K562	Whole Cell	PolyA+	1	K562 whole cell polyA+ CAGE Plus strand start sites Rep 1 from ENCODE/RIKEN ...	schema 2011
<input checked="" type="checkbox"/> Plus Signal	K562	Whole Cell	PolyA+	2	K562 whole cell polyA+ CAGE Plus strand start sites Rep 2 from ENCODE/RIKEN ...	schema 2011
<input checked="" type="checkbox"/> Plus Signal	K562	Chromatin	Total RNA	1	K562 chromatin total CAGE Plus strand start sites Generation 0 from ENCODE/RIKEN ...	schema 2009
<input checked="" type="checkbox"/> Plus Signal	K562	Cytosol	PolyA-	1	K562 cytosol polyA- CAGE Plus strand start sites Generation 0 from ENCODE/RIKEN ...	schema 2009
<input checked="" type="checkbox"/> Plus Signal	K562	Cytosol	PolyA+	1	K562 cytosol polyA+ CAGE Plus strand start sites Rep 1 from ENCODE/RIKEN ...	schema 2011
<input checked="" type="checkbox"/> Plus Signal	K562	Cytosol	PolyA+	2	K562 cytosol polyA+ CAGE Plus strand start sites Rep 2 from ENCODE/RIKEN ...	schema 2011

Metzker ML. *Sequencing technologies - the next generation*. Nature Reviews: Genetics. 2010 Jan;11(1):31-46

Allen Brain
hide

Burge RNA-seq
hide

ENC Exon
Array...
hide

ENC RNA-seq
hide

GNF Atlas 2
hide

Illumina WG-
6

Finding ENCODE metadata descriptions

Encyclopedia of DNA Elements

About ENCODE Data

The [Encyclopedia of DNA Elements](#) (ENCODE) Consortium is an international collaboration of research groups funded by the National Human Genome Research Institute (NHGRI). The goal of ENCODE is to build a comprehensive parts list of functional elements in the human genome, including elements that act at the protein and RNA levels, and regulatory elements that control cells and circumstances in which a gene is active.

WARNING: This is a test site. Data and tools here are under construction, have not been quality reviewed, and are subject to change at any time. For high-quality reviewed annotations on our production server, visit <http://encodeproject.org>.

ENCODE data are now available for the entire human genome. *All ENCODE data are free and available for immediate use via:*

- [Search](#) for displayable tracks and downloadable files

ENCODE Data Coordination Center at UCSC

[Home](#) - [Data Policy](#) - [Help](#)

ENCODE Common Cell Types

To facilitate integration of data between the contributing research groups, the ENCODE Consortium has identified common cell types for use by ENCODE contributors. These cell types are divided into two *Tiers*. **Tier1** cells are of higher priority, and should be used within experiments before **Tier2** cells. Rationale for the selection is described on the [NHGRI ENCODE Common Cell Types](#) page. Additional cell types beyond Tier1 and Tier2 may be used, but must be registered with the DCC before submitting data. These additional cell types are designated Tier3.

Click the link in the **Documents** column of the table below to access the cell culture protocol document.

Common Cell Types: Tier 1 and Tier 2

Cell, tissue or DNA sample: Cell line or tissue used as the source of experimental material.

cell	Tier	Description	Lineage	Karyotype	Sex	Documents	Vendor ID	Term ID	Label
GM12878	1	lymphoblastoid	International HapMap Project - CEPH/Utah - European Caucasian; Epstein-Barr Virus	relatively normal	F	protocol	Coriell GM12878	BTO:0002026 (non-specific)	GM12878
H1-hESC	1	Human Embryonic Stem Cells	embryonic stem cells	normal	M	protocol	Cellular Dynamics	CL:0000007	H1-hESC
K562	1	leukemia	"The continuous cell line K-562 was established by Lozzio and Lozzio from the pleural effusion of a 53-year-old female with chronic myelogenous leukemia in terminal blast crises." - ATCC	cancer	F	protocol	ATCC CCL-243	BTO:0000664	K562
K562b	1	leukemia (UCDavis)	"The continuous cell line K-562 was established by Lozzio and Lozzio from the pleural effusion of a 53-year-old female with chronic myelogenous leukemia	cancer	F	Yale TFBS(Farnham)	ATCC CCL-243	BTO:0000664	K562b

Visualizing: Genome Browser tricks that every ENCODE user should know



Turning ENCODE subtracks and views on and off

Home Genomes Genome Browser Blat Tables Gene Sorter PCR Session FAQ Help

SYDH Histone Track Settings [Downloads](#) [Subtracks↓](#) [Description↓](#)

← Peaks
← Signal

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

Select views ([help](#)):
Peaks **Signal** ← View on/off

Select subtracks by factor and cell line:

