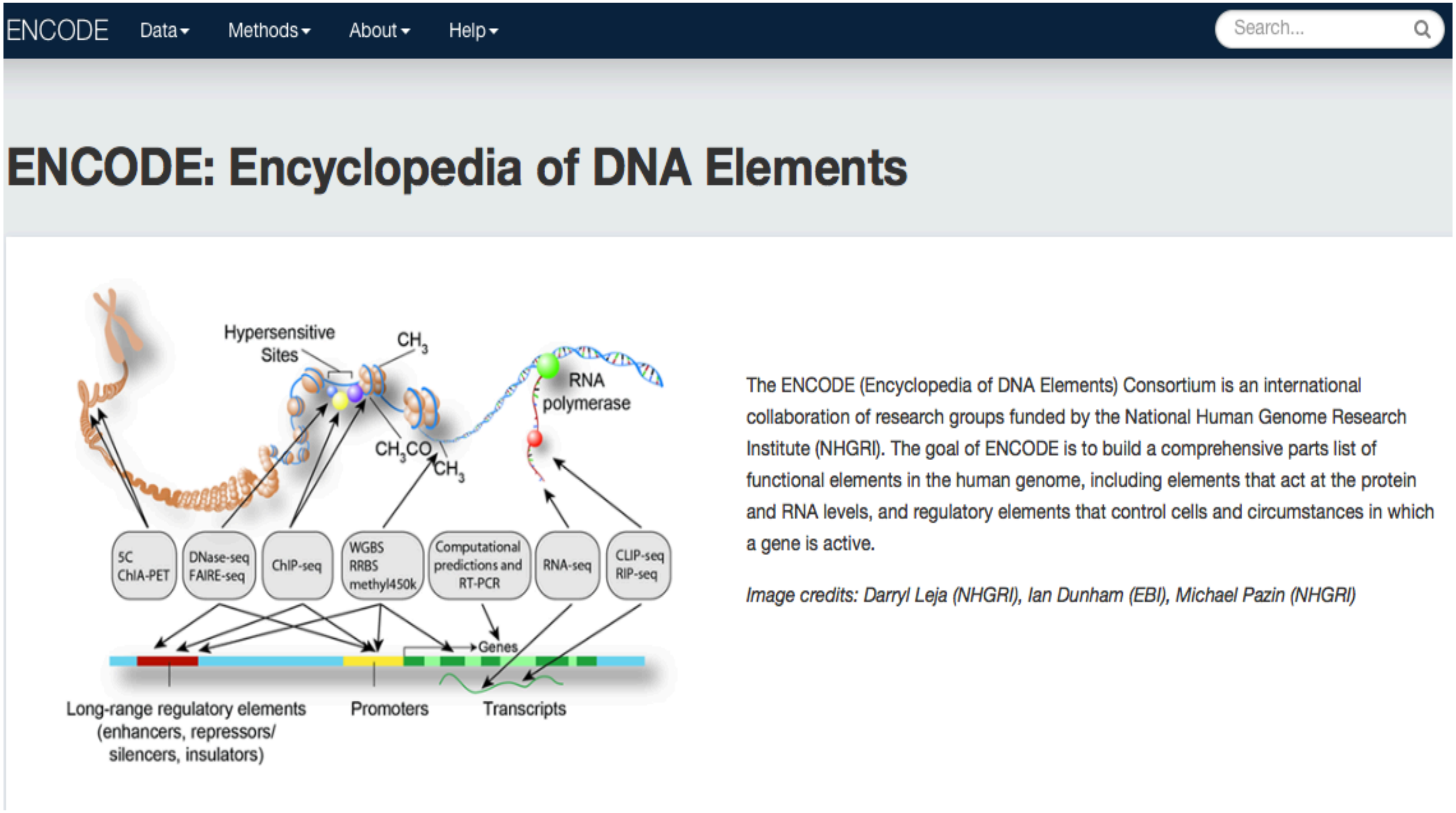
ENCODE workshop
SOT 2016

Feng Yue, PhD
Dept. of Biochemistry & Molecular Biology
School of Medicine
Penn 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)**

www.encodeproject.org

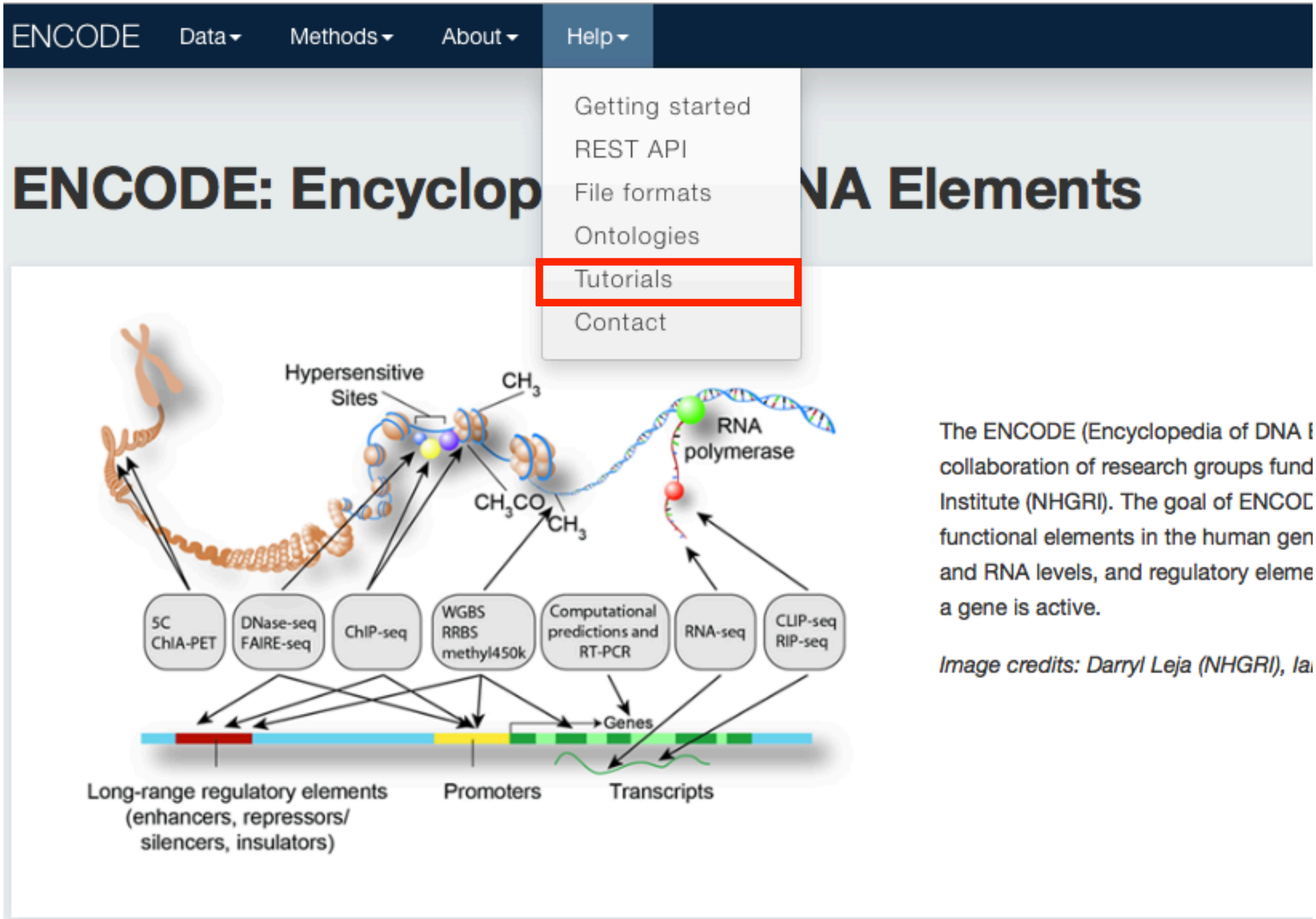


ENCODE: Encyclopedia of DNA Elements

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 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)

