

PENNSSTATE HERSHEY



Milton S. Hershey  
Medical Center



# ENCODE workshop

11/12/2016,  
Washington University

Feng Yue, Ph.D.  
Director, Bioinformatics Division  
Institute for Personalized Medicine  
Assistant Professor,  
Dept. of Biochemistry and Molecular Biology  
Pennsylvania State University

# Outline

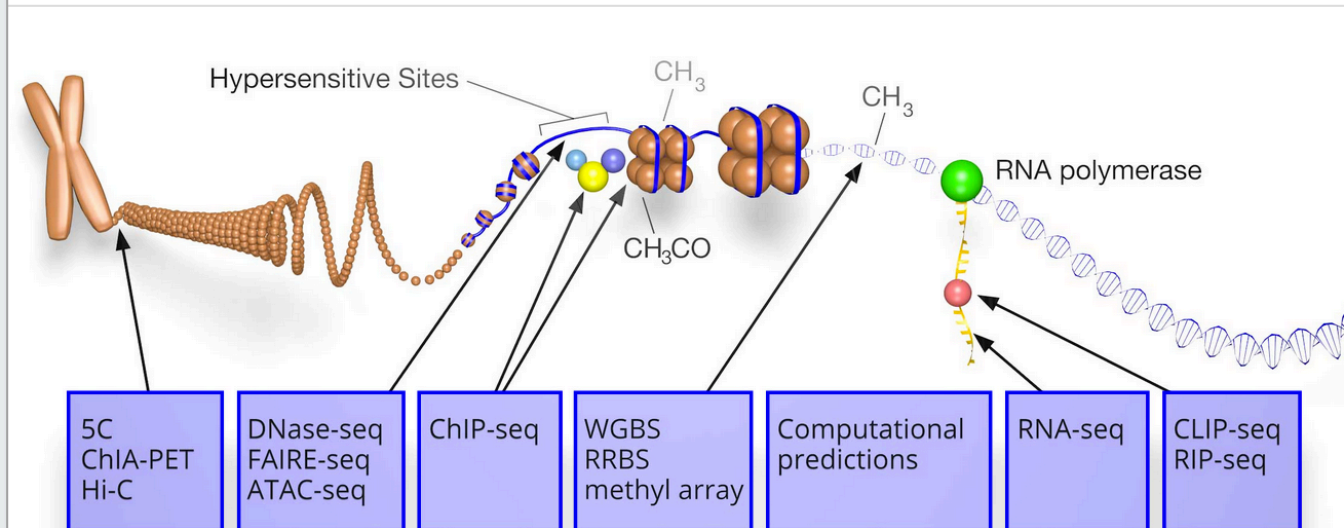


- 1. Search and download through ENCODE portal**
- 2. Search and visualize ENCODE annotations (gene expression, candidate promoters, enhancers ...)**
- 3. Visualize ENCODE chromatin interaction data (Hi-C, ChIA-PET)**
- 4. Annotate genetic variants by HapReg and RegulomeDB.**

[www.encodeproject.org](http://www.encodeproject.org)



# ENCODE: Encyclopedia of DNA Elements



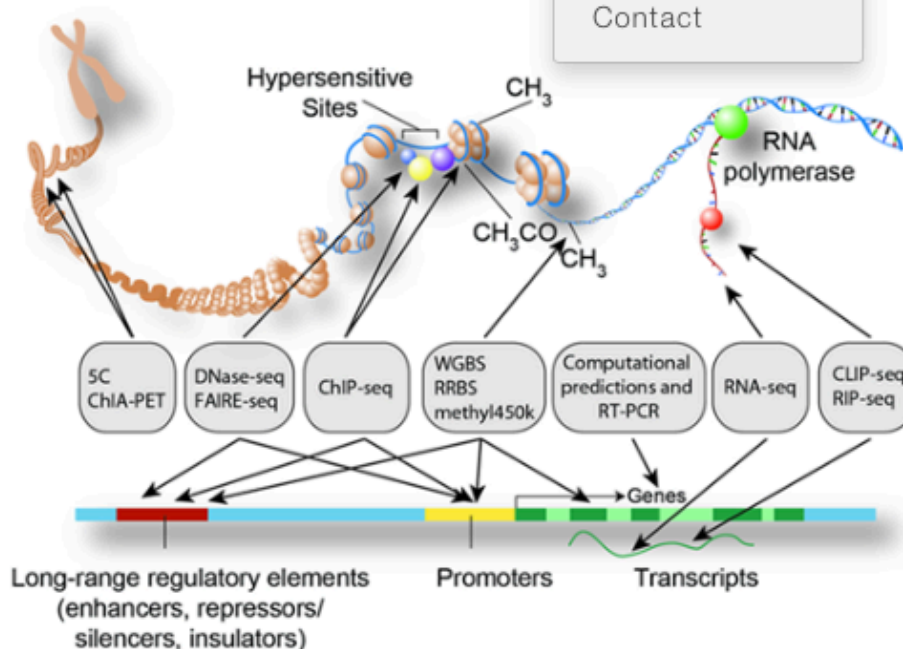
The ENCODE (Encyclopedia of DNA Elements) Consortium is an international collaboration of research groups funded by the National Human Genome Research Institute (NHGRI). The goal of ENCODE is to build a comprehensive parts list of functional elements in the human genome, including elements that

# All the tutorial materials are available on the ENCODE portal

ENCODE Data Methods About Help

## ENCODE: Encyclopedia of DNA Elements

- Getting started
- REST API
- File formats
- Ontologies
- Tutorials**
- Contact



The diagram illustrates the flow of genomic information from DNA to RNA. At the top, a DNA double helix is shown with various regulatory elements: Long-range regulatory elements (enhancers, repressors/silencers, insulators), Promoters, and Transcripts. Below the DNA, several experimental and computational methods are listed in boxes: SC ChIA-PET, DNase-seq FAIRE-seq, ChIP-seq, WGBS RRBS methyl450k, Computational predictions and RT-PCR, RNA-seq, and CLIP-seq RIP-seq. Arrows indicate the relationship between these methods and the genomic features. For example, SC ChIA-PET, DNase-seq FAIRE-seq, ChIP-seq, and WGBS RRBS methyl450k are linked to Long-range regulatory elements and Promoters. Computational predictions and RT-PCR are linked to Promoters. RNA-seq and CLIP-seq RIP-seq are linked to Transcripts. The diagram also shows a 3D model of a DNA loop with Hypersensitive Sites, CH<sub>3</sub>, and CH<sub>3</sub>CO groups, and a RNA polymerase molecule transcribing a gene.

Long-range regulatory elements (enhancers, repressors/silencers, insulators) Promoters Transcripts

RNA polymerase

Genes

SC ChIA-PET DNase-seq FAIRE-seq ChIP-seq WGBS RRBS methyl450k Computational predictions and RT-PCR RNA-seq CLIP-seq RIP-seq

Hypersensitive Sites CH<sub>3</sub> CH<sub>3</sub>CO CH<sub>3</sub>

The ENCODE (Encyclopedia of DNA Elements) is a large-scale collaboration of research groups funded by the National Human Genome Research Institute (NHGRI). The goal of ENCODE is to identify all functional elements in the human genome, including DNA elements, RNA levels, and regulatory elements, to determine when and where a gene is active.

*Image credits: Darryl Leja (NHGRI), Ian Dunham (NHGRI)*

# Tutorials & Workshops

## Upcoming tutorials and workshops

- Please check this page at a later time for information on future workshops.

## Workshop materials

- **NEW: [Asia Pacific Bioinformatics Conference](#), San Francisco, January 10, 2016.**
  - J. Michael Cherry, Introduction to ENCODE and the ENCODE DCC
  - Aditi Narayanan, The ENCODE Portal: Searching for Metadata and Data
  - J. Seth Strattan, ENCODE Data Availability and Standardized Processing
- **[Presentations and Tutorials from the American Society of Human Genetics 2015 Annual Meeting](#) in Baltimore, MD, October 6 - October 10, 2015. ENCODE presented the [Advanced Workshop on Integrative Analysis using ENCODE and Roadmap Epigenomics Data](#).**
- **[Tutorials and video](#) from the [ENCODE 2015: Research Applications and Users Meeting at the Bolger Center](#) in Potomoc, MD, June 29 - July 1, 2015**
  - Video and workshop materials from hands-on tutorial sessions on accessing, processing, analyzing, and utilizing ENCODE data and resources, along with presentations from leading experts in disease, biology, and computational fields explaining how they employ ENCODE resources in their work.

# ENCODE 2016: Research Applications and Users Meeting

On June 08 - June 10, 2015, the National Human Genome Research Institute (NHGRI) will sponsor the **ENCODE 2016: Research Applications and Users Meeting** at the Stanford University in Palo Alto, CA. This page features the some of the material from the upcoming hands-on workshops, and will be updated as more material becomes available.

\*Note: Dr. Freedman requested the ENCODE Consortium to not put his own slides in our tutorial page.

## The meeting will feature:

- Hands-on workshops on learning to navigate, analyze, and integrate ENCODE and mouse ENCODE data into your research.
- Leading-edge research applications from distinguished invited speakers
- Tutorials on newly-available informatics pipelines that greatly facilitate working with ENCODE data
- Short talks selected from abstracts

**Meeting materials:** [Agenda](#) | [Abstracts](#)

# How to search ENCODE datasets -- type “heart CTCF”

The screenshot shows the ENCODE website interface. At the top, there is a navigation bar with 'ENCODE' and menu items: 'Data', 'Methods', 'About', and 'Help'. A search bar on the right contains the text 'heart ctf' and is highlighted with a red box. Below the navigation bar, on the left, is a 'Data Type' filter menu with 'Experiment' selected and highlighted by a red box. The main content area shows 'Showing 4 of 4 results' and a list of search results. Each result includes a title, target, lab, project, and data type.

Title	Target	Lab	Project	Data Type
<b>ChIP-seq of heart (<i>Mus musculus</i>, adult 8 week)</b>	CTCF	Bing Ren, UCSD	ENCODE	Experiment
<b>ChIP-seq of cardiac fibroblast (<i>Homo sapiens</i>)</b>	CTCF	John Stamatoyannopoulos, UW	ENCODE	Experiment
<b>ChIP-seq of cardiac muscle cell (<i>Homo sapiens</i>)</b>	CTCF	John Stamatoyannopoulos, UW	ENCODE	Experiment
<b>Enhancers compete with a long non-coding RNA for regulation of the <i>Kcnq1</i> domain.</b>		Schultz BM, Gallicio GA, Cesaroni M, et al.		Publication

Additional details for the first three results:

- ChIP-seq of heart (*Mus musculus*, adult 8 week):** Target: CTCF, Lab: Bing Ren, UCSD, Project: ENCODE. Data Type: Experiment, ENCSR000CBI released.
- ChIP-seq of cardiac fibroblast (*Homo sapiens*):** Target: CTCF, Lab: John Stamatoyannopoulos, UW, Project: ENCODE. Data Type: Experiment, ENCSR000DTF released.
- ChIP-seq of cardiac muscle cell (*Homo sapiens*):** Target: CTCF, Lab: John Stamatoyannopoulos, UW, Project: ENCODE. Data Type: Experiment, ENCSR000DTI released.

Additional details for the fourth result:

- Enhancers compete with a long non-coding RNA for regulation of the *Kcnq1* domain:** Schultz BM, Gallicio GA, Cesaroni M, et al. *Nucleic acids research*. 2015 Jan;43(2):745-59. PMID:25539921 [PMCID:PMC4333379](#)

# Various filters are available

The screenshot displays a search interface with a left sidebar containing various filters and a main content area on the right. The filters on the left are: Assay category (DNA binding, 1), Assay (ChIP-seq, 1), Project (ENCODE, 1), Experiment status (released, 1), Genome assembly (visualization) (ho19, 2; mm9, 1), Organism (Mus musculus, 1), Target of assay (transcription factor, 1), Biosample type (tissue, 1), and Organ (heart, 1). The 'mm9' filter is highlighted with a red box. The main content area shows 'Showing 1 of 1 results' with a grid icon, followed by the title 'ChIP-seq of heart (Mus musculus, adult 8 week)' and details: Target: CTCF, Lab: Bing Ren, UCSD, and Project: ENCODE.

**Assay category**  
DNA binding 1

**Assay**  
ChIP-seq 1

**Project**  
ENCODE 1

**Experiment status**  
released 1

**Genome assembly (visualization)**  
ho19 2  
mm9 1

**Organism**  
Mus musculus 1

**Target of assay**  
transcription factor 1

**Biosample type**  
tissue 1

**Organ**  
heart 1

Showing 1 of 1 results

**ChIP-seq of heart (*Mus musculus*, adult 8 week)**  
Target: CTCF  
Lab: Bing Ren, UCSD  
Project: ENCODE

# Choose mm9 in genome assembly, then click visualization button

ENCODE Data Methods About Help heart ctf

Showing 1 of 1 results

**Visualize** Download

**ChIP-seq of heart (*Mus musculus*, adult 8 week)** Experiment  
ENCSR000CBI released

Target: CTCF  
Lab: Bing Ren, UCSD  
Project: ENCODE

**Assay category**  
DNA binding 1

**Assay**  
ChIP-seq 1

**Project**  
ENCODE 1

**Experiment status**  
released 1

**Genome assembly (visualization)**  
hg19 0  
**mm9 1**

**Organism**  
*Mus musculus* 1

# Visualizing data in UCSC genome browser

Genomes Genome Browser Tools Mirrors Downloads My Data Help About Us

**Mouse (*Mus musculus*) Genome Browser Gateway**

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.

group	genome	assembly	position	search term
Mammal	Mouse	July 2007 (NCBI37/mm9)	chr12:57,795,963-57,815,592	enter position, gene symbol or search terms

[Click here to reset](#) the browser user interface settings to their defaults.

## Mouse Genome Browser – mm9 assembly ([sequences](#))

The July 2007 mouse (*Mus musculus*) genome data were obtained from the Build 37 assembly by [NCBI](#) and the [Mouse Genome Sequencing Consortium](#). For more information about this assembly, see [MGSCv37](#) in the NCBI Assembly database.

### Sample position queries

A genome position can be specified by the accession number of a sequenced genomic region, an mRNA or EST, a chromosomal coordinate range, or keywords from the GenBank description of an mRNA. The following list shows examples of valid position queries for the Mouse genome. See the [User's Guide](#) for more information.