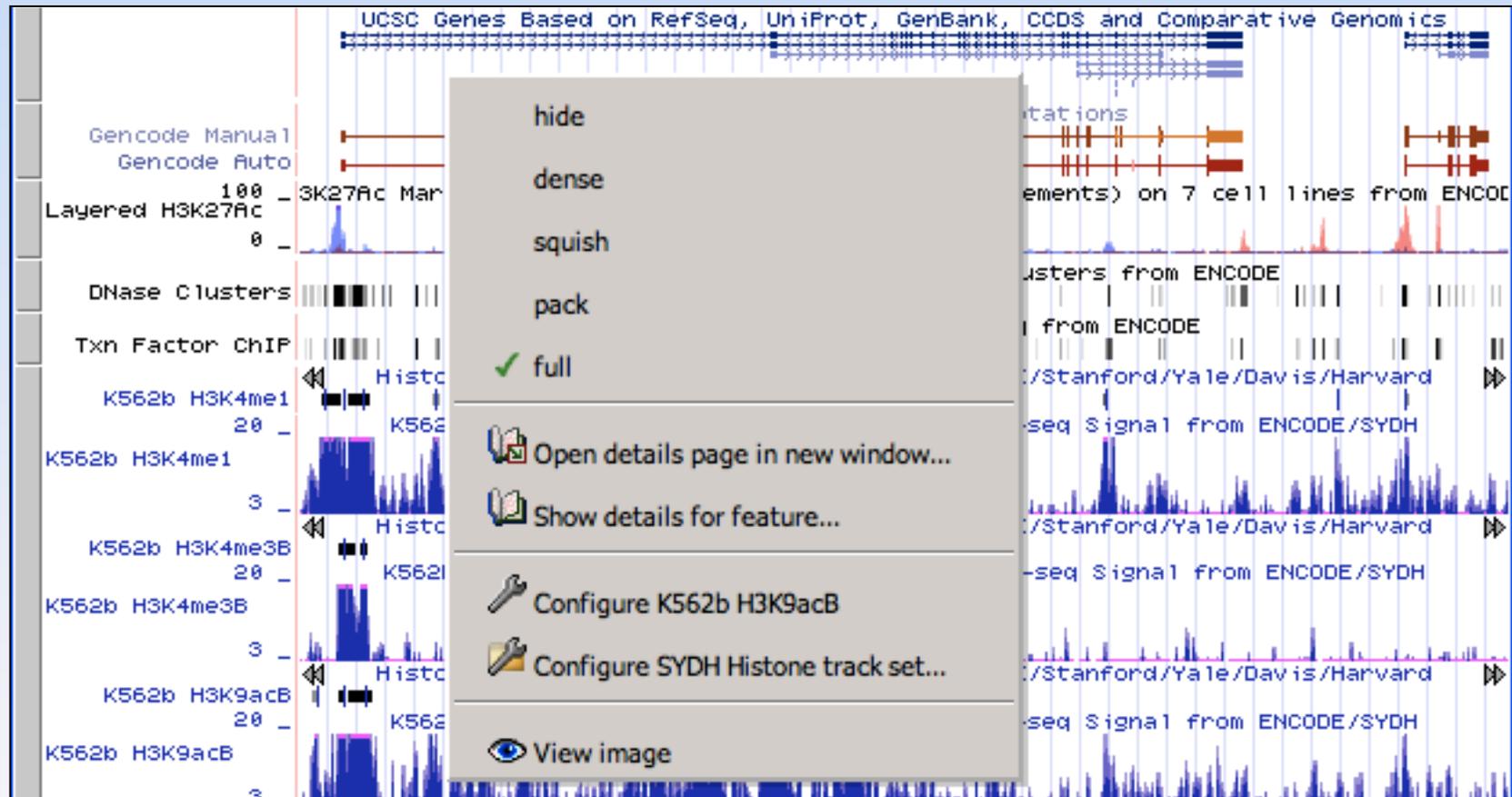
+ -	All	Factor	H3K4me3B	H3K4me1	H3K9acB	H3K9me3	H3K27me3B	H3K36me3B	Input
		Cell Line	+ -	+ -	+ -	+ -	+ -	+ -	+ -
		K562b (Tier 1)	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>		<input checked="" type="checkbox"/>		<input type="checkbox"/>
		NT2-D1	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>		<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>
		U2OS				<input checked="" type="checkbox"/>		<input checked="" type="checkbox"/>	<input type="checkbox"/>