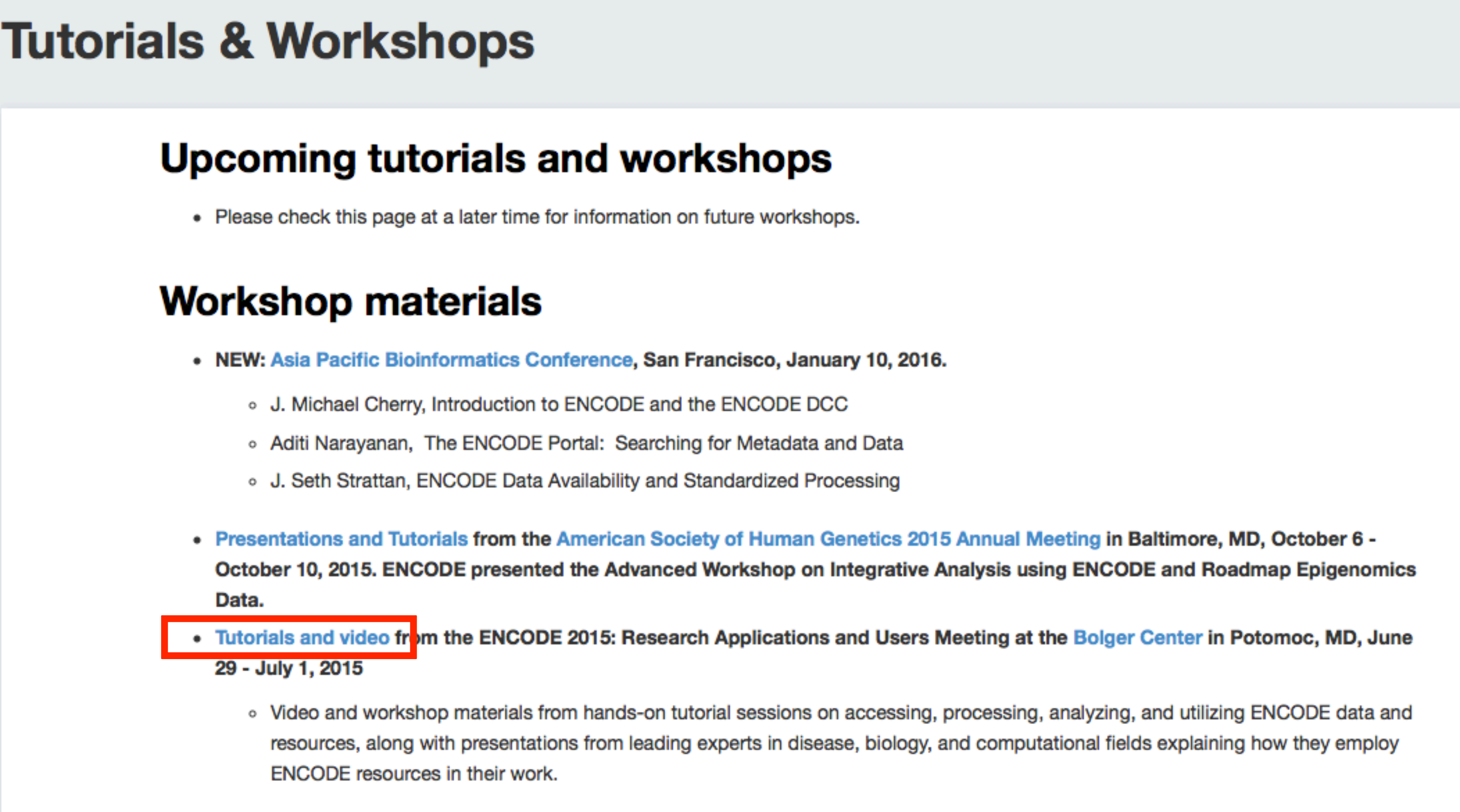
All the tutorial materials are available on the ENCODE portal



ENCODE: Encyclop NA Elements

The ENCODE (Encyclopedia of DNA E
collaboration of research groups fund
Institute (NHGRI). The goal of ENCOD
functional elements in the human gen
and RNA levels, and regulatory eleme
a gene is active.

Image credits: Darryl Leja (NHGRI), Ia



Tutorials & Workshops

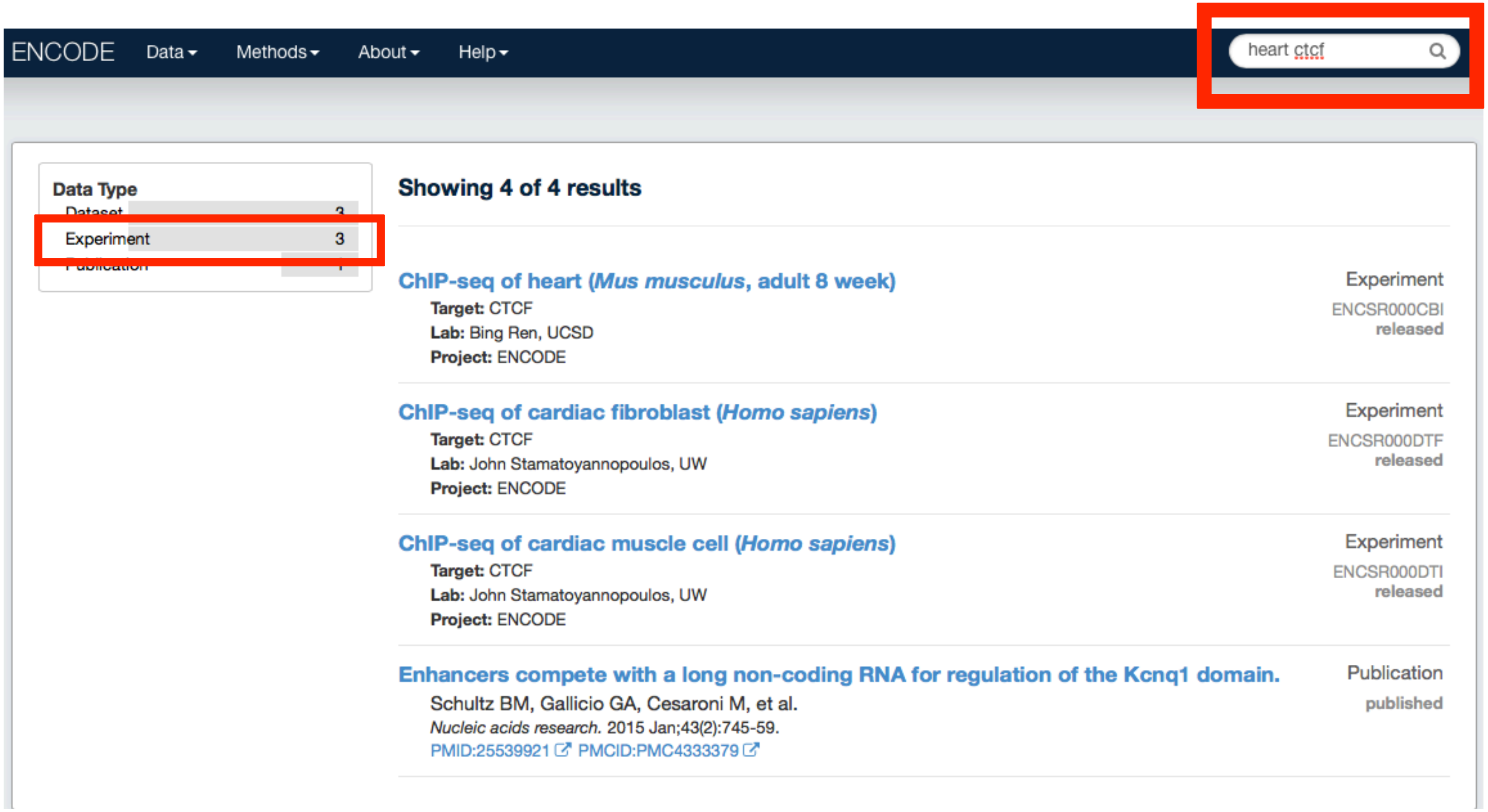
Upcoming tutorials and workshops

- 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.