#### Request:

#### Genome Browser Response:

chr16	Displays all of chromosome 16
chr16:1-5000000	Displays first 5 million bases of chr 16
chr16:1000000+2000	Displays a region of chr 16 that spans 2000 bases, starting with position 1000000
AC072048.4	Displays region of contig AC072048.4
D16Mit120	Displays region around STS marker DMit16120 from the MGI consensus genetic map, including 100,000 bases on each side as well
D16Mit120;D16Mit149	Displays region between STS markers D16Mit120 and D16Mit149
BC012683	Displays alignment location of mRNA with GenBank accession BC012683
AW045217	Displays alignment location of EST with GenBank accession AW045217
Ncam2	Displays region of genome with official MGI mouse genetic nomenclature Ncam2
pseudogene mRNA	Lists transcribed pseudogenes but not cDNAs
zinc finger	Lists many zinc finger mRNAs
kruppel zinc finger	Lists only kruppel-like zinc fingers
huntington	Lists candidate genes associated with Huntington's disease
Smith,D.	Lists mRNAs deposited by co-author D. Smith



*Mus musculus*  
(Photo courtesy of [The Jackson Laboratory](#))

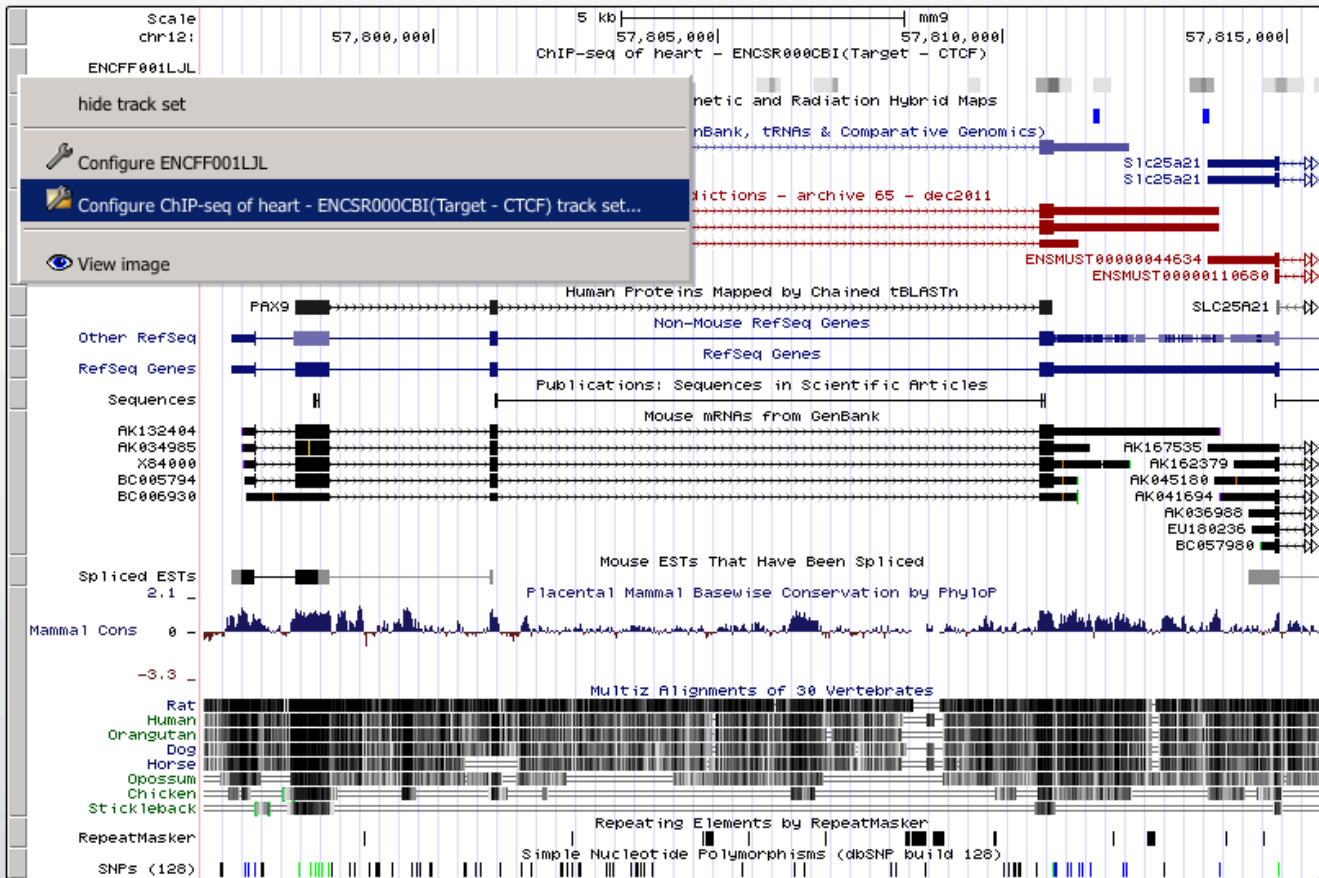
# Visualizing data in UCSC genome browser

UCSC Genome Browser on Mouse July 2007 (NCBI37/mm9) Assembly

move <<< << < > >> >>> ZOOM in 1.5x 3x 10x base ZOOM out 1.5x 3x 10x 100x

chr12:57,795,963-57,815,592 19,630 bp.  go

chr12 (qC1) 12qA1.1 qA2 12qA3 qB1 12qB3 qC1 qC2 12qC3 qD1 D2 12qD3 12qE 12qF1 qF2



move start < 2.0 > move end < 2.0 >  
Click on a feature for details. Click or drag in the base position track to zoom in. Click side bars for track options. Drag side bars or labels up or down to reorder tracks. Drag tracks left or right to new position.

link search default tracks default order hide all add custom tracks track hubs configure multi-region reverse resize refresh

Use drop-down controls below and press refresh to alter tracks displayed

# Visualizing data in UCSC genome browser

Genomes Genome Browser Tools Mirrors Downloads My Data Help About Us

## ChIP-seq of heart - ENCSR000CBI(Target - CTCF) Track Settings

### ChIP-seq of heart - ENCSR000CBI(Target - CTCF)

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

Select views ([help](#)):

Peaks  Signals  full

#### Peaks Configuration

Show only items with score at or above:  (range: 0 to 1000)

#### Signals Configuration

Type of graph:  [Graph configuration help](#)

Track height:  pixels (range: 8 to 100)

Data view scaling:  Always include zero:

Vertical viewing range: min:  max:  (range: 0 to 127)

Transform function: Transform data points by:

Windowing function:  Smoothing window:  pixels

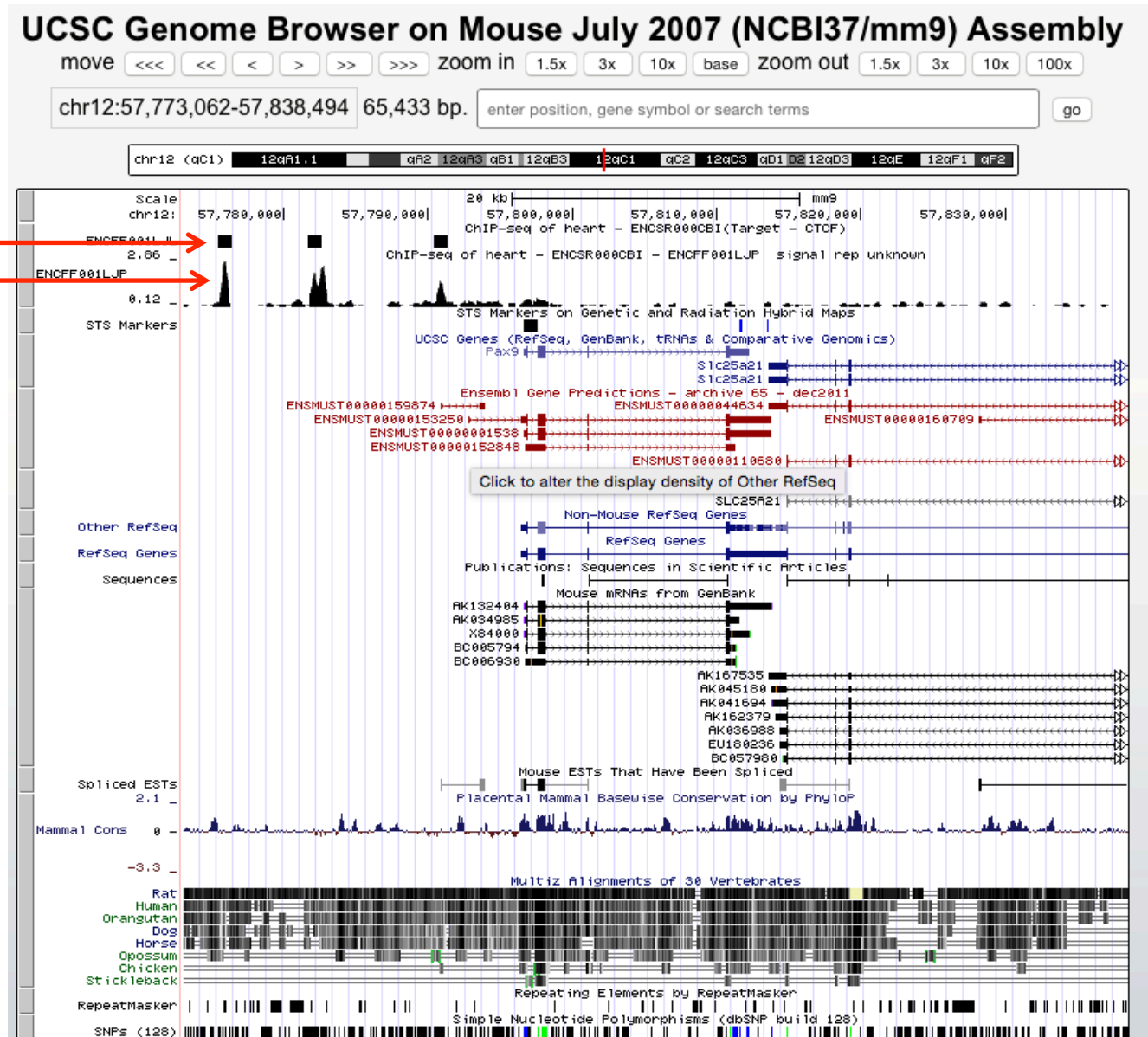
Negate values:

Draw y indicator lines: at y = 0.0:  at y =

List subtracks:  only selected/visible  all

	Views <sup>1</sup>	Track Name <sup>2</sup>
<input checked="" type="checkbox"/>	<input type="text" value="dense"/>	Peaks ChIP-seq of heart - ENCSR000CBI - ENCF001LJL broadPeak peaks rep unknown
<input checked="" type="checkbox"/>	<input type="text" value="full"/>	Signals ChIP-seq of heart - ENCSR000CBI - ENCF001LJP signal rep unknown

# Visualizing data in UCSC genome browser



CTCF binding sites

CTCF raw signals

# Downloading data: Option 1: batch download

ENCODE Data ▾ Methods ▾ About ▾ Help ▾

**Assay category**

DNA binding	1
-------------	---

**Assay**

ChIP-seq	1
----------	---

**Project**

ENCODE	1
--------	---

**Experiment status**

released	1
----------	---

**Genome assembly (visualization)**

hg19	2
mm9	1 <input type="radio"/>

**Organism**

<i>Mus musculus</i>	1
---------------------	---

Showing 1 of 1 results

**Visualize**  **Download**

---

**ChIP-seq of heart (*Mus musculus*, adult 8 week)**

Target: CTCF  
Lab: Bing Ren, UCSD  
Project: ENCODE

Experiment  
ENCSR000CBI  
released

# Option 1: batch download

## -- command-line based

ENCODE Data Methods About Help

ctcf heart

### Using batch download

Click the "Download" button below to download a "files.txt" file that contains a list of URLs to a file containing all the experimental metadata and links to download the file. The first line of the file will always be the URL to download the metadata file. Further description of the contents of the metadata file are described in the [Batch Download help doc](#).

The "files.txt" file can be copied to any server.  
The following command using cURL can be used to download all the files in the list:

```
xargs -n 1 curl -O -L < files.txt
```

Close Download

Assay category  
DNA binding

Assay  
ChIP-seq

Project  
ENCODE

RFA  
ENCODE2-M

Experiment status  
released

Genome assembly (visualization)  
hg19 2  
mm9 1

Organism  
Mus musculus 1

Experiment  
ENCSR000CBI  
released

# Option 2: download individual file

ENCODE Data ▾ Methods ▾ About ▾ Help ▾

**Assay category**  
DNA binding 1



**Assay**  
ChIP-seq 1

**Project**  
ENCODE 1

**Experiment status**  
released 1


**Genome assembly (visualization)**  
hg19 2  
mm9 1

**Organism**  
*Mus musculus* 1

Showing 1 of 1 results  

**ChIP-seq of heart (*Mus musculus*, adult 8 week)** Experiment  
ENCSR000CBI released

Lab: Bing Ren, UCSD  
Project: ENCODE


[Visualize](#)  [Download](#)

# Option 2: download individual file

EXPERIMENTS / [CHIP-SEQ](#) / [MUS MUSCULUS](#) / [HEART](#)

## Experiment summary for ENCSR000CBI






Status: released

Summary	Attribution	 ENCODE
<b>Assay:</b> CHIP-seq	<b>Lab:</b> Bing Ren, UCSD	
<b>Target:</b> <a href="#">CTCF</a>	<b>Award PI:</b> Bing Ren, UCSD	
<b>Biosample summary:</b> heart ( <i>Mus musculus</i> , adult 8 week male)	<b>Project:</b> ENCODE	
<b>Biosample Type:</b> tissue	<b>External resources:</b> <a href="#">UCSC-ENCODE-mm9:wgEncodeEM001684</a> <a href="#">GEO:GSM918756</a>	
<b>Replication type:</b> isogenic	<b>Date released:</b> 2011-07-22	
<b>Description:</b> CTCF ChIP-seq on 8-week mouse heart		
<b>Nucleic acid type:</b> DNA		
<b>Size range:</b> 200-500		
<b>Lysis method:</b> see document		
<b>Extraction method:</b> see document		
<b>Fragmentation method:</b> see document		
<b>Size selection method:</b> see document		
<b>Platform:</b> <a href="#">Illumina Genome Analyzer II</a>		
<b>Controls:</b> <a href="#">ENCSR000CAV</a>		