← Subtrack on/off

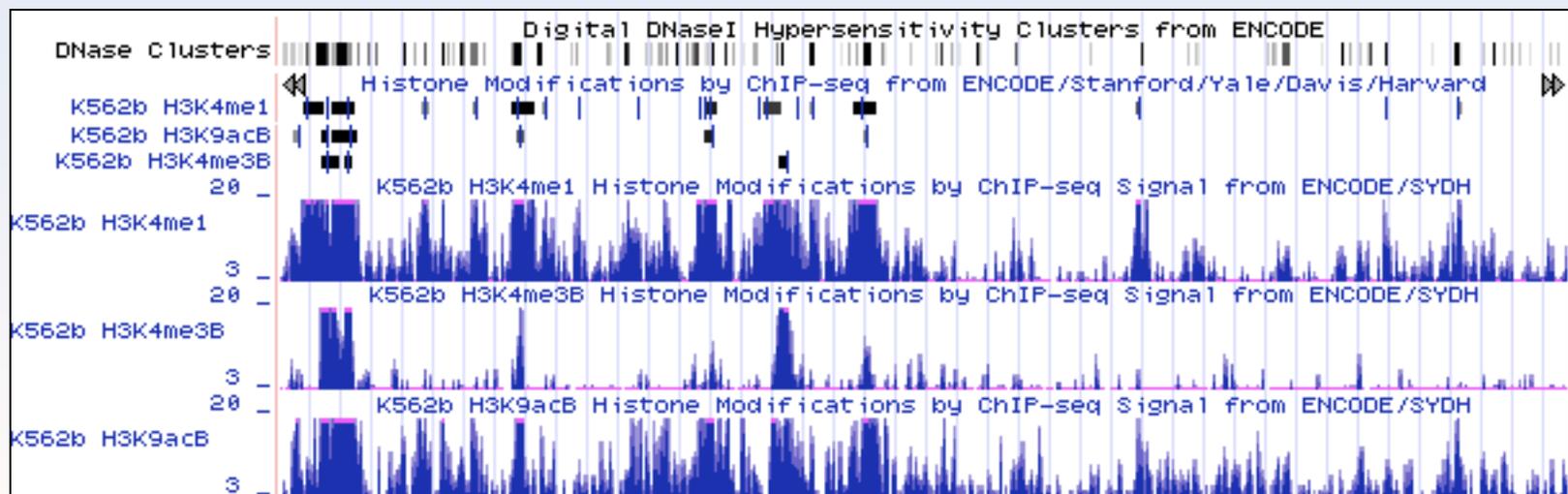
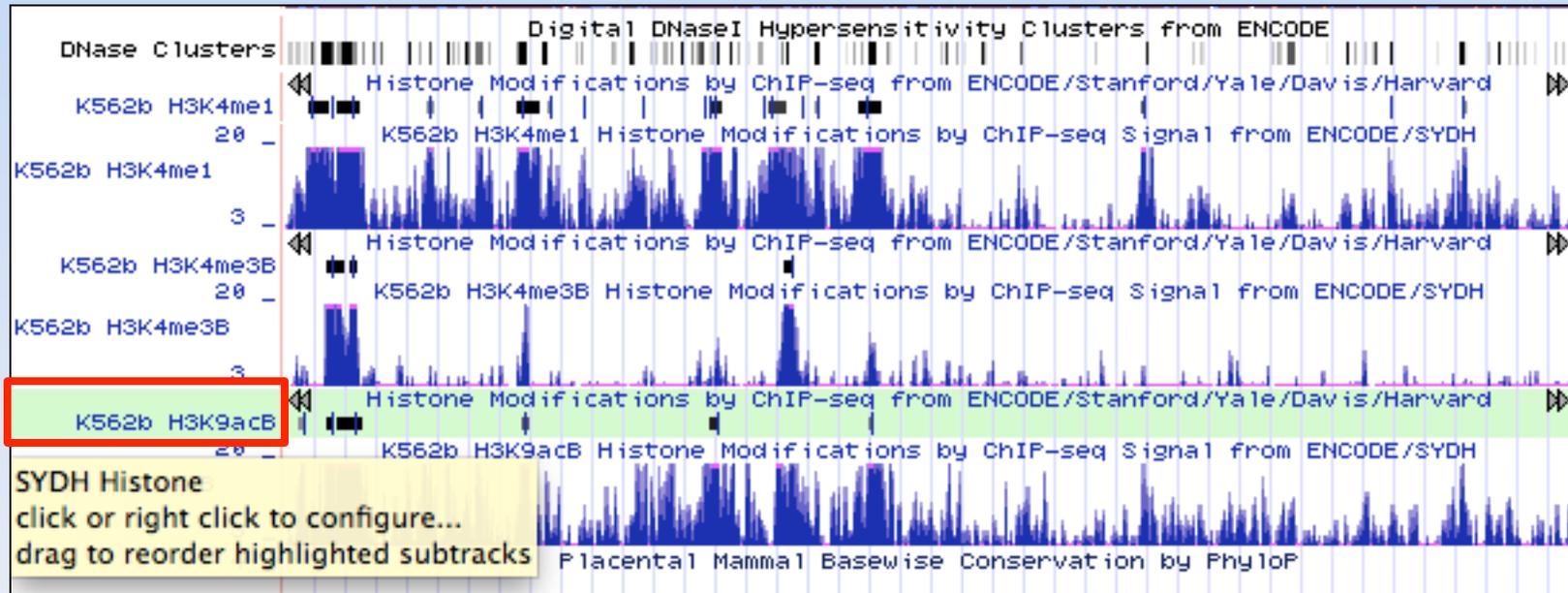
List subtracks: only selected/visible all (22 of 25 selected) [Top↑](#)

	Cell Line ^{↓1}	Factor ^{↓2}	Views ^{↓3}	Track Name ^{↓4}		Restricted Until ^{↓5}
<input checked="" type="checkbox"/>	K562b	H3K27me3B	Peaks	K562b H3K27me3B Histone Modifications by ChIP-seq Peaks from ENCODE/SYDH ...	schema	2011-03-11
<input checked="" type="checkbox"/>	K562b	H3K27me3B	Signal	K562b H3K27me3B Histone Modifications by ChIP-seq Signal from ENCODE/SYDH ...	schema	2011-03-11
<input checked="" type="checkbox"/>	K562b	H3K4me1	Peaks	K562b H3K4me1 Histone Modifications by ChIP-seq Peaks from ENCODE/SYDH ...	schema	2011-02-20
<input checked="" type="checkbox"/>	K562b	H3K4me1	Signal	K562b H3K4me1 Histone Modifications by ChIP-seq Signal from ENCODE/SYDH ...	schema	2011-02-20
<input checked="" type="checkbox"/>	K562b	H3K4me3B	Peaks	K562b H3K4me3B Histone Modifications by ChIP-seq Peaks from ENCODE/SYDH ...	schema	2011-03-11
<input checked="" type="checkbox"/>	K562b	H3K4me3B	Signal	K562b H3K4me3B Histone Modifications by ChIP-seq Signal from ENCODE/SYDH ...	schema	2011-03-11
<input checked="" type="checkbox"/>	K562b	H3K9acB	Peaks	K562b H3K9acB Histone Modifications by ChIP-seq Peaks from ENCODE/SYDH ...	schema	2011-03-11

Right-click to the subtrack display menu



Subtrack Drag and Drop



Sessions: the easy way to save and share your work

The screenshot shows the UCSC Genome Browser interface with a navigation bar at the top containing links like Home, Genomes, Blat, Tables, Gene Sorter, PCR, DNA, Convert, Ensembl, NCBI, PDF/PS, and Session (highlighted with a red box). Below the navigation bar, a window titled "hg19) Assembly" is open, displaying an email composition form. The form includes buttons for "Send", "Save Now", and "Discard". The "From:" field is filled with "Melissa Cline <cline@soe.ucsc.edu>". The "To:" field is empty. Below the "To:" field are links for "Add Cc" and "Add Bcc". The "Subject:" field contains "UCSC browser session hg19-demo". There are also links for "Attach a file" and "Insert: Invitation". At the bottom of the form, there are links for "Rich formatting »" and "Check Spelling ▾". The email body text reads: "Here is a UCSC browser session I'd like to share with you: http://genome.ucsc.edu/cgi-bin/hgTracks?hgS_doOtherUser=submit&hgS_otherUserName=Melissa%20Cline%20Demo%20Account&hgS_otherUserSessionName=hg19%2Ddemo". Below the email composition window, a table lists session details:

session name	created on	use this session	delete this session	share with others?	link to session	send to mail
hg19-demo	2011-06-25	<input type="button" value="use"/>	<input type="button" value="delete"/>	<input checked="" type="checkbox"/>	Browser	<input style="border: 2px solid red;" type="button" value="Email"/>