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



Data Type

Dataset	3
Experiment	3
Publication	1

Showing 4 of 4 results

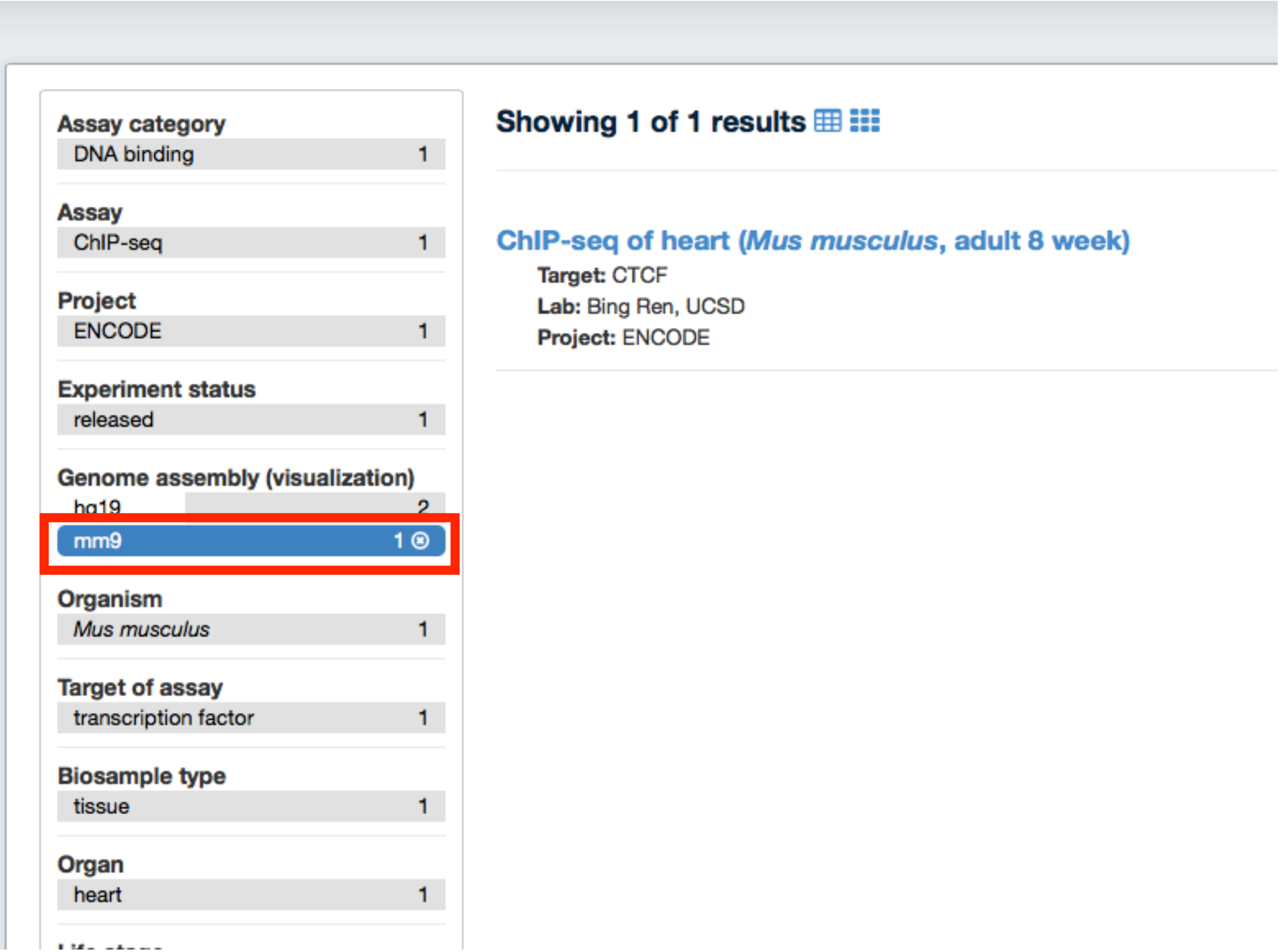
ChIP-seq of heart (*Mus musculus*, adult 8 week) — Experiment, ENCSR000CBI, released
Target: CTCF
Lab: Bing Ren, UCSD
Project: ENCODE

ChIP-seq of cardiac fibroblast (*Homo sapiens*) — Experiment, ENCSR000DTF, released
Target: CTCF
Lab: John Stamatoyannopoulos, UW
Project: ENCODE

ChIP-seq of cardiac muscle cell (*Homo sapiens*) — Experiment, ENCSR000DTI, released
Target: CTCF
Lab: John Stamatoyannopoulos, UW
Project: ENCODE

Enhancers compete with a long non-coding RNA for regulation of the Kcnq1 domain. — Publication, published
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



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

Target of assay	
transcription factor	1

Biosample type	
tissue	1

Organ	
heart	1

Showing 1 of 1 results

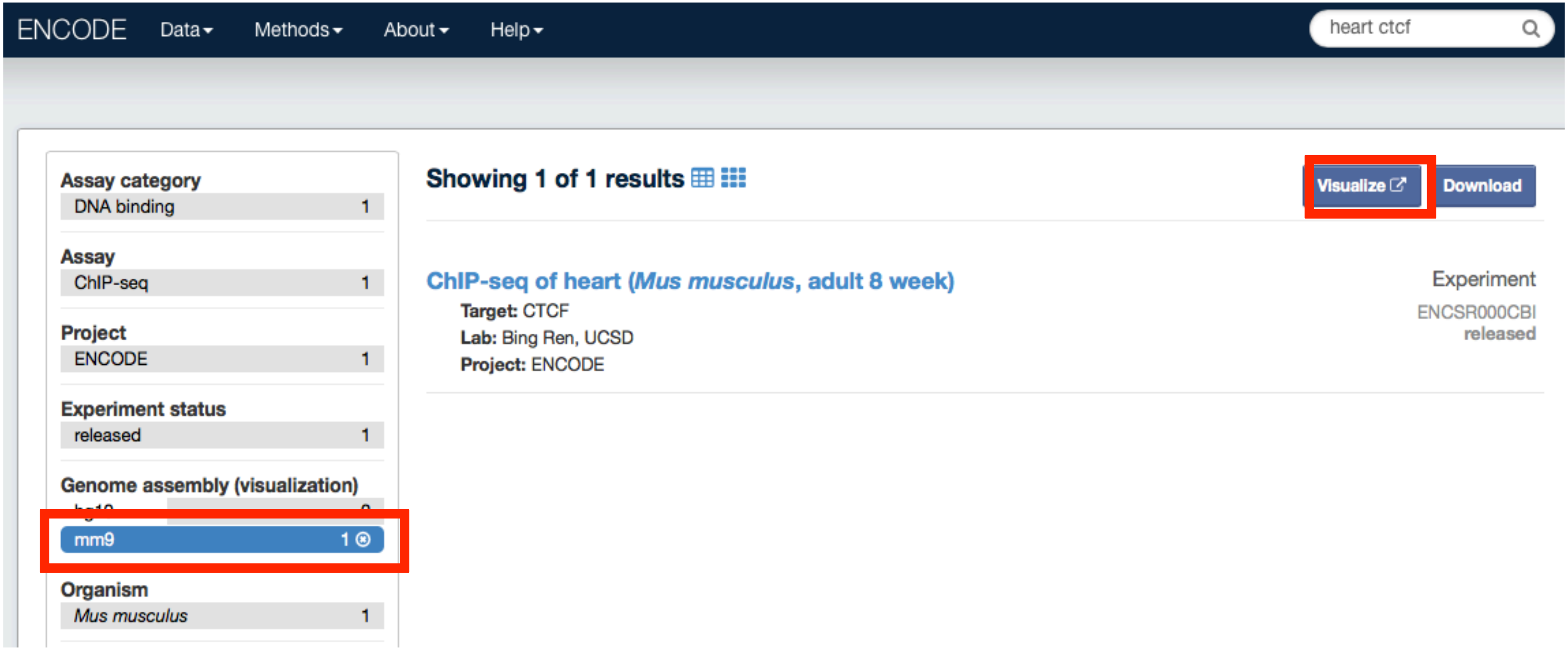
ChIP-seq of heart (*Mus musculus*, adult 8 week)