# Option 2: download individual file

File summary										Visualize Data ▾
Raw data files										
Accession	File type	Biological replicate	Library	Run type	Read	Lab	Date added	File size		
ENCF001LJW <a href="#">↓</a>	fastq	1	ENCLB676HUM	SE 36nt		Bing Ren, UCSD	2012-02-14	534 MB		
ENCF001LJV <a href="#">↓</a>	fastq	2	ENCLB858LXJ	SE 36nt		Bing Ren, UCSD	2012-02-14	694 MB		
Processed data files										
Accession	File type	Output type	Biological replicate	Mapping assembly	Lab	Date added	File size			
ENCF001LJK <a href="#">↓</a>	bam	alignments	1	mm9	Bing Ren, UCSD	2012-02-14	358 MB			
ENCF001LJM <a href="#">↓</a>	bam	alignments	2	mm9	Bing Ren, UCSD	2012-02-14	335 MB			
ENCF001LJL <a href="#">↓</a>	bigBed broadPeak	peaks		mm9	Bing Ren, UCSD	2012-02-14	741 kB			
ENCF001LJP <a href="#">↓</a>	bigWig	signal		mm9	Bing Ren, UCSD	2012-02-14	62.4 MB			
ENCF001YAF <a href="#">↓</a>	bed broadPeak	peaks		mm9	Bing Ren, UCSD	2012-02-14	444 kB			


# Experimental protocols are available too!



Processed data files									
Accession	File type	Output type	Biological replicate	Mapping assembly	Lab	Date added	File size		
ENCFF001LJK 	bam	alignments	1	mm9	Bing Ren, UCSD	2012-02-14	358 MB		
ENCFF001LJM 	bam	alignments	2	mm9	Bing Ren, UCSD	2012-02-14	335 MB		
ENCFF001LJL 	bigBed broadPeak	peaks		mm9	Bing Ren, UCSD	2012-02-14	741 kB		
ENCFF001LJP 	bigWig	signal		mm9	Bing Ren, UCSD	2012-02-14	62.4 MB		
ENCFF001YAF 	bed broadPeak	peaks		mm9	Bing Ren, UCSD	2012-02-14	444 kB		

### Documents

**General protocol**


**Description excerpt:** The library preparation protocol used in the CHIP-seq by the Ren lab during...





 [Ren\\_CHIP\\_Library\\_Preparation\\_v090113.pdf](#) 

**General protocol**


**Description excerpt:** The general chromatin immunoprecipitation protocol used by the Ren lab during...





 [Ren\\_Chip\\_Protocol\\_v090113.pdf](#) 

**General protocol**

**Description:** Mouse heart dissection and CHIP-seq preparation protocol



 [Heart\\_Ren\\_protocol.pdf](#) 



# Assignment



1. Search H3K4me3 in human heart on the ENCODE portal, and display the tracks on UCSC genome browser.

# Outline for the workshop

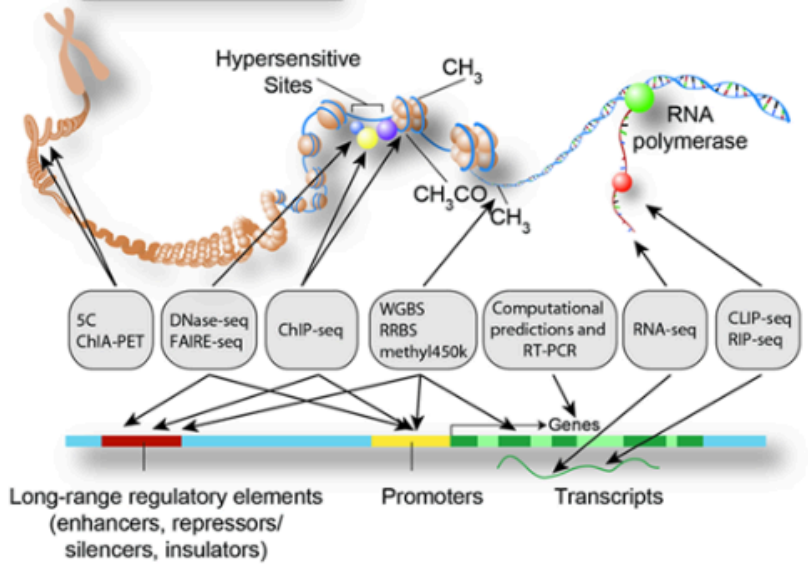


1. Search and download through ENCODE portal
2. Search and visualize ENCODE annotations (gene expression, promoters, enhancers ...)
3. Visualize ENCODE chromatin interaction data (Hi-C, ChIA-PET).
4. Search and Visualize Roadmap Epigenomics project data through WashU Epigenome Browser.

# ENCODE annotations

- Assays
- Biosamples
- Antibodies
- Annotations
- Release policy

## ENCODE Encyclopedia of DNA Elements



The ENCODE (Encyclopedia of DNA Elements) 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.

Image credits: Darryl Leja (NHGRI), Ian Dunham (EBI), Michael Pazin (NHGRI)

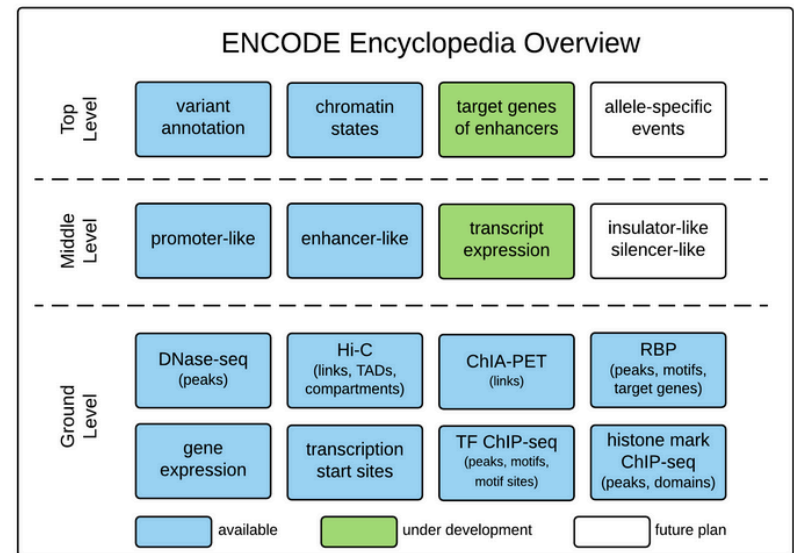
# ENCODE element query website

## ENCODE Encyclopedia, Version 3: Genomic annotations

### Introduction

The ENCODE Consortium not only produces data, but also analyzes the data in an integrative fashion. The ENCODE Encyclopedia organizes the most salient analysis products into annotations, and provides tools to search and visualize them. The Encyclopedia has three levels of annotations:

- Ground level annotations are typically derived directly from the experimental data.
- Middle level annotations integrate multiple types of experimental data and multiple ground level annotations.
- Top level annotations integrate a broad range of experimental data and ground and middle level annotations.



# Query gene expression and regulatory elements

## Ground Level Annotations

### Gene expression (RNA-seq)

The expression levels of genes annotated by GENCODE 19 in over 100 human cell types and 70 mouse cell types.

[ [Long RNA-seq Data](#) | [Query](#) | [Download](#) | [Method](#) ]

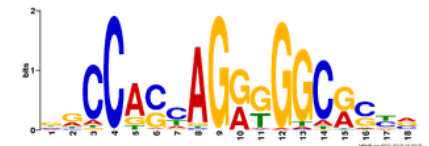


### Transcription factor binding (TF ChIP-seq)

Peaks (enriched genomic regions) of TFs computed from ~900 human and mouse ChIP-seq experiments.

[ [Raw Data](#) | [Peaks](#) ]

Visualize sequence motifs and other information [ [Factorbook](#) [↗](#) ]



CTCF Motif from [Factorbook](#) [↗](#)

### Histone mark enrichment (ChIP-seq)

Peaks of a variety of histone marks computed from ~600 ChIP-seq experiments.

[ [Raw Data](#) | [Peaks](#) ]



H3K27ac from e11.5 Neural Tube

# Click human tab, and search for the expression for Sox2



Welcome to YUE Lab  
Computational and Functional Genomics/Epigenomics

HOME

MOUSE

HUMAN

DOWNLOAD

LINKS

CONTACT

## Query human ENCODE data!

### Option 1: Search gene expression across ~ 60 human cell types (total 108 datasets)

Human (hg19) ▾

Gene name(Sox2, Nanog ...)

submit!

### Option 2: Search cis-elements in a given genomic region

human (hg19) ▾

chr1 ▾

start:

end:

submit!

### Option 3: search cis-elements surrounding a gene

Human (hg19) ▾

Gene name(Sox2, Nanog ...)

Extended region (default +/- 100kb)

kb

submit!

### Option 4: search cis-elements LINKED to a gene based on DNaseI HSS specificity

Human (hg19) ▾

Gene name(Sox2, Nanog ...)

submit!

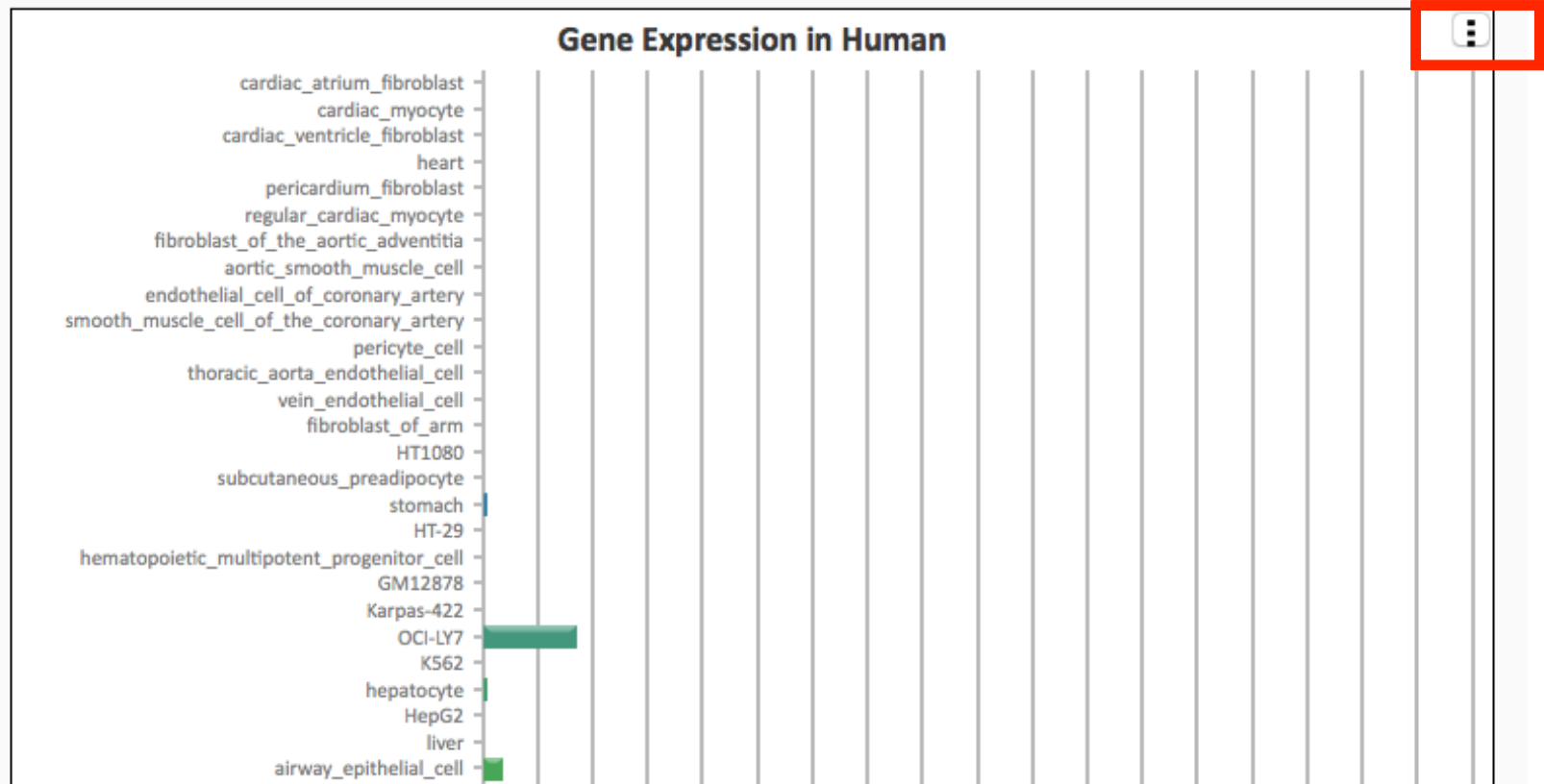
# Visualize and save the gene expression profile

## Gene Expression Across Multiple Tissues/Cell Lines.

Human (hg19)

Gene **SOX2** [NM\_003106, ENSG00000181449, ENST00000325404]

*Click on the upper left button on the chart to save it as an image.*



# You can choose the cell types to display!

Location:  Cell  Cytosol  Nucleus

Method:  PolyA<sup>+</sup> RNA  Total RNA

Update Graph

Export to CSV

Select the values to display:	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>	Select All <input checked="" type="checkbox"/>
Tissue	TPM1	TPM2	Average	Display
cardiac atrium fibroblast.ENCSR110BDY.cell.longRNA total	0	0	0.000	<input checked="" type="checkbox"/>
cardiac myocyte.ENCSR379YAE.cell.longRNA total	0.01	0.01	0.010	<input checked="" type="checkbox"/>
cardiac ventricle fibroblast.ENCSR369RVN.cell.longRNA total	0	0	0.000	<input checked="" type="checkbox"/>
heart.ENCSR000AEZ.cell.longRNA total	0.03	0.03	0.030	<input checked="" type="checkbox"/>
heart.ENCSR000AHH.cell.longRNA total	0.1	0.1	0.100	<input type="checkbox"/>
pericardium fibroblast.ENCSR362HMX.cell.longRNA total	0	0	0.000	<input checked="" type="checkbox"/>
regular cardiac myocyte.ENCSR000AAH.cell.longRNA total	0	0	0.000	<input checked="" type="checkbox"/>
fibroblast of the aortic adventitia.ENCSR000CUJ.cell.longRNA total	0	0	0.000	<input checked="" type="checkbox"/>
aortic smooth muscle cell.ENCSR000AAA.cell.longRNA total	0	0	0.000	<input checked="" type="checkbox"/>
endothelial cell of coronary artery.ENCSR000AAF.cell.longRNA total	0	0	0.000	<input checked="" type="checkbox"/>
smooth muscle cell of the coronary artery.ENCSR000AAG.cell.longRNA total	0	0	0.000	<input checked="" type="checkbox"/>
pericyte cell.ENCSR000CTX.cell.longRNA total	0	0	0.000	<input checked="" type="checkbox"/>
thoracic aorta endothelial cell.ENCSR000CUK.cell.longRNA total	0	0	0.000	<input checked="" type="checkbox"/>
vein endothelial cell.ENCSR000CUG.cell.longRNA total	0	0	0.000	<input checked="" type="checkbox"/>
fibroblast of arm.ENCSR797BPP.cell.longRNA total	0	0	0.000	<input checked="" type="checkbox"/>
HT1080.ENCSR535VTR.cell.longRNA total	0	0	0.000	<input checked="" type="checkbox"/>