Downloading data with less pain



1. Via the Downloads button on the track details page

Home Genomes Genome Browser Blat Tables Gene Sorter PCR Session FAQ Help

UTA TFBS Track Settings [Downloads](#) [Subtracks](#) [Description](#)

Home Genomes Blat Tables Gene Sorter PCR Session FAQ Help

UTA TFBS Downloadable Files [File Search](#) [Description](#)

Open Chromatin TFBS by ChIP-seq from ENCODE/Open Chrom(UT Austin)

Data is **RESTRICTED FROM USE** in publication until the restriction date noted for the given data file.

Supporting documents:

- [files.txt](#) is a tab-separated file with the name and metadata for each download.
- [md5sum.txt](#) is a list of the md5sum output for each download.
- [Supplemental materials](#) contains additional files provided by the laboratory related to these downloads.

Filter files by: (select multiple categories and items - [help](#))

352 files	Cell Line ¹	Antibody Target ²	View ³	Size ⁴	File Type ⁵	Submitted	RESTRICTED Until	Additional Details
Download	Fibrobl	CTCF	Alignments	5.6 MB	bam.bai	2010-10-13	2011-07-13	replicate=1; origAssembly=hg19; dataVersion=1
Download	Fibrobl	CTCF	Alignments	5.7 MB	bam.bai	2010-10-13	2011-07-13	replicate=2; origAssembly=hg19; dataVersion=1
Download	Fibrobl	CTCF	Alignments	377 MB	bam	2010-10-13	2011-07-13	replicate=1; origAssembly=hg19; dataVersion=1
Download	Fibrobl	CTCF	Alignments	382 MB	bam	2010-10-13	2011-07-13	replicate=2; origAssembly=hg19; dataVersion=1
Download	Fibrobl	CTCF	Base_Overlap_Signal	172 MB	bigWig	2010-10-13	2011-07-13	origAssembly=hg19; dataVersion=ENCODE Ja
Download	Fibrobl	CTCF	Peaks	1.2 MB	narrowPeak	2010-10-13	2011-07-13	origAssembly=hg19; dataVersion=ENCODE Ja
Download	Fibrobl	CTCF	Signal	4.1 GB	bigWig	2010-10-13	2011-07-13	origAssembly=hg19; dataVersion=ENCODE Ja
Download	Fibrobl	Input	Alignments	6.1 MB	bam.bai	2010-10-14	2011-07-14	origAssembly=hg19; dataVersion=ENCODE Ja
Download	Fibrobl	Input	Alignments	745 MB	bam	2010-10-14	2011-07-14	origAssembly=hg19; dataVersion=ENCODE Ja

2. Via the File Selection tool

The screenshot displays the ENCODE Data Coordination Center at UCSC website. The main heading is "ENCODING Encyclopedia of DNA Elements". The page is titled "ENCODING Data Coordination Center at UCSC". The left sidebar contains navigation links for Human, Mouse, and General categories. The main content area is titled "ENCODING Downloads" and contains a search tool for downloadable ENCODE files in the Human Feb. 2009 (GRCh37/hg19) Assembly. The search tool includes fields for Track Name, Description, Group, and Data Format. The "and Group:" field is highlighted with a red box and set to "Expression". Below the search tool, there are two search criteria: "+ and Cell, tissue or DNA sample is among Urothelia" and "+ and Antibody or target protein is among Any". The "Cell, tissue or DNA sample" criterion is also highlighted with a red box. The search results table shows two entries: "wgEncodeEH001076" and "UT189", both with a size of 2.2 MB and data type of broadPeak. The table headers are "Treatment", "UCSC Accession", "Size", "File Type", and "Additional Details".

ENCODING Encyclopedia of DNA Elements

About ENCODE Data

ENCODING Data Coordination Center at UCSC

Home - Help

ENCODING Downloads

This page contains links to directories containing raw and processed data for ENCODE data released as part of the ENCODE production phase (September 2007-present). For bulk download, retrieval by [FTP](#) is recommended. All files here are covered by the [ENCODING data release policy](#). Preview of unreleased data for

Home Genomes Blat Tables Gene Sorter PCR Session FAQ Help

Search for Downloadable ENCODE Files in the Human Feb. 2009 (GRCh37/hg19) Assembly

Track Name: contains

and Description: contains

and Group: is Expression

and Data Format: is Any

ENCODING terms

+ and Cell, tissue or DNA sample is among Urothelia Cell, tissue or DNA sample

+ and Antibody or target protein is among Any Antibody or target protein

search clear cancel

	Treatment ¹	UCSC Accession ²	Size ³	File Type ⁴	Additional Details ⁵
Download		wgEncodeEH001076	2.2 MB	broadPeak	grant=Crawford; lab=Duke; dataType=AffyExonArray; view=SimpleSignal; cell=Urothelia; date
Download	UT189	wgEncodeEH001077	2.2 MB	broadPeak	grant=Crawford; lab=Duke; dataType=AffyExonArray; view=SimpleSignal; cell=Urothelia; date

Publishing: the ENCODE data release policy



Every ENCODE subtrack has a “Restricted Until” date

Home Genomes Genome Browser Blat Tables Gene Sorter PCR Session FAQ Help

LICR Histone Track Settings [Downloads](#) [Subtracks↓](#) [Description↓](#)

Histone Modifications by ChIP-seq from ENCODE/LICR (▲[All Expression and Regulation tracks](#))

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

Select views ([help](#)):

Data Release Policy

Data users may freely use ENCODE data, but may not, without prior consent, submit publications that use an unpublished ENCODE dataset until nine months following the release of the dataset. This date is listed in the *Restricted Until* column on the track configuration page and the download page. The full data release policy for ENCODE is available [here](#).