Target: CTCF

Lab: Bing Ren, UCSD

Project: ENCODE

Choose mm9 in genome assembly, then click visualization button



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

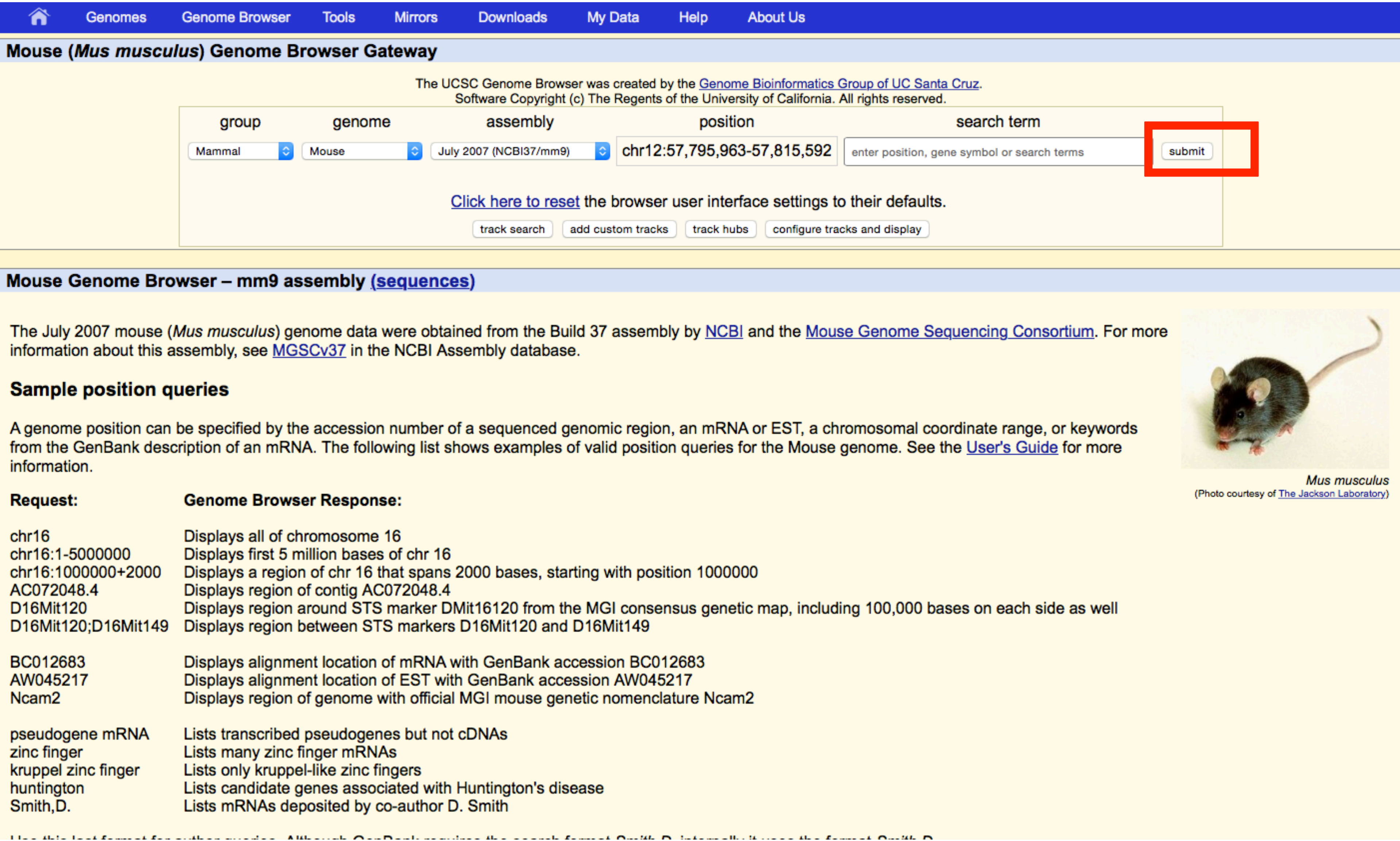
Showing 1 of 1 results

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

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

Experiment
ENCSR000CBI
released

Visualizing data in UCSC genome browser



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	submit

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.

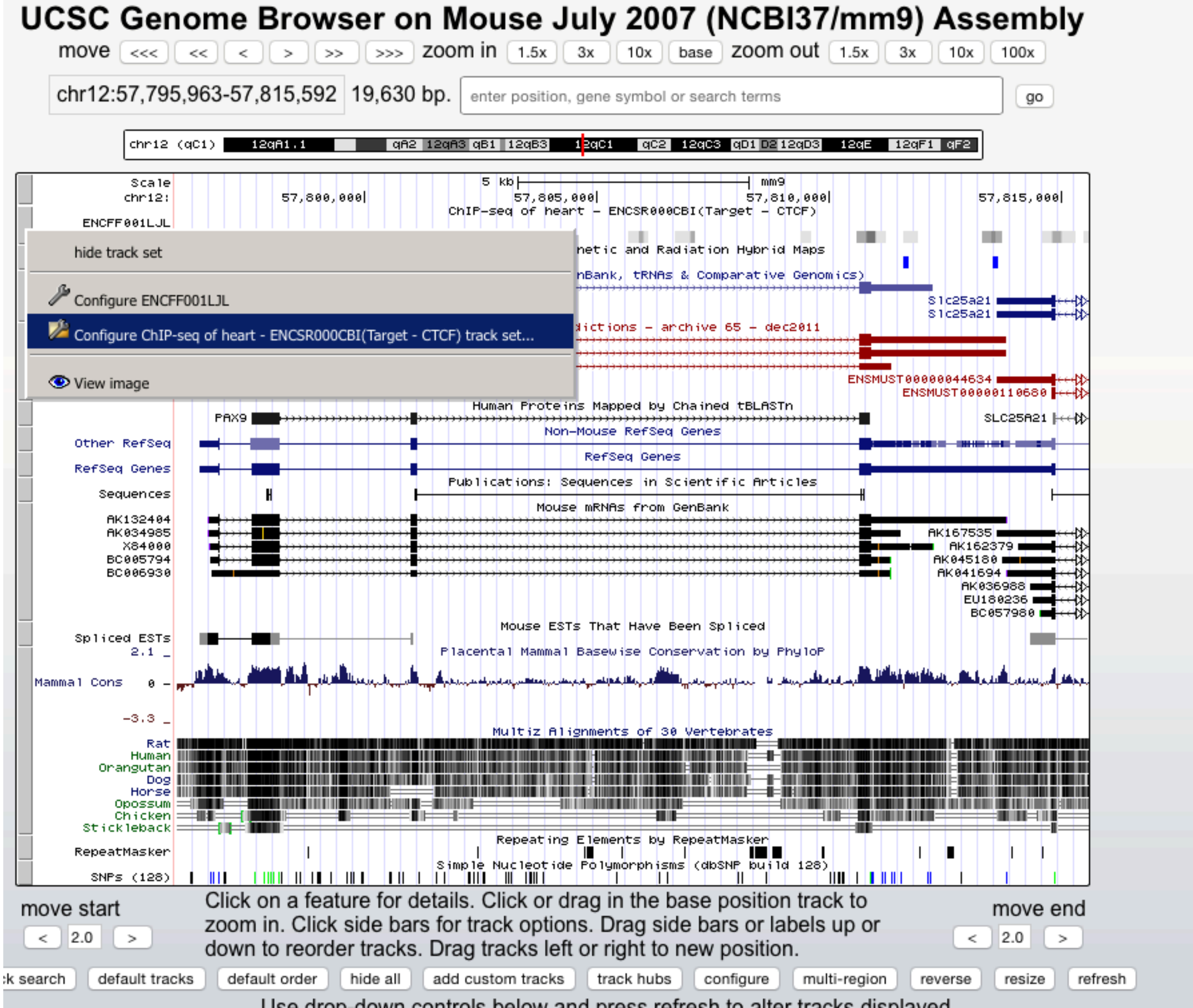
Mus musculus
(Photo courtesy of The Jackson Laboratory)