# You can choose the cell types to display!



Welcome to YUE Lab  
Computational and Functional Genomics/Epigenomics



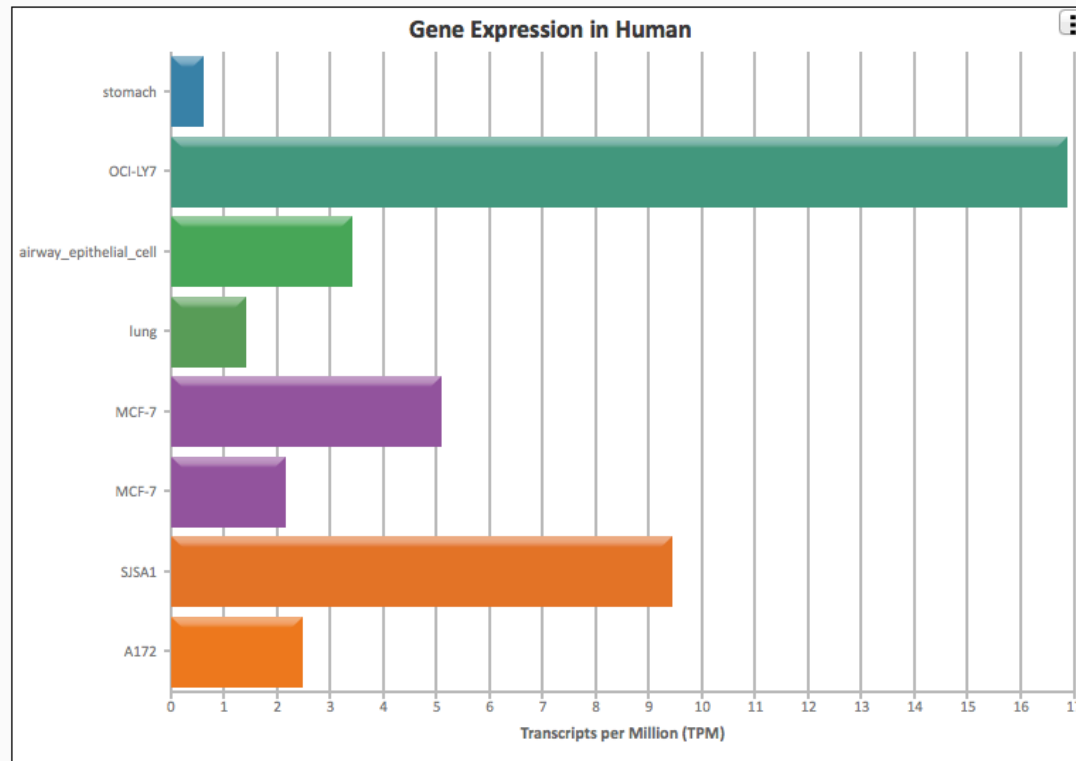
[ABOUT](#) [MOUSE](#) [HUMAN](#) [DOWNLOAD](#) [LINKS](#) [CONTACT](#)

## Gene Expression Across Multiple Tissues/Cell Lines.

Human (hg19)

Gene **SOX2** [NM\_003106, ENSG00000181449, ENST00000325404]

*Click on the upper left button on the chart to save it as an image.*





# Welcome to YUE Lab

## Computational and Functional Genomics/Epigenomics

[HOME](#) [MOUSE](#) [HUMAN](#) [DOWNLOAD](#) [LINKS](#) [CONTACT](#)

### Query human ENCODE data!

#### Option 1: Search gene expression across ~ 60 human cell types (total 108 datasets)

Human (hg19) Gene name(Sox2, Nanog ...)

#### Option 2: Search cis-elements in a given genomic region

human (hg19) chr3 start:  end:

#### Option 3: search cis-elements surrounding a gene

Human (hg19) Gene name(Sox2, Nanog ...)

Extended region (default +/- 100kb)  kb

#### Option 4: search cis-elements LINKED to a gene based on DNaseI HSS specificity

Human (hg19) Gene name(Sox2, Nanog ...)



# Welcome to YUE Lab

## Computational and Functional Genomics/Epigenomics

[HOME](#)[MOUSE](#)[HUMAN](#)[DOWNLOAD](#)[LINKS](#)[CONTACT](#)

### Candidate cis-elements in your queried region.

Species: human\_hg19 chr3:181000000-181200000

#### DNaseI Hypersensitive Sites:

Coordinate	Tissue/cell type
chr3:181010880-181011030	NH-A,HSMM,HSMMtube,HUVEC,NHLF,NHEK,HMEC,GM12878
chr3:181011740-181011890	HMEC
chr3:181016320-181016470	HMEC
chr3:181018360-181018510	NH-A
chr3:181034320-181034470	NH-A,HUVEC,HeLa-S3,NHLF,HSMMtube,HMEC,NHEK
chr3:181042360-181042510	H1-hESC
chr3:181042880-181043030	H1-hESC
chr3:181044545-181044695	H1-hESC,HSMM,HUVEC,HepG2,NHEK,GM12878,HeLa-S3,HSMMtube,NHLF,HMEC
chr3:181045285-181045435	H1-hESC
chr3:181048640-181048790	NH-A,NHLF



# Welcome to YUE Lab

Computational and Functional Genomics/Epigenomics

- HOME
- MOUSE
- HUMAN
- DOWNLOAD
- LINKS
- CONTACT

## Query human ENCODE data!

### Option 1: Search gene expression across ~ 60 human cell types (total 108 datasets)

Human (hg19) Gene name(Sox2, Nanog ...)  submit!

### Option 2: Search cis-elements in a given genomic region

human (hg19) chr1 start:  end:  submit!

### Option 3: search cis-elements surrounding a gene

Human (hg19) Gene name(Sox2, Nanog ...)  ←

Extended region (default +/- 100kb)  ← kb submit!

### Option 4: search cis-elements LINKED to a gene based on DNaseI HSS specificity

Human (hg19) Gene name(Sox2, Nanog ...)  submit!



# Welcome to YUE Lab

## Computational and Functional Genomics/Epigenomics

<a href="#">ABOUT</a>	<a href="#">MOUSE</a>	<a href="#">HUMAN</a>	<a href="#">DOWNLOAD</a>	<a href="#">LINKS</a>	<a href="#">CONTACT</a>
-----------------------	-----------------------	-----------------------	--------------------------	-----------------------	-------------------------

### Candidate cis-elements in your queried region.

Species: human\_hg19

Gene name: sox2

Gene coordinate: chr3:181429711-181432223

Padding: +/-20000 bp.

### DNaseI Hypersensitive Sites:

Coordinate	Tissue/cell type
chr3:181411285-181411435	HMEC
chr3:181413225-181413375	HeLa-S3,H1-hESC,HUVEC,NHLF,HSMMtube,NH-A,NHEK,K562,HepG2,HMEC,HSMM
chr3:181414005-181414155	HSMM,HSMMtube,NH-A,HUVEC,NHEK,Osteobl,NHLF
chr3:181416105-181416255	H1-hESC,NHEK,HMEC
chr3:181417425-181417575	HUVEC,NH-A,NHLF,H1-hESC,NHEK,Osteobl
chr3:181418320-181418470	HUVEC,K562,H1-hESC,HSMMtube,GM12878,HepG2,HSMM,HMEC,NH-A,NHDF-Ad,HeLa-S3,Osteobl,NHLF,NHEK
chr3:181419345-181419495	HUVEC

**TF binding Site:**

Coordinate	TF	Bindings in tissues
chr3:181413225-181413375	multiple	BACH1(H1-hESC), CHD1(H1-hESC), CTBP2(H1-hESC), CTCF(AG09319), CTCF(AoAF), CTCF(Fibrobl), CTCF(GM12874), CTCF(GM12892), CTCF(GM19238), CTCF(GM19239), CTCF(GM19240), CTCF(Gliobla), CTCF(H1-hESC), CTCF(HA-sp), CTCF(HAc), CTCF(HBMEC), CTCF(HCM), CTCF(HMEC), CTCF(HRPEpiC), CTCF(HUVEC), CTCF(K562), CTCF(NHEK), CTCF(NHLF), CTCF(SAEC), E2F6(K562), EZH2(NHDF-Ad), HDAC2(H1-hESC), MAX(K562), POLR2A(HUVEC), RAD21(H1-hESC), RBBP5(H1-hESC), ZBTB7A(K562)
chr3:181414005-181414155	EZH2	EZH2(H1-hESC)
chr3:181416105-181416255	BATF	BATF(GM12878)
chr3:181417425-181417575	JUND, SP1, TBP	JUND(H1-hESC), SP1(H1-hESC), TBP(H1-hESC)
chr3:181418320-181418470	multiple	CEBPB(A549), CEBPB(HeLa-S3), CEBPB(K562), CTCF(A549), CTCF(AG04449), CTCF(AG04450), CTCF(AG09309), CTCF(AG09319), CTCF(AG10803), CTCF(AoAF), CTCF(BE2_C), CTCF(BJ), CTCF(Caco-2), CTCF(Dnd41), CTCF(ECC-1), CTCF(Fibrobl), CTCF(GM06990), CTCF(GM12864), CTCF(GM12865), CTCF(GM12872), CTCF(GM12873), CTCF(GM12874), CTCF(GM12875), CTCF(GM12878), CTCF(GM12891), CTCF(GM12892), CTCF(GM19238), CTCF(GM19239), CTCF(GM19240), CTCF(Gliobla), CTCF(H1-hESC), CTCF(HA-sp), CTCF(HAc), CTCF(HBMEC), CTCF(HCFaa), CTCF(HCM), CTCF(HCPEpiC), CTCF(HCT-116), CTCF(HEEpiC), CTCF(HEK293), CTCF(HFF), CTCF(HFF-Myc), CTCF(HL-60), CTCF(HMEC), CTCF(HMF), CTCF(HPAF), CTCF(HPF), CTCF(HRE), CTCF(HRPEpiC). CTCF(HSMM). CTCF(HSMMtube).



# Welcome to YUE Lab

Computational and Functional Genomics/Epigenomics

- HOME
- MOUSE
- HUMAN
- DOWNLOAD
- LINKS
- CONTACT

## Query human ENCODE data!

### Option 1: Search gene expression across ~ 60 human cell types (total 108 datasets)

Human (hg19) Gene name(Sox2, Nanog ...)  submit!

### Option 2: Search cis-elements in a given genomic region

human (hg19) chr1 start:  end:  submit!

### Option 3: search cis-elements surrounding a gene

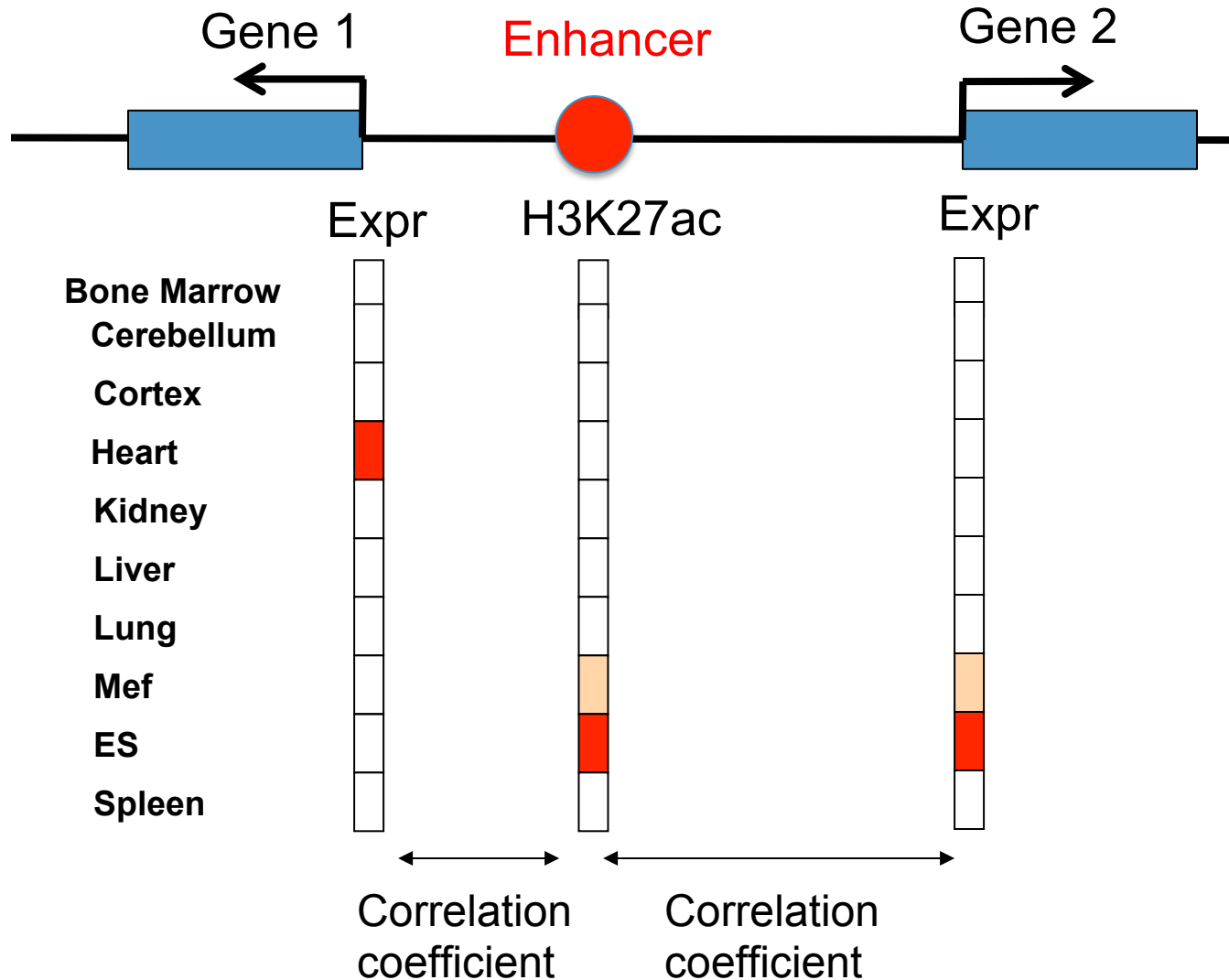
Human (hg19) Gene name(Sox2, Nanog ...)