Cell Line ¹	Factor ²	Views ³	Track Name ⁴	Restricted Until ⁵
<input checked="" type="checkbox"/>	Bone Marrow H3K4me1	Peaks	Bone Marrow H3K4me1 Histone Modifications by ChIP-seq Peaks from ENCODE/LICR ... schema	2011-10-19
<input checked="" type="checkbox"/>	Bone Marrow H3K4me1	Signal	Bone Marrow H3K4me1 Histone Modifications by ChIP-seq Signal from ENCODE/LICR ... schema	2011-10-19
<input checked="" type="checkbox"/>	Bone Marrow H3K4me3	Peaks	Bone Marrow H3K4me3 Histone Modifications by ChIP-seq Peaks from ENCODE/LICR ... schema	2011-10-19
<input checked="" type="checkbox"/>	Bone Marrow H3K4me3	Signal	Bone Marrow H3K4me3 Histone Modifications by ChIP-seq Signal from ENCODE/LICR ... schema	2011-10-19
<input type="checkbox"/>	Bone Marrow Input	Signal	Bone Marrow Input Histone Modifications by ChIP-seq Signal from ENCODE/LICR ... schema	2012-01-26
<input checked="" type="checkbox"/>	Cerebellum H3K4me1	Peaks	Cerebellum H3K4me1 Histone Modifications by ChIP-seq Peaks from ENCODE/LICR ... schema	2011-12-07
<input checked="" type="checkbox"/>	Cerebellum H3K4me1	Signal	Cerebellum H3K4me1 Histone Modifications by ChIP-seq Signal from ENCODE/LICR ... schema	2011-12-07

Key points of the ENCODE data release policy

- Anyone is free to download and analyze data.
- One cannot submit publications involving ENCODE data unless
 - the data has been at the ENCODE DCC for at least nine months, or
 - the data producers have published on the data, or
 - the data producers have granted permission to publish.
- Please acknowledge the ENCODE project and the data producers in all publications.

Additional Resources

- The OpenHelix ENCODE tutorial at <http://www.openhelix.com/ENCODE/>
- [A User's Guide to the Encyclopedia of DNA Elements \(ENCODE\)](#) in PLoS Biology
- The ENCODE Project page at <http://www.genome.gov/10005107>
- The UCSC Genome Browser User's Guide at <http://genome.ucsc.edu/goldenPath/help/>