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

Visualizing data in UCSC genome browser

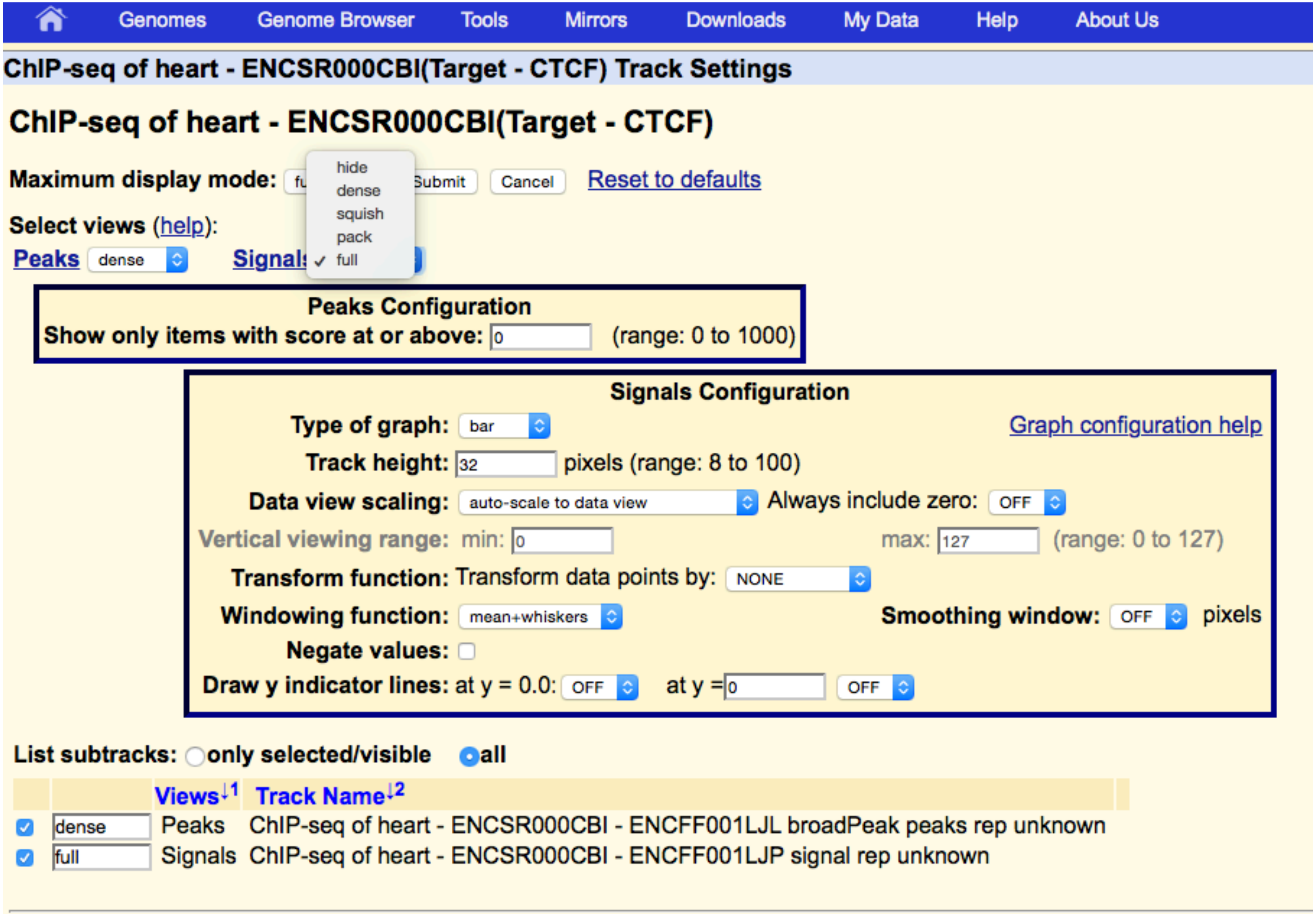


UCSC Genome Browser on Mouse July 2007 (NCBI37/mm9) Assembly

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

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.

Visualizing data in UCSC genome browser



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

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

Maximum display mode: full

Select views (help):

Peaks: dense   Signals: full

Peaks Configuration

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

Signals Configuration

Type of graph: bar

Track height: 32 pixels (range: 8 to 100)

Data view scaling: auto-scale to data view   **Always include zero:** OFF

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

Transform function: Transform data points by: NONE

Windowing function: mean+whiskers   **Smoothing window:** OFF pixels

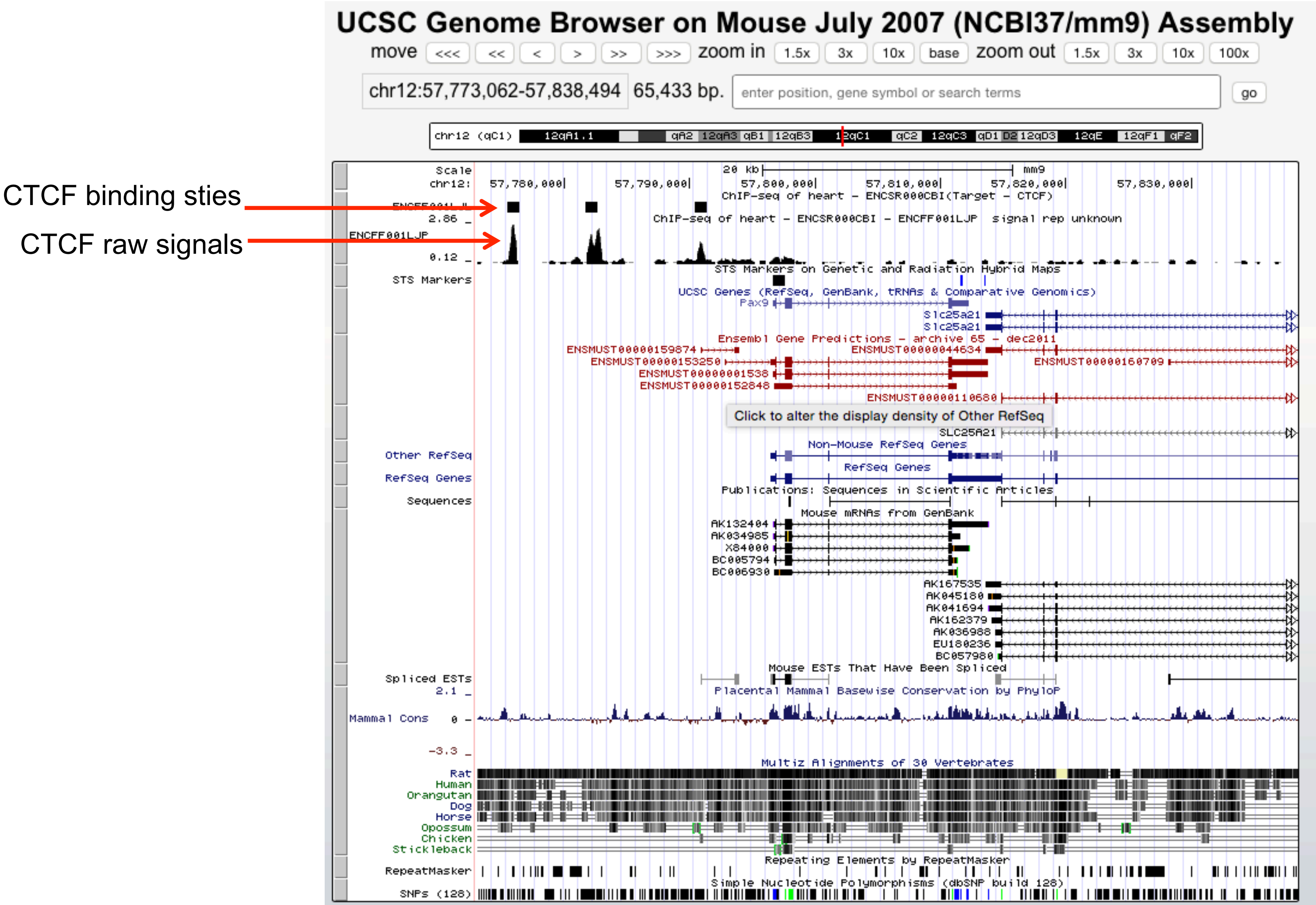
Negate values: ☐

Draw y indicator lines: at y = 0.0: OFF   at y = 0   OFF

List subtracks: ○ only selected/visible ● all

		Views↓1	Track Name↓2
☑	dense	Peaks	ChIP-seq of heart - ENCSR000CBI - ENCFF001LJL broadPeak peaks rep unknown
☑	full	Signals	ChIP-seq of heart - ENCSR000CBI - ENCFF001LJP signal rep unknown