Extended region (default +/- 100kb)  kb submit!

### Option 4: search cis-elements LINKED to a gene based on DNaseI HSS specificity

Human (hg19) Gene name(Sox2, Nanog ...) SOX2 submit!

# Linking enhancers to target promoters





# Welcome to YUE Lab

## Computational and Functional Genomics/Epigenomics

[ABOUT](#)[MOUSE](#)[HUMAN](#)[LINKS](#)[CONTACT](#)

### Cis-elements linked to your queried gene.

Species: human\_hg19

Gene name: SOX2

### Cis-element lined by DNaseI Hypersensitive Sites Linkage:

Proximal DHS (TSS)	start	end	Gene	Distal DHS	start	end	correlation
chr3	181429540	181429690	SOX2	chr3	181051260	181051410	0.739482
chr3	181429540	181429690	SOX2	chr3	181422060	181422210	0.74818
chr3	181429540	181429690	SOX2	chr3	181430060	181430210	0.757448
chr3	181429540	181429690	SOX2	chr3	181430620	181430770	0.735273
chr3	181429540	181429690	SOX2	chr3	181444460	181444610	0.900111
chr3	181429540	181429690	SOX2	chr3	181634020	181634170	0.761618
chr3	181429540	181429690	SOX2	chr3	181753200	181753350	0.74952
chr3	181429540	181429690	SOX2	chr3	181902380	181902530	0.802853

## Cis-elements linked to your queried gene.

Species: mouse\_mm9

Gene name: Sox2

### Cis-element lined by DNaseI Hypersensitive Sites Linkage:

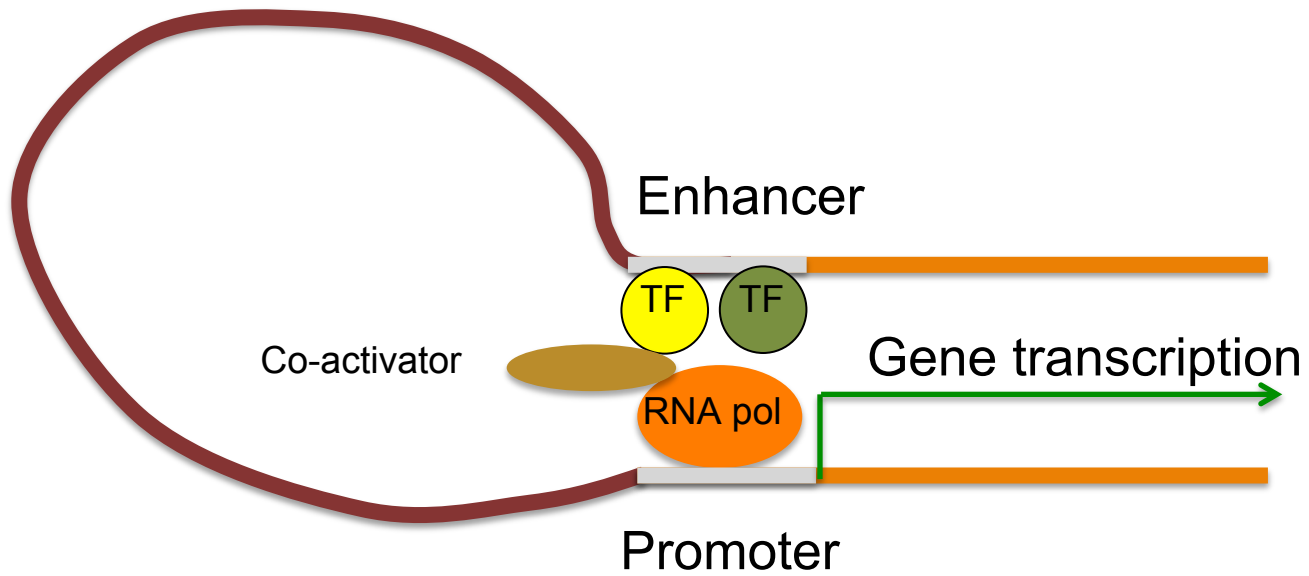
Proximal DHS(TSS)	start	end	Gene	Distal DHS	start	end	correlation
chr3	34548800	34548950	Sox2	chr3	34049080	34049230	0.746471
chr3	34548800	34548950	Sox2	chr3	34051020	34051170	0.751643
chr3	34548800	34548950	Sox2	chr3	34096080	34096230	0.851654
chr3	34548800	34548950	Sox2	chr3	34097280	34097430	0.700074
chr3	34548800	34548950	Sox2	chr3	34130700	34130850	0.756613
chr3	34548800	34548950	Sox2	chr3	34133200	34133350	0.815104
chr3	34548800	34548950	Sox2	chr3	34133520	34133670	0.717251
chr3	34548800	34548950	Sox2	chr3	34140140	34140290	0.802445
chr3	34548800	34548950	Sox2	chr3	34192360	34192510	0.713653
chr3	34548800	34548950	Sox2	chr3	34221860	34222010	0.737236
chr3	34548800	34548950	Sox2	chr3	34224720	34224870	0.787703
chr3	34548800	34548950	Sox2	chr3	34227900	34228050	0.723365

# Outline for the workshop

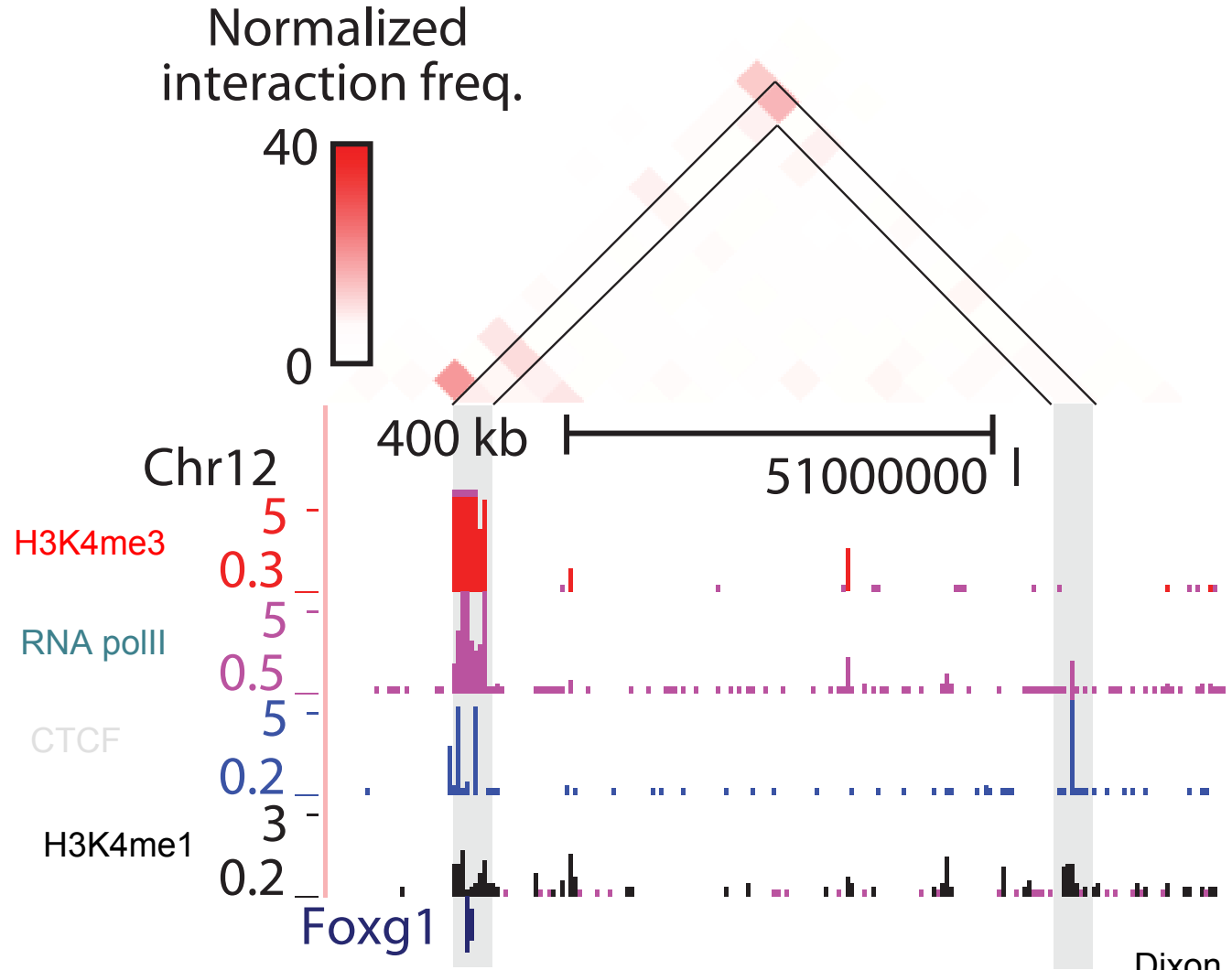


1. Search and download through ENCODE portal
2. Search and visualize ENCODE annotations (gene expression, promoters, enhancers ...)
3. Visualize ENCODE chromatin interaction data (Hi-C, ChIA-PET).
4. Search and Visualize Roadmap Epigenomics project data through WashU Epigenome Browser.

# 3dgenome.org



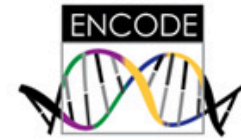
# Hi-C measures genome-wide chromatin interaction



# http://3dgenome.org



Welcome to YUE Lab  
Computational and Functional Genomics/Epigenomics



- HOME
- HI-C INTERACTIONS
- VIRTUAL 4C
- DOWNLOAD
- TUTORIAL
- CITATION
- CONTACT

## Welcome to the Interactive Hi-C Data Browser!

Accessing Hi-C Data

**Step 1. Choose the assembly of Hi-C data:**  
Species  Assembly

**Step 2. Choose the source of the data:**  
 Browse Available Hi-C Data  
Tissue  Resolution

OR  
 Use Your Own Data

Data URL

**Step 3. Choose the region to show Hi-C interactions:**

Option 1: Search by Gene Name	Option 2: Search by Location
Gene/Genomic Feature <input type="text"/> <input type="button" value="Show Interaction"/>	Chromosome <input type="text" value="chr1"/> Start <input type="text"/> End <input type="text"/> <input type="button" value="Show Interaction"/>

# Users can query by gene name or genomic loci



Welcome to YUE Lab  
Computational and Functional Genomics/Epigenomics

- HOME
- HI-C INTERACTIONS
- VIRTUAL 4C
- DOWNLOAD
- TUTORIAL
- CITATION
- CONTACT

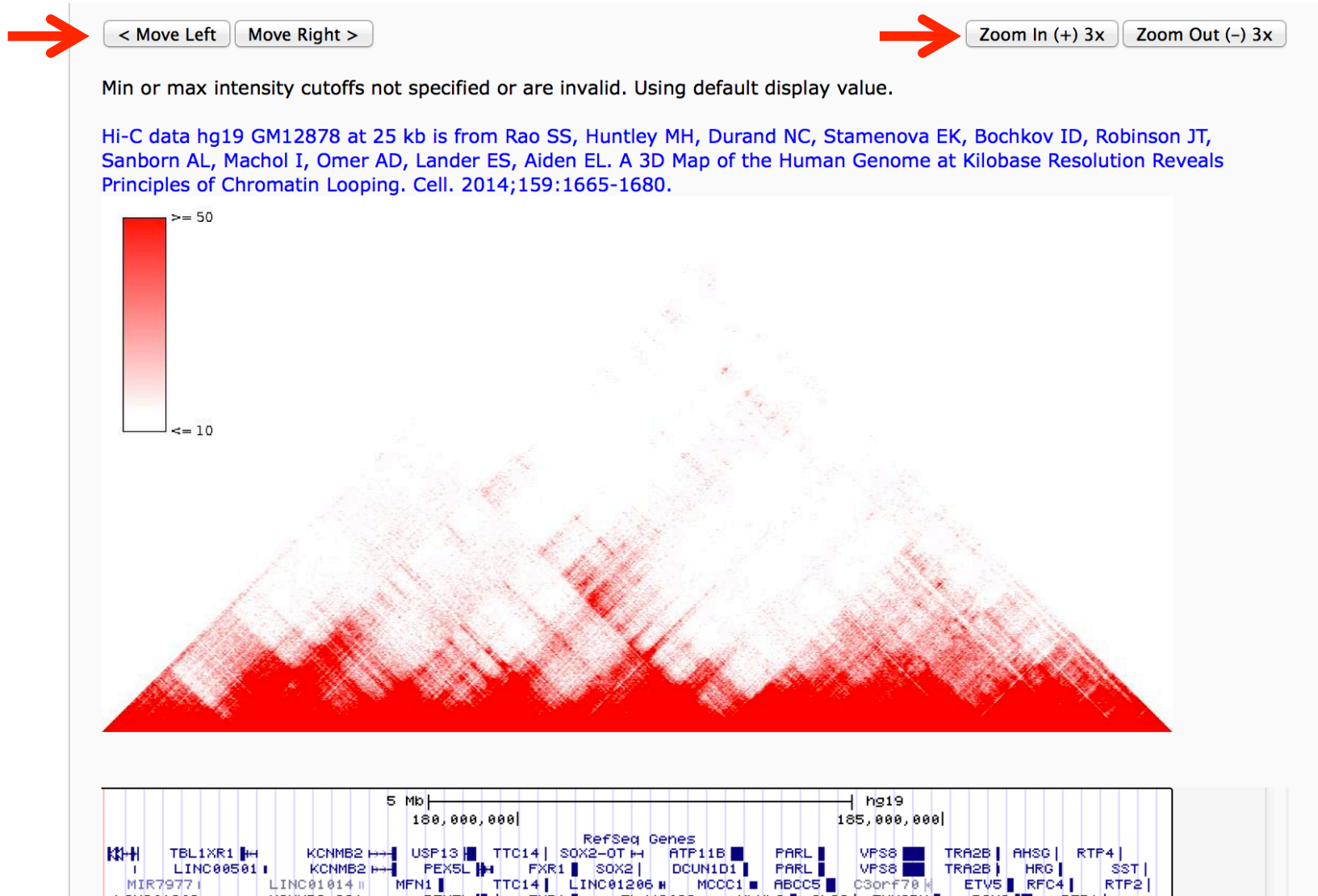
## Welcome to Hi-C Data Browser!