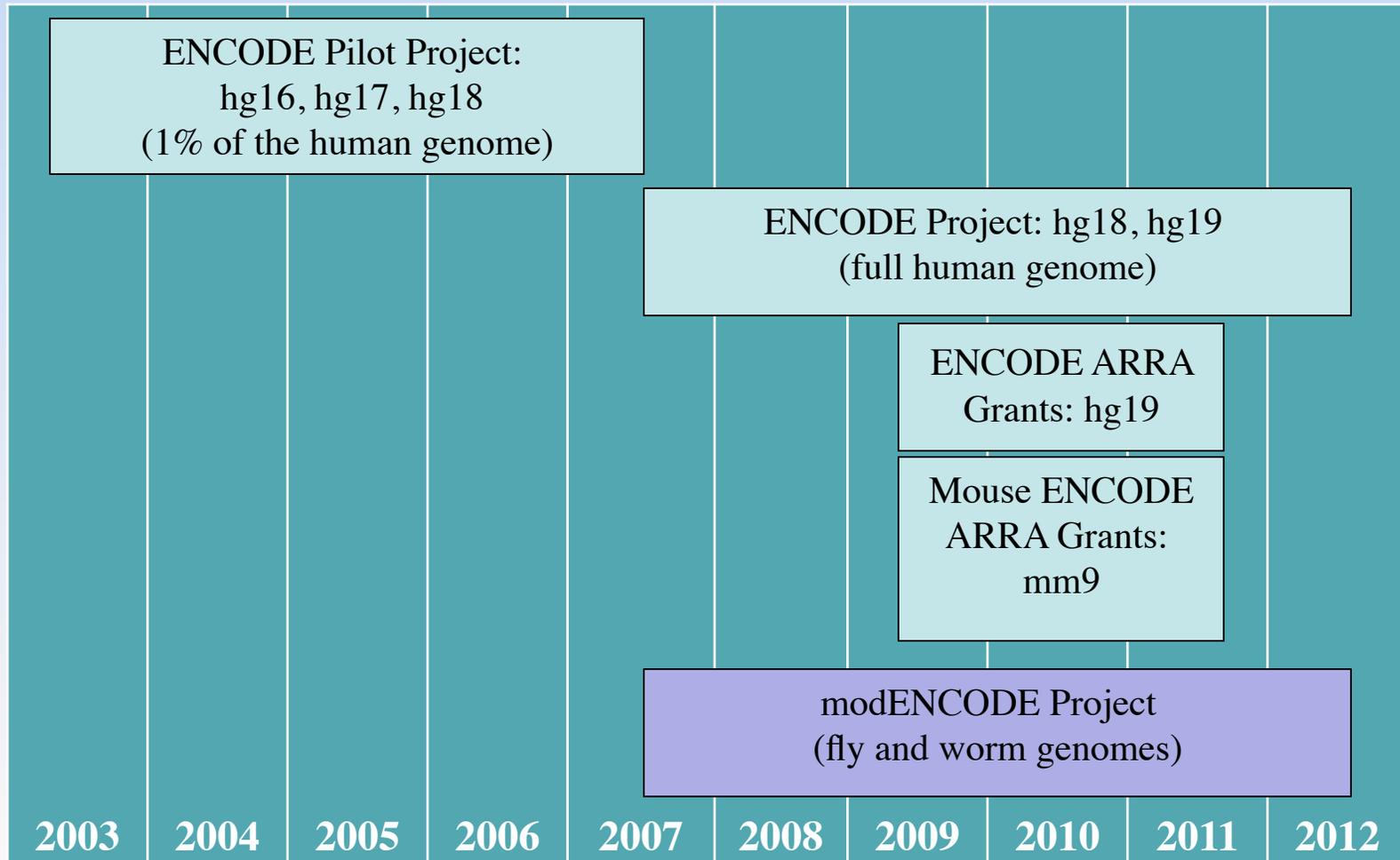
Acknowledgements

- **Cat Herding:** *Kate Rosenbloom and Jim Kent*
- **Cats:** *Galt Barber, Tim Dreszer, Katrina Learned, Venkat Malladi, Brian Raney, Cricket Sloan, Vanessa Swing, and Matt Wong*
- **Systems:** *Jorge Garcia, Victoria Lin, and Erich Weiler*
- *The UCSC Genome Browser staff*
- *The ENCODE consortium*
- **Funding:** *NHGRI*