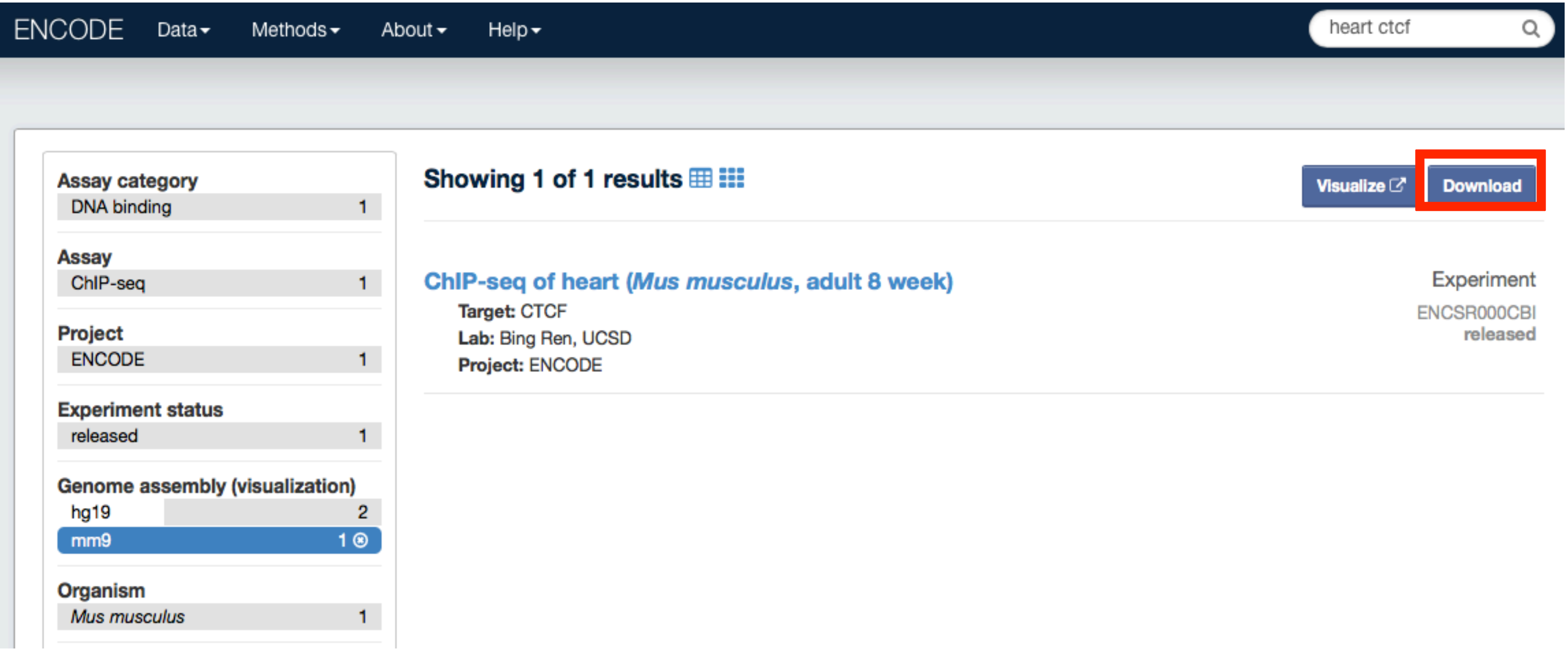
Visualizing data in UCSC genome browser



CTCF binding sties

CTCF raw signals

Downloading data:
Option 1: batch download



heart ctcf

Showing 1 of 1 results

Assay category

DNA binding	1

Assay

ChIP-seq	1

Project

ENCODE	1

Experiment status

released	1

Genome assembly (visualization)