Choose the regions to show Hi-C interactions:

Species  Assembly  Tissue  Resolution

Option 1: Search by Gene Name	Option 2: Search by Location
<p>Gene name (Sox2, Nanog ...)</p> <input type="text" value="sox2"/> <ul style="list-style-type: none"><li>SOX2</li><li>SOX2-OT</li><li>SOX20</li><li>SOX21</li><li>SOX22</li></ul>	<p>Chromosome <input type="text" value="chr1"/></p> <p>Start <input type="text"/></p> <p>End <input type="text"/></p> <p><input type="button" value="Show Interaction"/></p>
<p>Offsets Min <input type="text"/></p>	<p>Max <input type="text"/></p>

# Navigation bars to explore Hi-C data



< Move Left

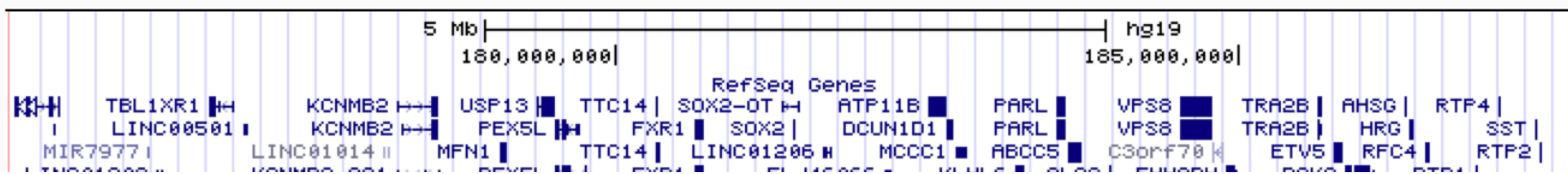
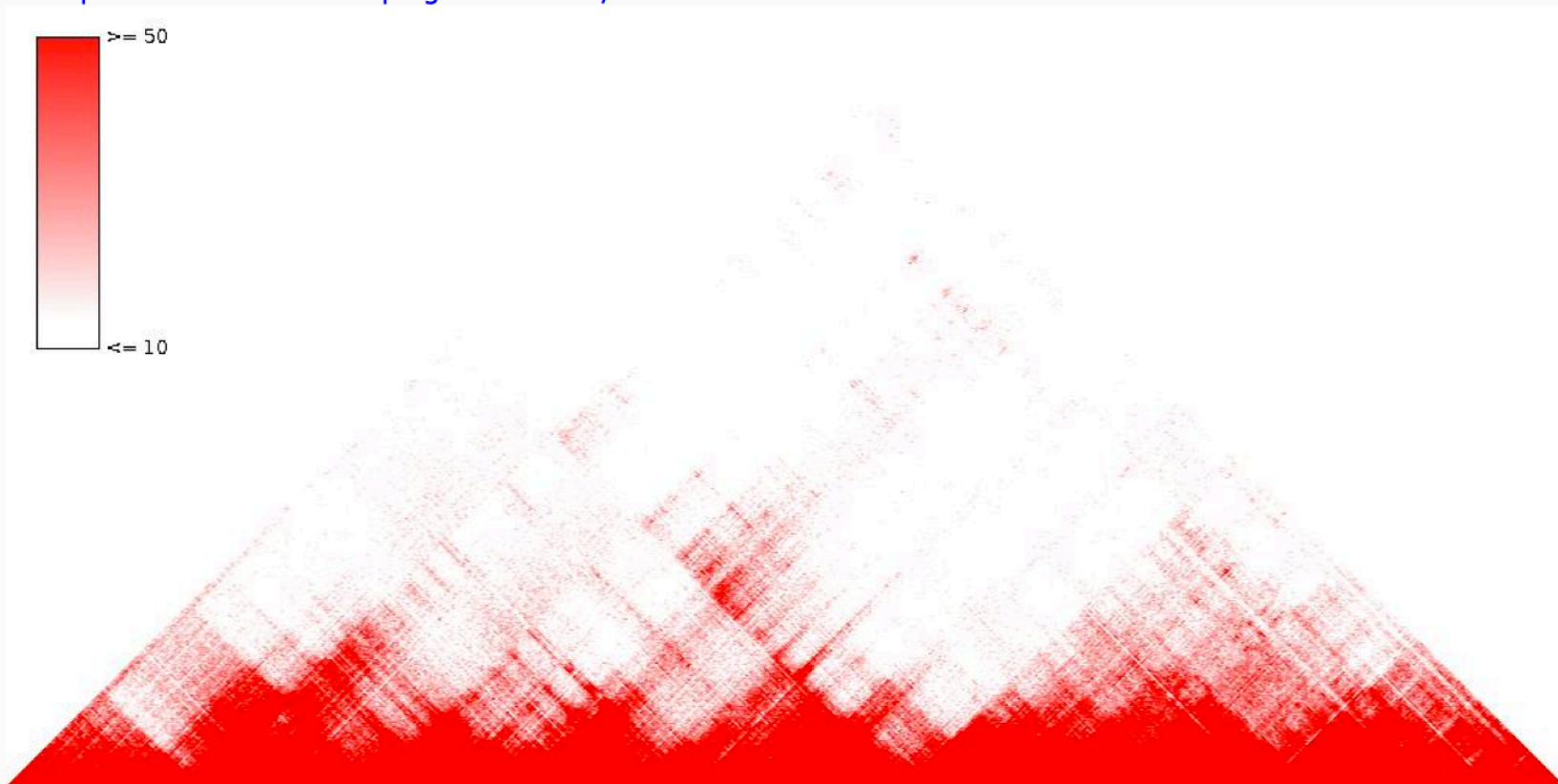
Move Right >

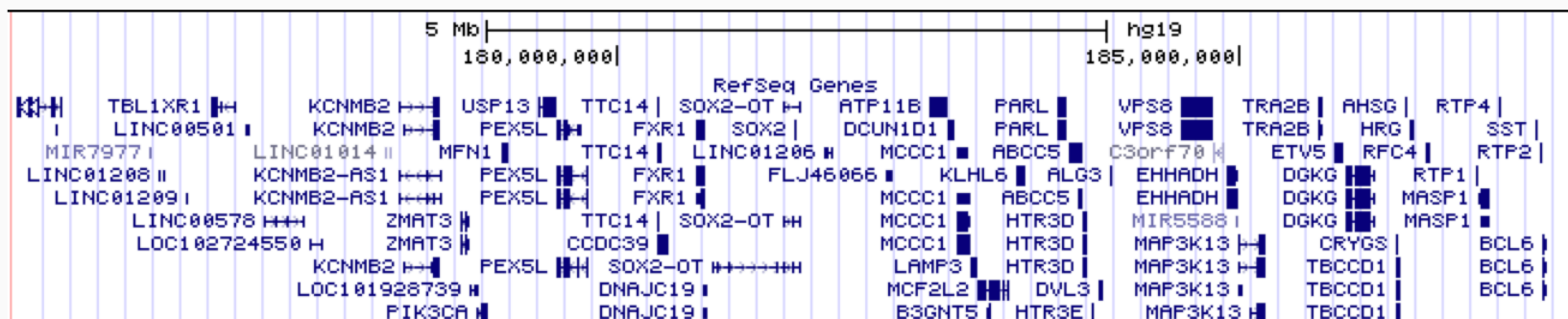
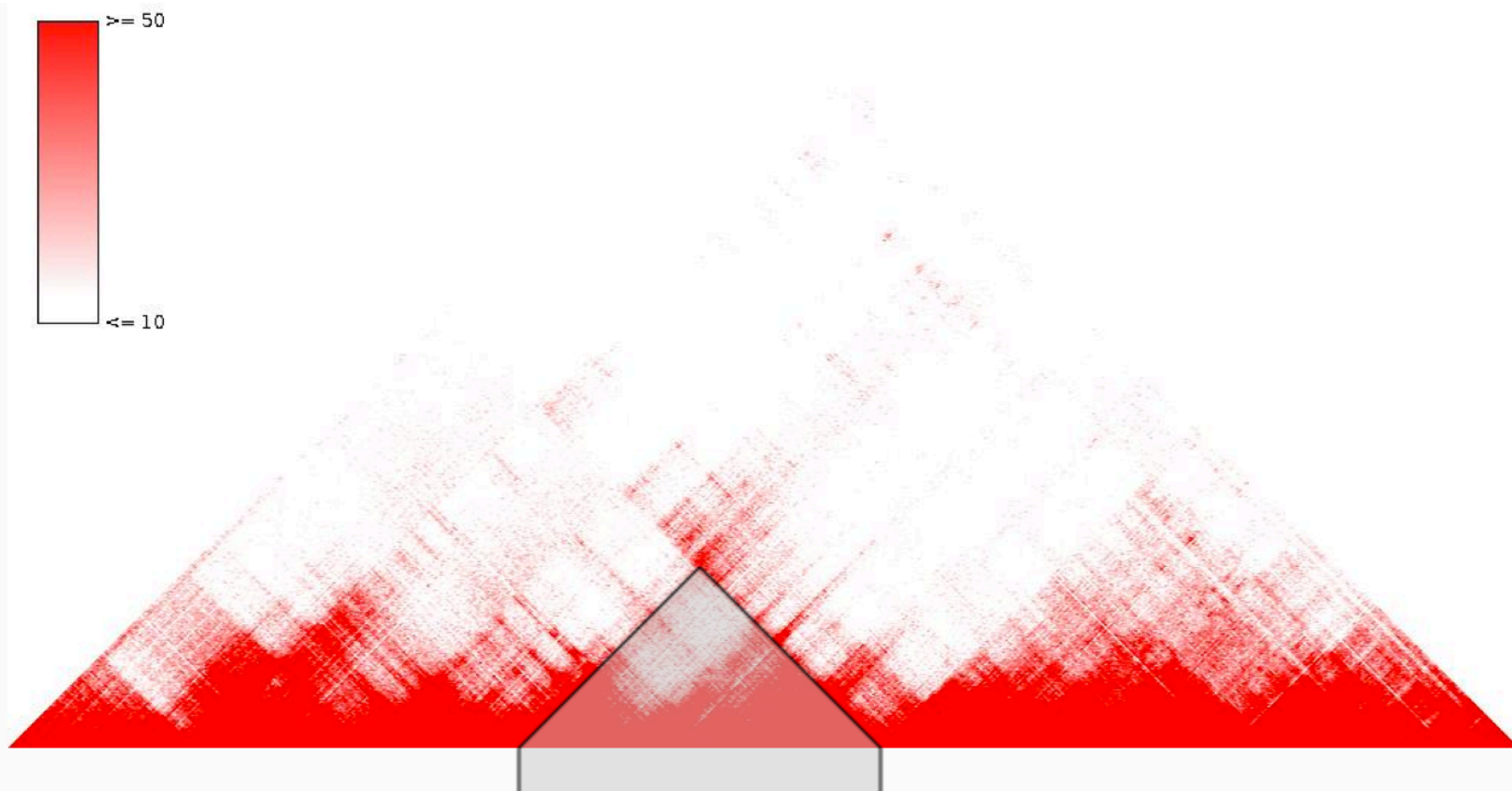
Zoom In (+) 3x

Zoom Out (-) 3x

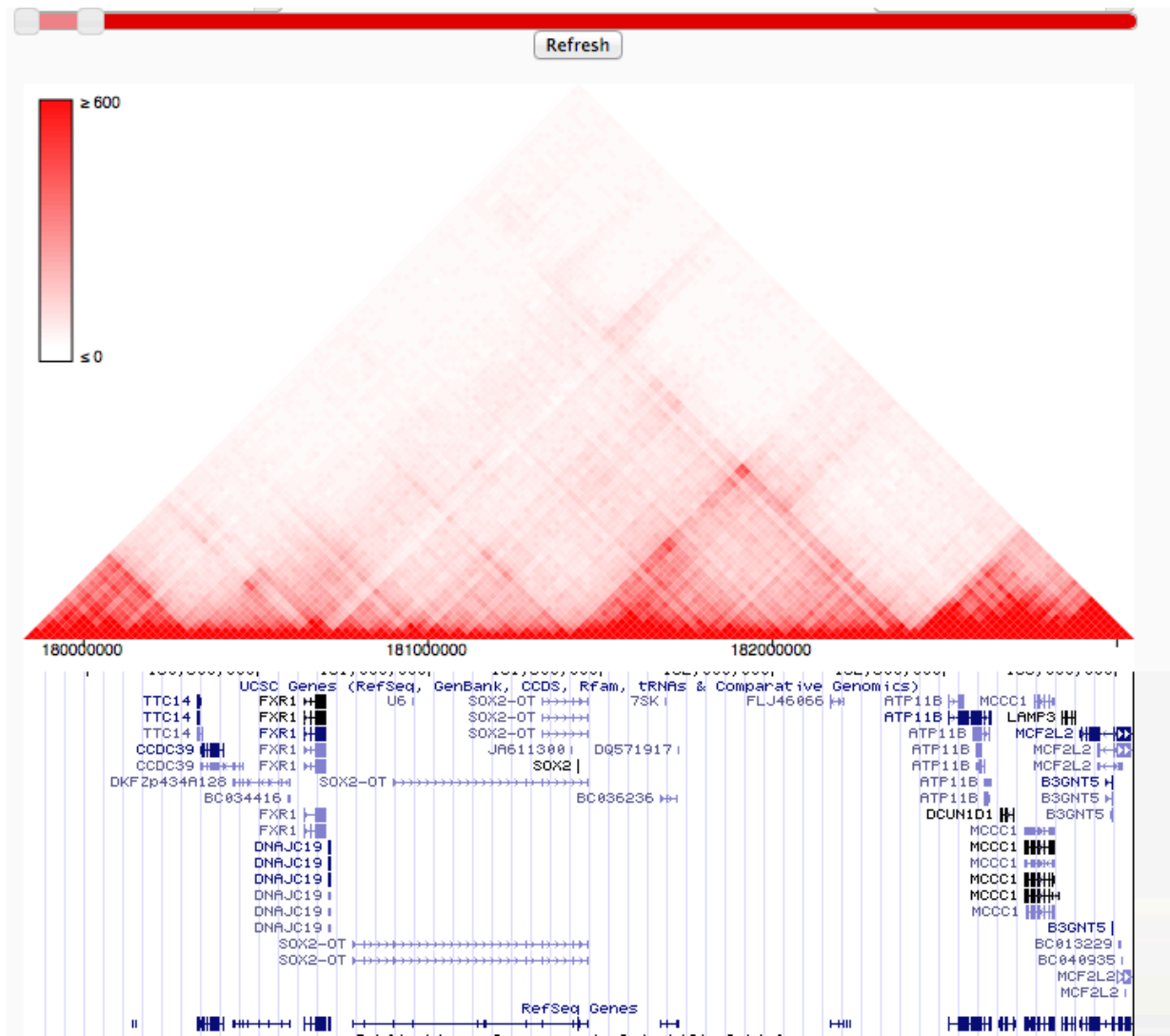
Min or max intensity cutoffs not specified or are invalid. Using default display value.