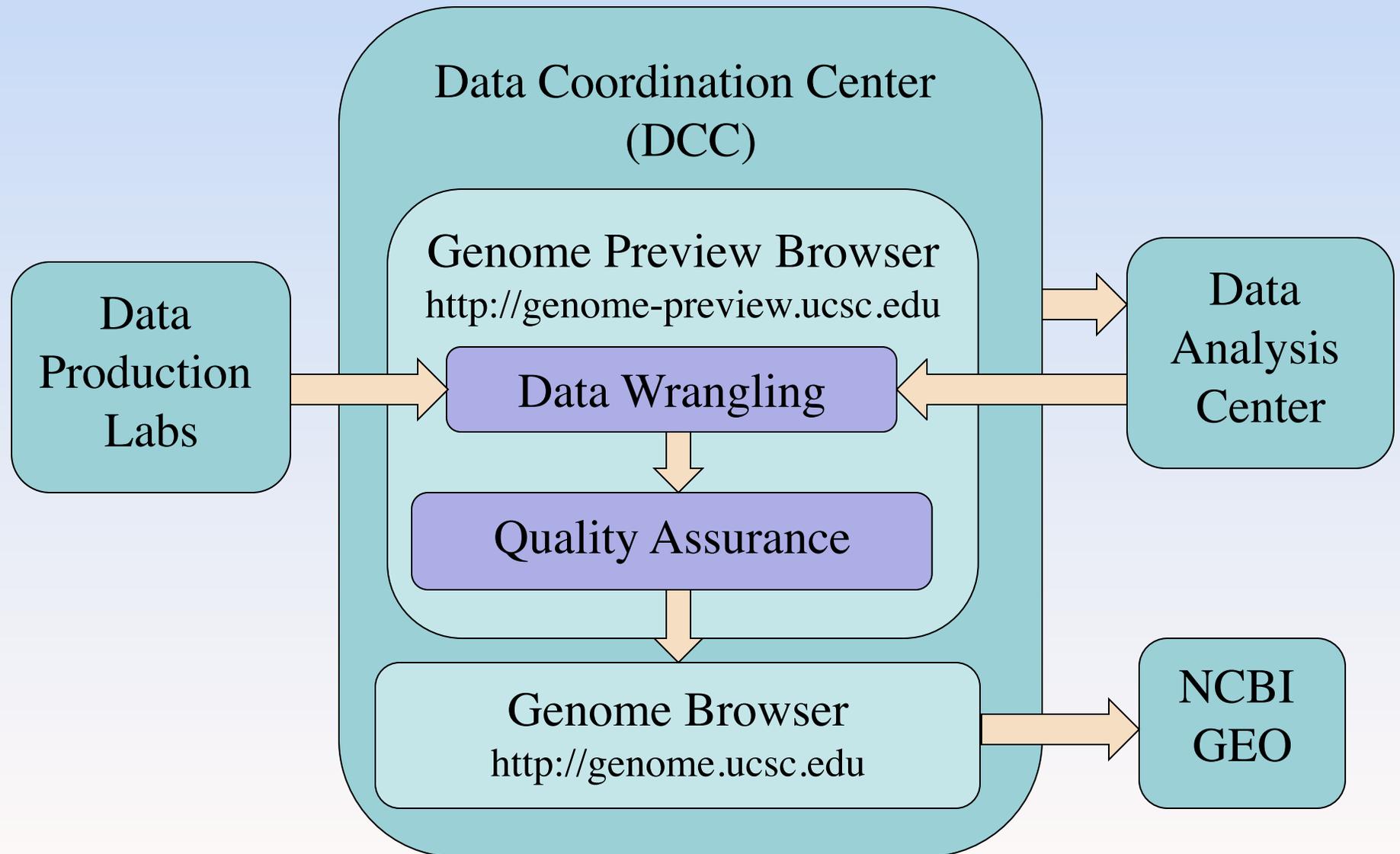
The End

Bonus Slides

ENCODE and related projects



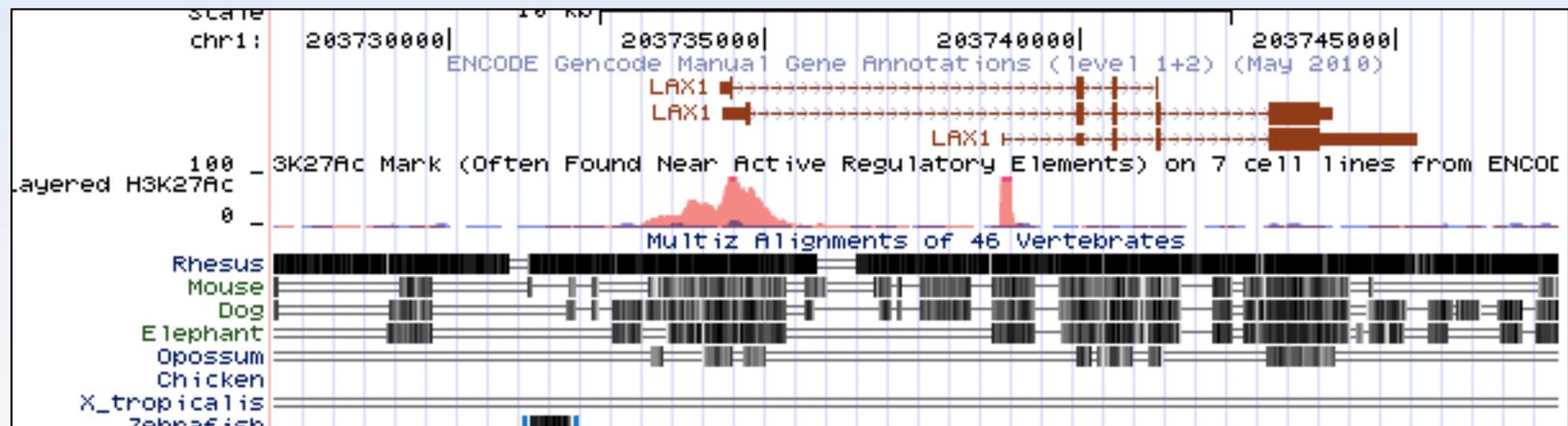
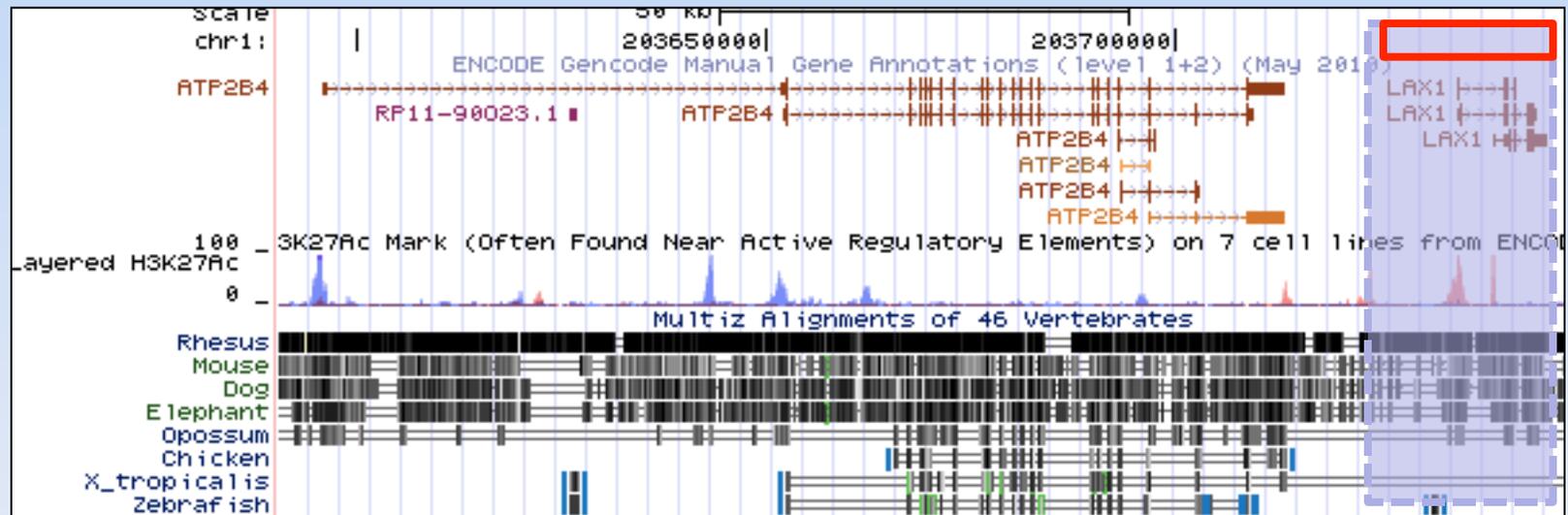
ENCODE Data Lifecycle