hg19	2
mm9	1 ⊗

Organism

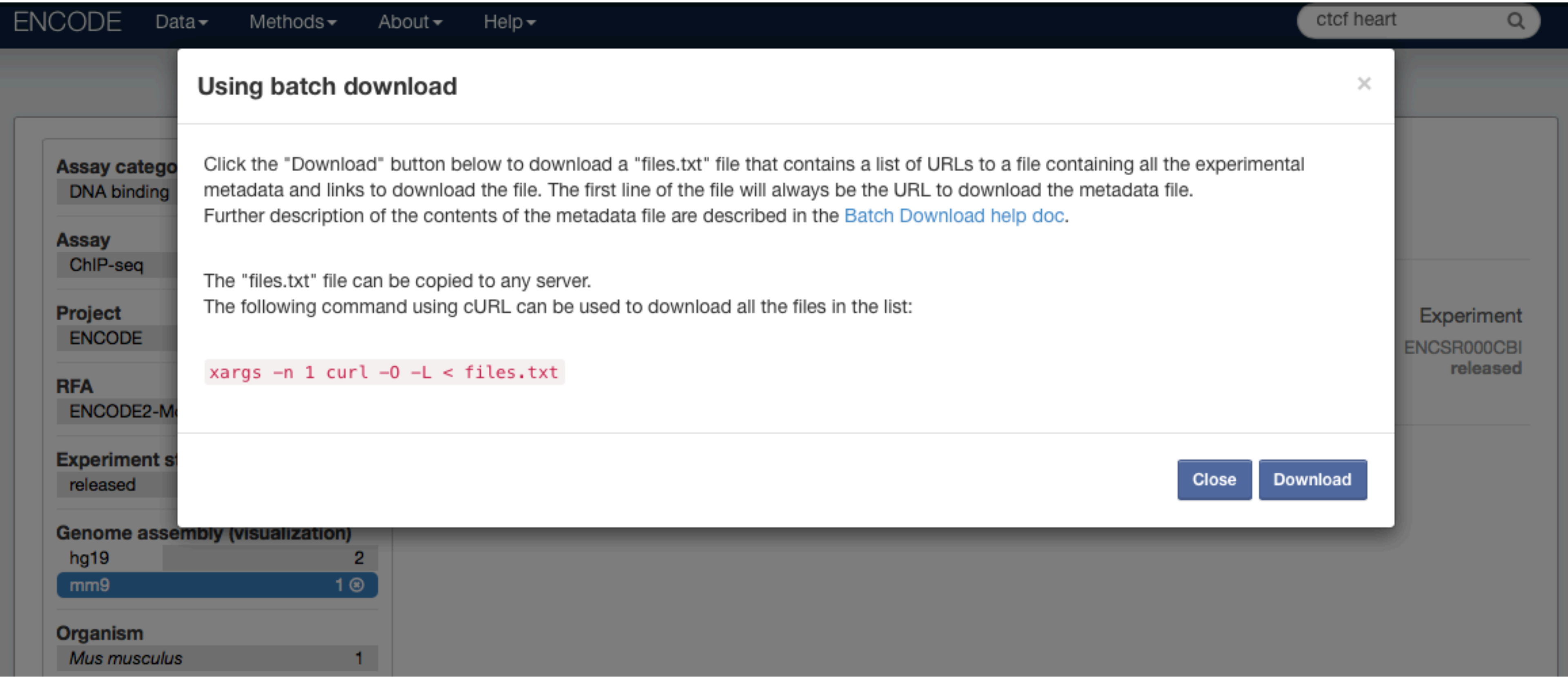
Mus musculus	1

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



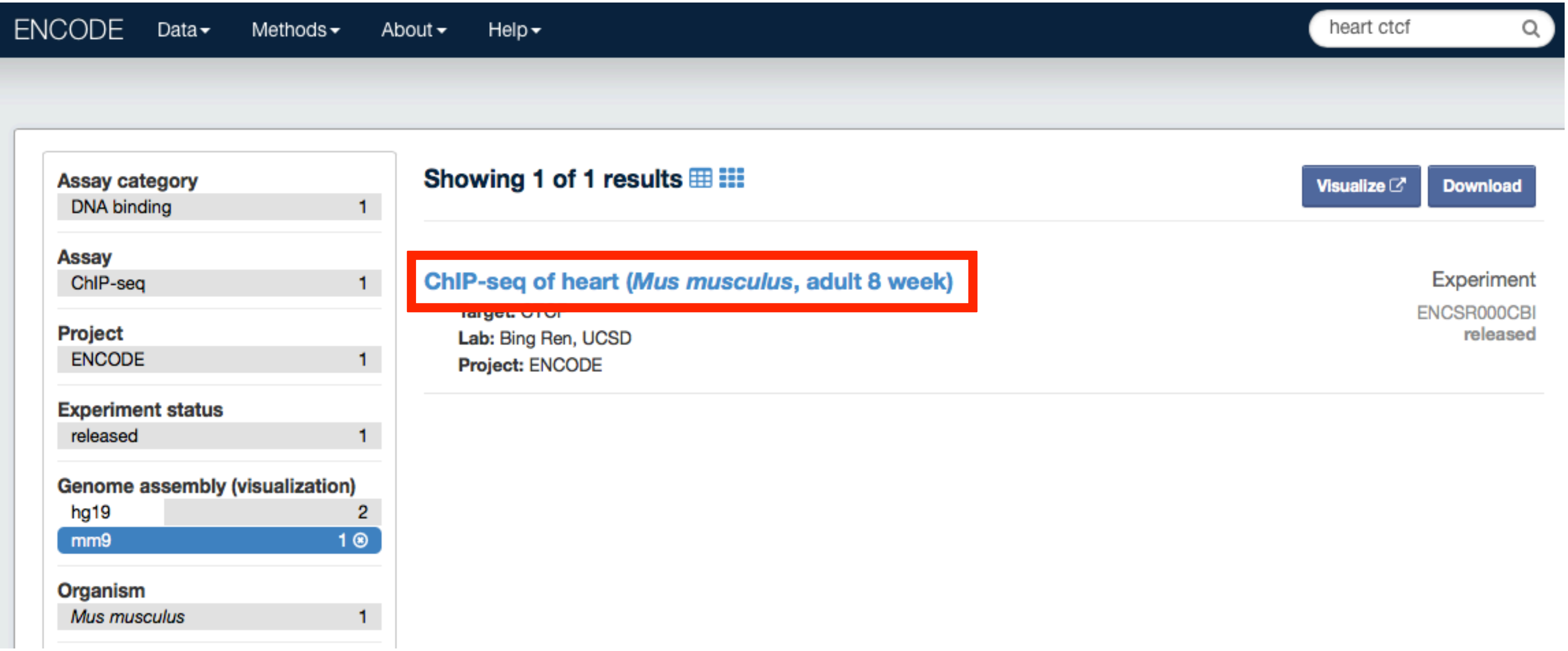
Using batch download

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
```

Option 2: download individual file



Showing 1 of 1 results

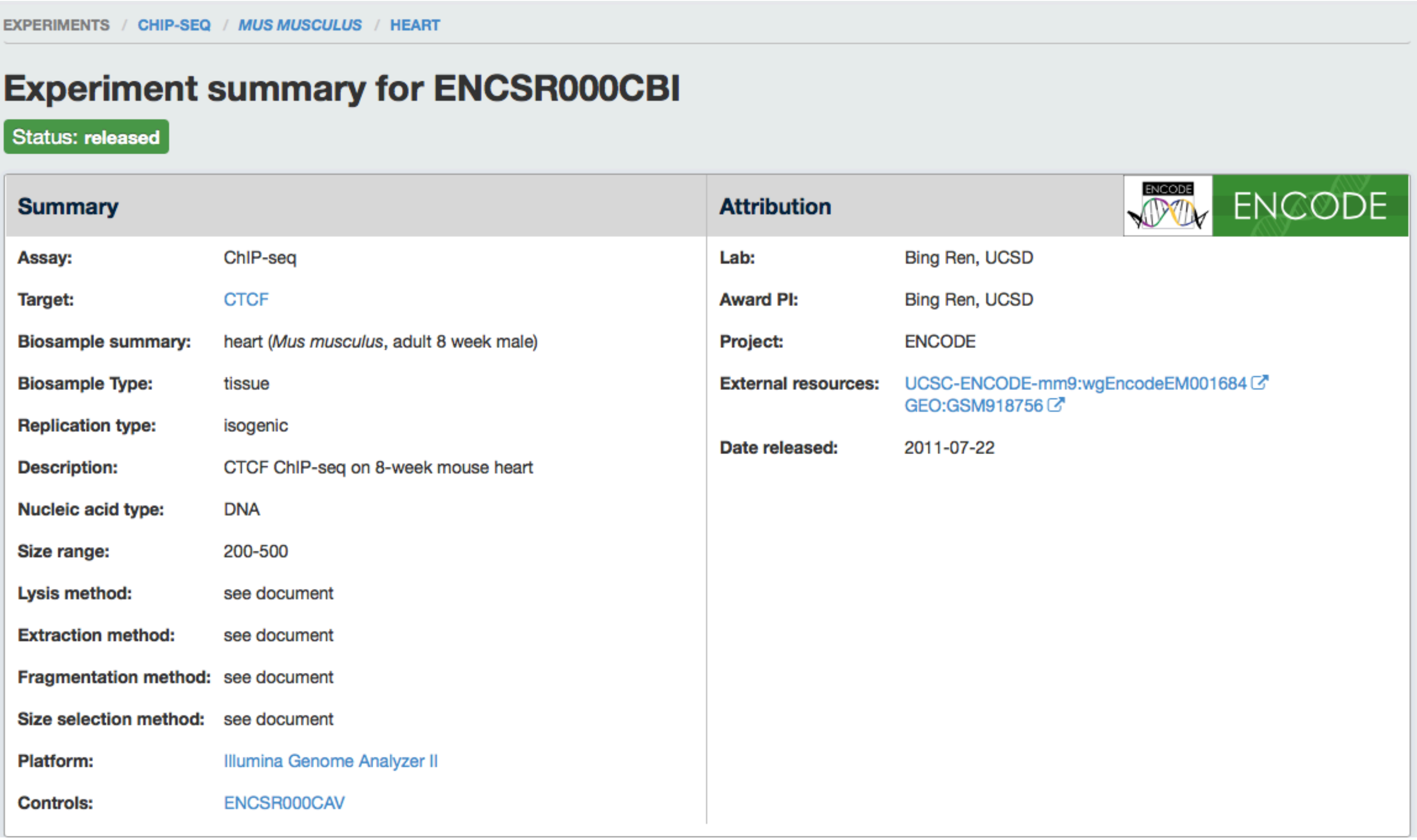
Facet	Value	Count
Assay category	DNA binding	1
Assay	ChIP-seq	1
Project	ENCODE	1
Experiment status	released	1
Genome assembly (visualization)	hg19	2
	mm9	1
Organism	*Mus musculus*	1

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

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

Experiment
ENCSR000CBI
released

Option 2: download individual file



EXPERIMENTS / CHIP-SEQ / *MUS MUSCULUS* / HEART

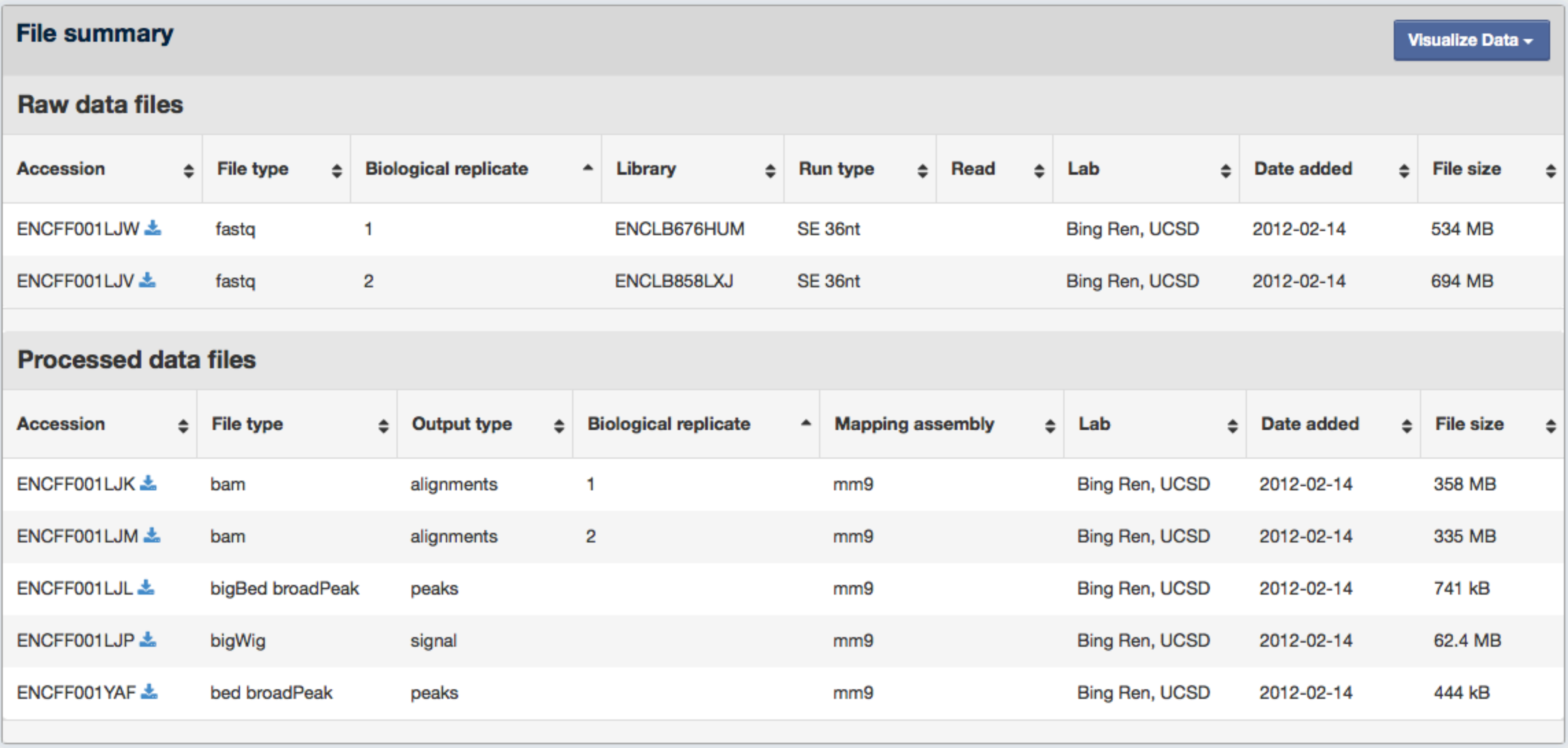
Experiment summary for ENCSR000CBI

Status: released

Summary	
Assay:	ChIP-seq
Target:	CTCF
Biosample summary:	heart (*Mus musculus*, adult 8 week male)
Biosample Type:	tissue
Replication type:	isogenic