Hi-C data hg19 GM12878 at 25 kb is from Rao SS, Huntley MH, Durand NC, Stamenova EK, Bochkov ID, Robinson JT, Sanborn AL, Machol I, Omer AD, Lander ES, Aiden EL. A 3D Map of the Human Genome at Kilobase Resolution Reveals Principles of Chromatin Looping. Cell. 2014;159:1665-1680.





# Built-in genome browser session



# You can use your own Hi-C data - no need to upload!



Welcome to the Interactive Hi-C Data Browser!

Accessing Hi-C Data

**Step 1. Choose the assembly of Hi-C data:**  
Species  Assembly

**Step 2. Choose the source of the data:**  
 Browse Available Hi-C Data  
Tissue  Resolution   
OR  
 Use Your Own Data

Data URL

**Step 3. Choose the region to show Hi-C interactions:**

Option 1: Search by Gene Name	Option 2: Search by Location
Gene/Genomic Feature <input type="text"/> <input type="button" value="Show Interaction"/>	Chromosome <input type="text" value="chr1"/> Start <input type="text"/> End <input type="text"/> <input type="button" value="Show Interaction"/>

Optional: User-defined UCSC Browser Session

UCSC Genome Browser Session ID

To retain your UCSC custom tracks and settings, save this URL and paste it here for the next time.  
We recommend creating a session at UCSC and paste the URL instead. (Guide: <http://genome.ucsc.edu/goldenPath/help/hgSessionHelp.html>)

# Use virtual 4C to hypothesize the enhancer promoter interaction



Welcome to YUE Lab  
Computational and Functional Genomics/Epigenomics

HOME

HI-C INTERACTIONS

VIRTUAL 4C

DOWNLOAD

TUTORIAL

CITATION

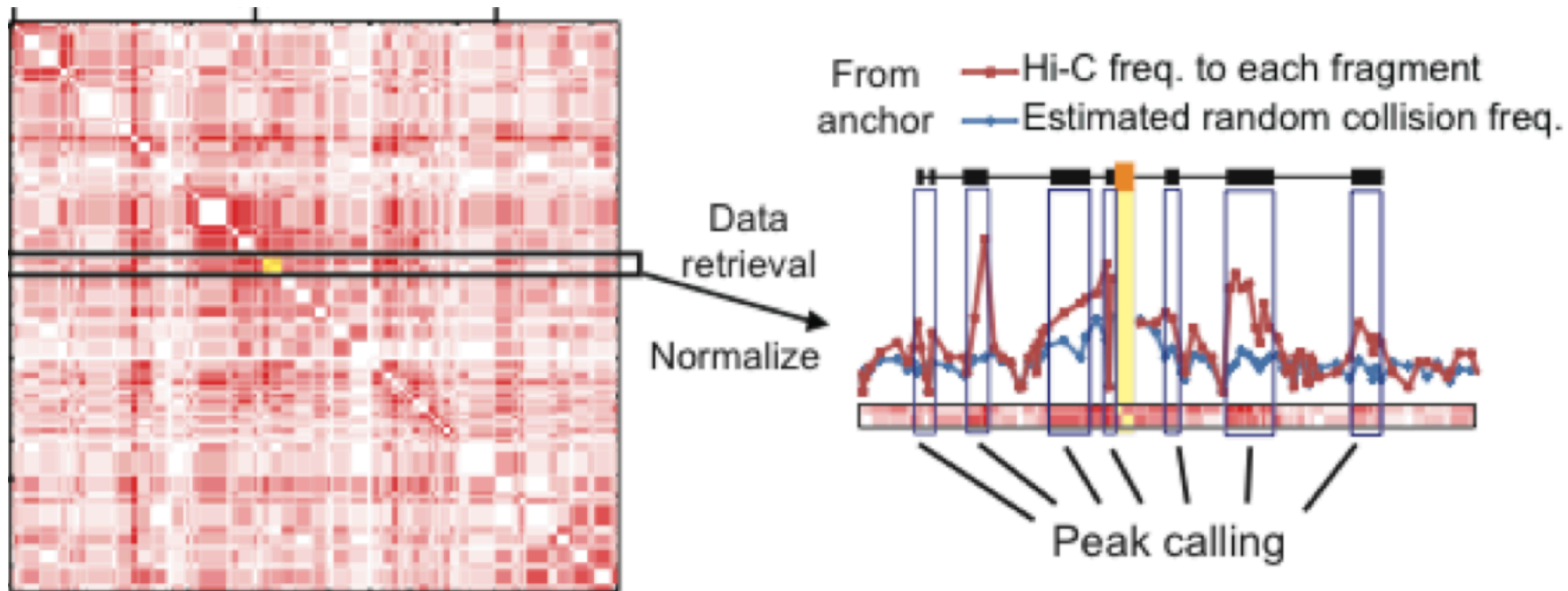
CONTACT

## Virtual 4C created from Hi-C data

Species  Assembly  Tissue

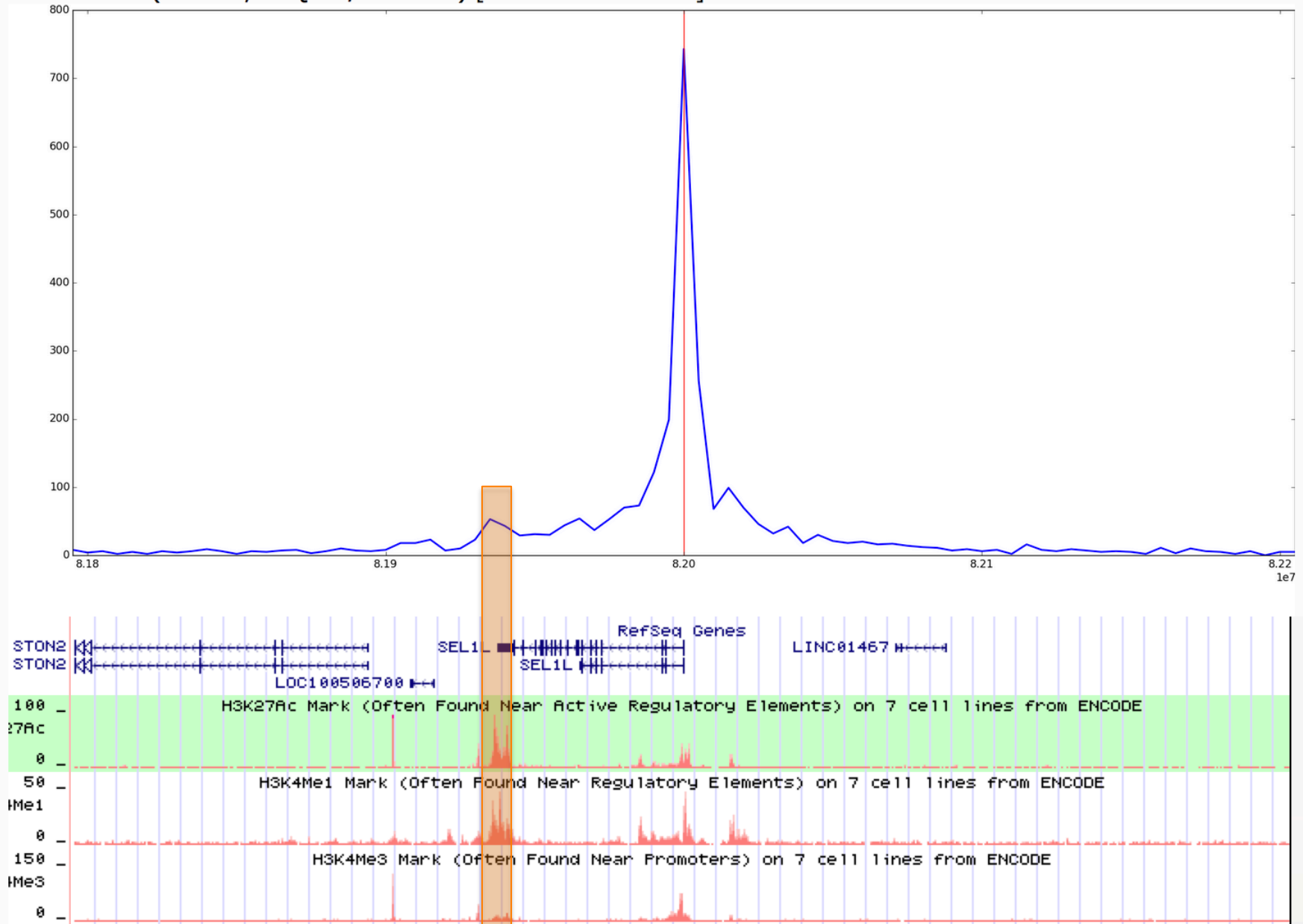
Gene Name:  Extended region in kb (default  $\pm$  500kb):

# Virtual 4C is derived from Hi-C matrix



Available smallest resolution for hg19 GM12878 is 5 kb from Rao SS, Huntley MH, Durand NC, Stamenova EK, Bochkov ID, Robinson JT, Sanborn AL, Machol I, Omer AD, Lander ES, Aiden EL. A 3D Map of the Human Genome at Kilobase Resolution Reveals Principles of Chromatin Looping. Cell. 2014;159:1665-1680.

Gene **SEL1L** (TSA305, UNQ128/PRO1063) [ENSG00000071537] with TSS at chr14:82000205.



# Outline



- 1. Search and download through ENCODE portal**
- 2. Search and visualize ENCODE annotations (gene expression, candidate promoters, enhancers ...)**
- 3. Visualize ENCODE chromatin interaction data (Hi-C, ChIA-PET)**
- 4. Annotate genetic variants by HapReg and RegulomeDB.**

# Online tools for variants annotation



<http://www.regulomedb.org/>

**HaploReg**

<http://www.broadinstitute.org/mammals/haploreg/haploreg.php>



RegulomeDB has been updated to Version 1.1. This includes bringing our database up-to-date with current ENCODE releases: [Xie et al. \(2013\)](#) and Boyle et al. (2014). We have also added Chromatin States from the Roadmap Epigenome Consortium (unpublished) as well as updates to DNase footprinting, PWMs, and DNA Methylation.

*Enter dbSNP IDs, 0-based coordinates, BED files, VCF files, GFF3 files (hg19).*

**Type variant name here**

**Submit**

# Scoring scheme for regulomeDB

**Table 2.** RegulomeDB variant classification scheme

Category scheme	
Category	Description
1a	Likely to affect binding and linked to expression of a gene target eQTL + TF binding + matched TF motif + matched DNase footprint + DNase peak
1b	eQTL + TF binding + any motif + DNase footprint + DNase peak
1c	eQTL + TF binding + matched TF motif + DNase peak
1d	eQTL + TF binding + any motif + DNase peak
1e	eQTL + TF binding + matched TF motif
1f	eQTL + TF binding/DNase peak
Likely to affect binding	
2a	TF binding + matched TF motif + matched DNase footprint + DNase peak
2b	TF binding + any motif + DNase footprint + DNase peak
2c	TF binding + matched TF motif + DNase peak
Less likely to affect binding	
3a	TF binding + any motif + DNase peak
3b	TF binding + matched TF motif
Minimal binding evidence	
4	TF binding + DNase peak
5	TF binding or DNase peak
6	Motif hit

Lower scores indicate increasing evidence for a variant to be located in a functional region. Category 1 variants have equivalents in other categories with the additional requirement of eQTL information.

# Variants annotation by regulomeDB

rs17293632 is Associated with IBD and Crohn's Disease

Date Added to Catalog (since 11/25/08)	First Author/Date/Journal/Study	Disease/Trait	Initial Sample Description	Replication Sample Description	Region	Reported Gene(s)	Mapped Gene(s)	Strongest SNP-Risk Allele	Context	Risk Allele Frequency in Controls	P-value	OR or beta-coefficient and [95% CI]	Platform [SNPs passing QC]	CNV
02/12/13	Justin L November 01, 2012 <i>Nature</i> Host-microbe interactions have shaped the genetic architecture of inflammatory bowel disease.	Inflammatory bowel disease	12,924 European ancestry cases, 21,442 European ancestry controls	25,683 European ancestry cases, 17,015 European ancestry controls	15q22.33	SMAD3	SMAD3	rs17293632-T	intron	0.235	$6 \times 10^{-16}$	1.067 [1.032-1.102]	Affymetrix & Illumina [1.23 million] (Imputed)	N
10/19/12	Franke A November 21, 2010 <i>Nat Genet</i> Genome-wide meta-analysis increases to 71 the number of confirmed Crohn's disease susceptibility loci.	Crohn's disease	6,333 European ancestry cases, 15,056 European ancestry controls	15,694 European ancestry cases, 14,026 European ancestry controls, 414 European ancestry trios	15q22.33	SMAD3	SMAD3	rs17293632-T	intron	0.233	$3 \times 10^{-19}$	1.12 [1.07-1.16]	Affymetrix & Illumina [953,241] (Imputed)	N

Slide provided by Jill Moore (U. Mass)

# Variants annotation by regulomeDB

The search has evaluated 1 input line(s) and found 1 SNP(s).