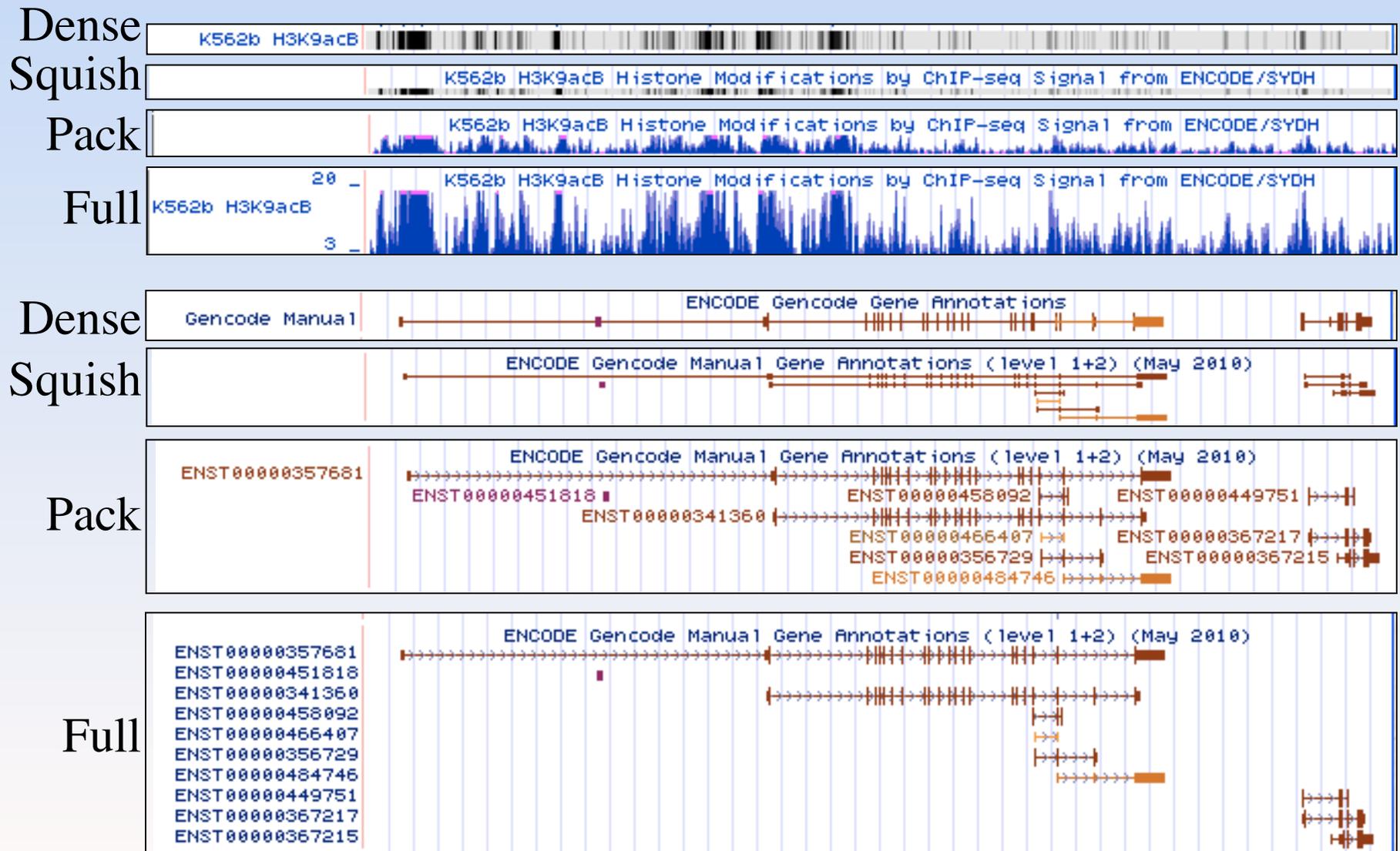
ENCODE Track Organization

<u>Supertracks</u> 8 in hg19	<u>Tracks</u> 45 in hg19	<u>Subtracks</u> 1584 in hg19
ENC RNA-Seq	Caltech RNA-seq	GM12878 - strand-specific, GM12878 - paired, H1hESC - strand-specific, ...
	CSHL Long RNA-seq	GM12878-Nucleus, K562-Nucleoplasm, HepG2-Cytosol, HSMM-WholeCell, ...
	Riken CAGE-Loc	GM12878-Nucleus, K562-Polysome, NHEK-Cytosol, BJ-WholeCell, ...
ENC Histone	Broad Histone	GM12878-CTCF, K562-H3K4me1, HepG2-H3K9ac, HSMM-H2A.Z, ...
	UW Histone	GM12878-H3K4me3, K562-H3K27me3, HUVEC-H3K36me3, BJ-H3K4me3, ...
...

Rectangular region-select



Display settings explained



3. Via NCBI GEO

The screenshot shows the NCBI GEO DataSets search interface. At the top, there are navigation links for 'NCBI', 'Resources', and 'How To'. The search bar contains the text 'encode [project]' and is highlighted with a red box. Below the search bar, the results are displayed in a summary view. The first result is 'GSE29692 record: DNaseI Hypersensitivity by Digital DNaseI from ENCODE/University of Washington [Homo sapiens]'. The summary text describes the ENCODE Project and the use of Digital DNaseI methodology. The page also includes a 'Filter your results' sidebar on the right, showing options for 'All (2386)', 'DataSets (0)', 'Platforms (0)', 'Samples (2330)', and 'Series (56)'. At the bottom, there is a 'Find related data' section with a 'Database' dropdown and a 'Search details' section showing the search query 'encode[project]'.

NCBI Resources How To My NCBI Sign In

GEO DataSets Functional Genomics Studies

Search: GEO DataSets Save search Limits Advanced search Help

encode [project] Search Clear

Display Settings: Summary, 20 per page, Sorted by Default order Send to:

Results: 1 to 20 of 2386 << First < Prev Page 1 of 120 Next > Last >>

1: GSE29692 record: DNaseI Hypersensitivity by Digital DNaseI from ENCODE/University of Washington [Homo sapiens] Links

Summary: (Submitter supplied) This track is produced as part of the ENCODE Project. This track shows DNaseI sensitivity measured genome-wide in different cell lines using the Digital DNaseI methodology (see below), and DNaseI hypersensitive sites. DNaseI has long been used to map general chromatin accessibility and DNaseI hypersensitivity is a universal feature of active cis-regulatory sequences. The use of this method has led to the discovery of functional regulatory elements that include enhancers, insulators, promoters, locus control regions and novel elements. For each experiment (cell type) this track shows DNaseI sensitivity as a continuous function using sequencing tag density (Raw Signal), and discrete loci of DNaseI sensitive zones (HotSpots) and hypersensitive sites (Peaks)." For data usage terms and conditions, please refer to <http://www.genome.gov/27528022> and <http://www.genome.gov/Pages/Research/ENCODE/ENCODEDataReleasePolicyFinal2008.pdf>
Project: ENCODE
1 related Platform

Type: Genome binding/occupancy profiling by high throughput sequencing

Supplementary Files: BAM BIGWIG BROADPEAK NARROWPEAK SRA Experiment SRA Study [download...](#)

Samples: 149

- GSM736493: Stam_HCT-116_2
- GSM736496: Stam_GM12878_1
- GSM736499: Stam_HRGEC_1
- GSM736503: Stam_HPF_2
- GSM736506: Stam_A549_2

Filter your results:

- All (2386)
- DataSets (0)
- Platforms (0)
- [Samples \(2330\)](#)
- [Series \(56\)](#)

[Manage Filters](#)

Top Organisms [Tree]

- Homo sapiens (2378)
- Mus musculus (9)

Find related data

Database: Select

Find items

Search details

encode[project]

Search

www.ncbi.nlm.nih.gov/guide/