## Summary of SNP analysis

Show  entries

Coordinate (0-based)	dbSNP ID	? Regulome DB Score	Other Resources
chr15:67442595	rs17293632	2a	<a href="#">UCSC</a>   <a href="#">ENSEMBL</a>   <a href="#">dbSNP</a>

Showing 1 to 1 of 1 entries

BED

GFF

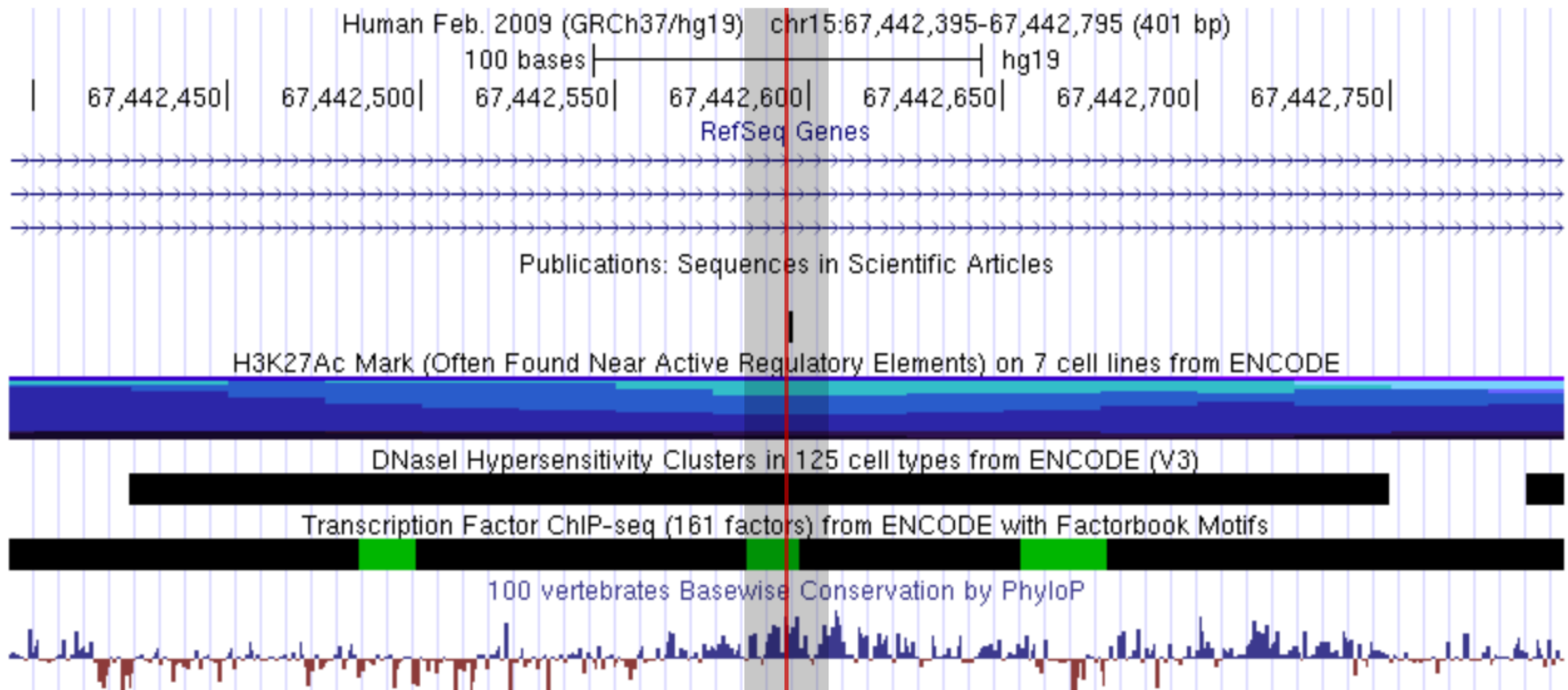
Full Output

# Variants annotation by regulomeDB

## Data supporting chr15:67442595 (rs17293632)

Score: 2a

Likely to affect binding

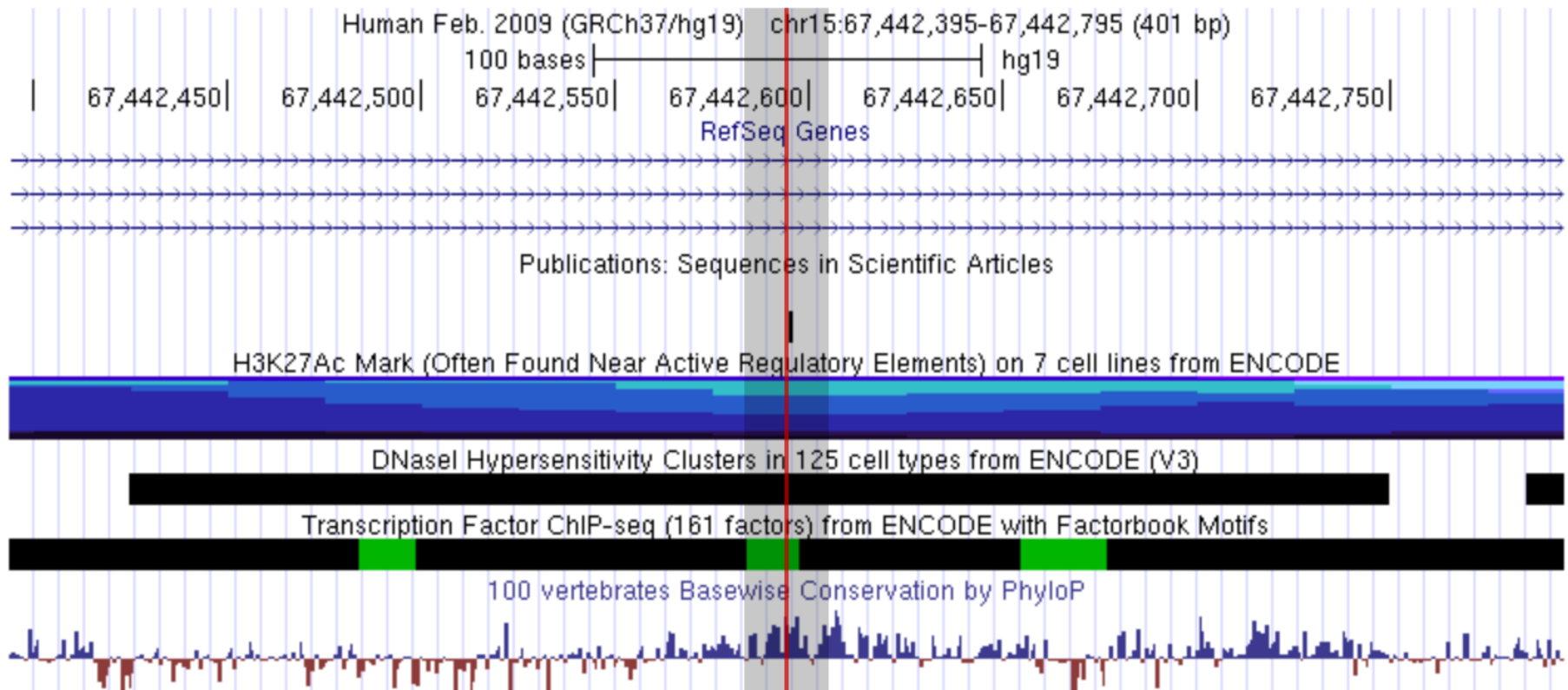


# Variants annotation by regulomeDB

## Data supporting chr15:67442595 (rs17293632)

Score: 2a

Likely to affect binding



# Variants (rs17293632) annotation by HaploReg



## HaploReg v4.1

HaploReg is a tool for exploring annotations of the noncoding genome at variants on haplotype blocks, such as candidate regulatory SNPs at disease-associated loci. Using LD information from the 1000 Genomes Project, linked SNPs and small indels can be visualized along with chromatin state and protein binding annotation from the Roadmap Epigenomics and ENCODE projects, sequence conservation across mammals, the effect of SNPs on regulatory motifs, and the effect of SNPs on expression from eQTL studies. HaploReg is designed for researchers developing mechanistic hypotheses of the impact of non-coding variants on clinical phenotypes and normal variation.

**Update 2015.11.05: Version 4.1** GWAS and eQTL have been updated; a simpler pruning strategy is applied when combining GWAS; and links out to other NHGRI/EBI GWAS hits and GRASP QTL hits are provided.

**Update 2015.09.15: Version 4.0** now includes many recent eQTL results including the GTEx pilot, four different options for defining enhancers using Roadmap Epigenomics data, and a complete set of source files for download and local analysis. Older versions available: [v3](#), [v2](#), [v1](#).

[Build Query](#) [Set Options](#) [Documentation](#)

Use one of the three methods below to enter a set of variants. If an  $r^2$  threshold is specified (see the Set Options tab), results for each variant will be shown in a separate table along with other variants in LD. If  $r^2$  is set to NA, only queried variants will be shown, together in one table.

Query (comma-delimited list of rsIDs OR a single region as chrN:start-end):  
or, upload a text file (one refSNP ID per line):  No file selected.  
or, select a GWAS:

# Variants (rs17293632) annotation by HaploReg

HaploReg will list the potential functional annotation for all the variants in the same LD.

Query SNP: **rs17293632** and variants with  $r^2 \geq 0.8$

chr	pos (hg38)	LD (r <sup>2</sup> )	LD (D')	variant	Ref	Alt	AFR freq	AMR freq	ASN freq	EUR freq	SiPhy cons	Promoter histone marks	Enhancer histone marks	DNase	Proteins bound	Motifs changed	NHGR/EBI GWAS hits	GRASP QTL hits	Selected eQTL hits	GENCODE genes	dbSNP func annot
15	67149412	0.98	1	<a href="#">rs72743461</a>	C	A	0.03	0.14	0.03	0.22		ESC, IPSC, CRVX	23 tissues	10 tissues		AP-4,Ets,HEN1			1 hit	SMAD3	intronic
15	67150258	1	1	<b>rs17293632</b>	C	T	0.02	0.14	0.03	0.21		17 tissues	20 tissues	41 tissues	24 bound proteins	25 altered motifs	2 hits	315 hits	2 hits	SMAD3	intronic
15	67156025	0.95	0.98	<a href="#">rs56375023</a>	G	A	0.02	0.14	0.03	0.22			BLD, LNG						1 hit	SMAD3	intronic
15	67157967	0.93	0.97	<a href="#">rs17228058</a>	A	G	0.02	0.14	0.03	0.21		MUS	15 tissues	7 tissues		GR,NERF1a,PU.1	1 hit		2 hits	SMAD3	intronic
15	67163292	0.94	0.97	<a href="#">rs56062135</a>	C	T	0.03	0.14	0.03	0.21			10 tissues			ERalpha-a			1 hit	SMAD3	intronic
15	67171953	0.87	0.95	<a href="#">rs72743477</a>	A	G	0.03	0.13	0.02	0.21		MUS	7 tissues			4 altered motifs			1 hit	SMAD3	intronic
15	67174261	0.85	0.94	<a href="#">rs72743482</a>	A	G	0.02	0.13	0.03	0.21			13 tissues	4 tissues		Ncx,Sp4			1 hit	SMAD3	intronic

# Variants (rs17293632) annotation by HaploReg

## Sequence facts

chr	pos (hg19)	chr	pos (hg38)	Reference	Alternate	1000 Genomes Phase 1 Frequencies				Sequence constraint		dbSNP functional annotation
						AFR	AMR	ASN	EUR	by GERP	by SiPhy	
chr15	67442596	chr15	67150258	C	T	0.02	0.14	0.03	0.21	No	No	intronic

## Closest annotated gene

Source	Distance	Direction	ID/Link	Common name	Description
GENCODE	NA	Within gene	<a href="#">ENSG00000166949.10</a>	SMAD3	SMAD family member 3 [Source:HGNC Symbol;Acc:6769]
RefSeq	NA	Within gene	<a href="#">NM_001145103</a>	SMAD3	SMAD family member 3 [Source:HGNC Symbol;Acc:6769]

## Regulatory chromatin states from DNase and histone ChIP-Seq (Roadmap Epigenomics Consortium, 2015)

(Black = missing data)

Epigenome ID (EID)	Group	Mnemonic	Description	Chromatin states (Core 15-state model)	Chromatin states (25-state model using 12 imputed marks)	H3K4me1	H3K4me3	H3K27ac
E017	IMR90	LNG.IMR90	IMR90 fetal lung fibroblasts Cell Line	2_TssAFlnk	9_TxReg	H3K4me1_Enh	H3K4me3_Pro	H3K27ac_Enh
E002	ESC	ESC.WA7	ES-WA7 Cells	1_TssA	13_EnhA1	H3K4me1_Enh	H3K4me3_Pro	
E008	ESC	ESC.H9	H9 Cells	2_TssAFlnk	9_TxReg	H3K4me1_Enh	H3K4me3_Pro	H3K27ac_Enh
E001	ESC	ESC.I3	ES-I3 Cells	2_TssAFlnk	9_TxReg	H3K4me1_Enh	H3K4me3_Pro	
E015	ESC	ESC.HUES6	HUES6 Cells	2_TssAFlnk	9_TxReg	H3K4me1_Enh	H3K4me3_Pro	H3K27ac_Enh
E014	ESC	ESC.HUES48	HUES48 Cells	2_TssAFlnk	9_TxReg	H3K4me1_Enh	H3K4me3_Pro	H3K27ac_Enh

# Acknowledgement



**ENCODE consortium**

## **ENCODE DCC**

- Mike Cherry
- Seth Strattan
- Ben Hitz
- Aditi Kalpagam  
Narayanan

...

## **ENCODE EDAC**

- Zhiping Weng
- Jill Moore

...

**Dr. Bing Ren**

**Dr. Ting Wang**

**Yue Lab**

- Yanli Wang
- Lijun Zhang

**and many others ...**

## **Funding Resources**

- Leukemia Research Foundation,
- PhRMA Foundation,
- WW Smith Foundation,
- 1U01CA200060 (Wang, 4DN)
- 1R24DK106766 (Hardison)