Description:	CTCF ChIP-seq on 8-week mouse heart
Nucleic acid type:	DNA
Size range:	200-500
Lysis method:	see document
Extraction method:	see document
Fragmentation method:	see document
Size selection method:	see document
Platform:	Illumina Genome Analyzer II
Controls:	ENCSR000CAV

Attribution	
Lab:	Bing Ren, UCSD
Award PI:	Bing Ren, UCSD
Project:	ENCODE
External resources:	UCSC-ENCODE-mm9:wgEncodeEM001684 GEO:GSM918756
Date released:	2011-07-22

Option 2: download individual file



File summary

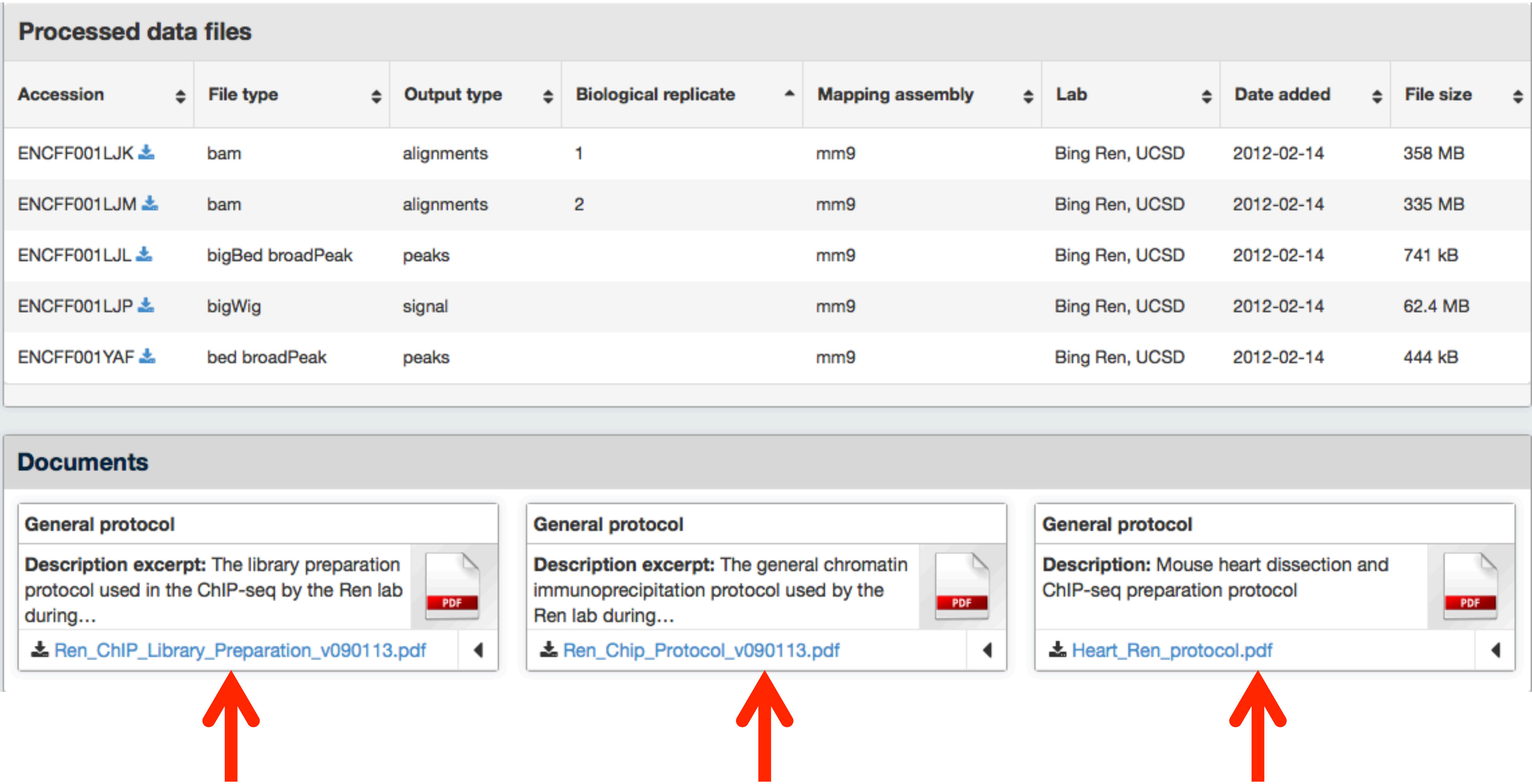
Raw data files

Accession	File type	Biological replicate	Library	Run type	Read	Lab	Date added	File size
ENCFF001LJW	fastq	1	ENCLB676HUM	SE 36nt		Bing Ren, UCSD	2012-02-14	534 MB
ENCFF001LJV	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
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

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

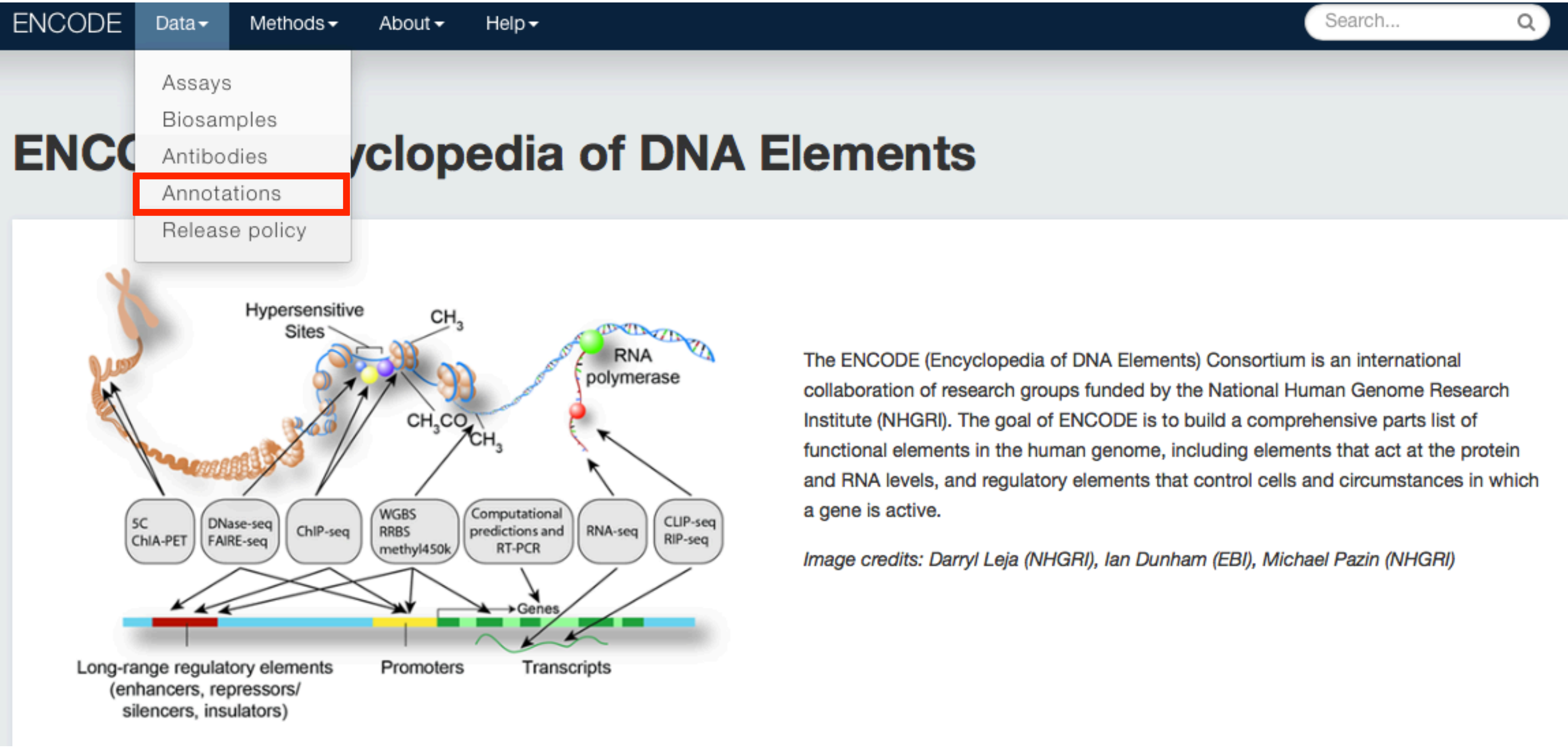
Description: Mouse heart dissection and ChIP-seq preparation protocol

Heart_Ren_protocol.pdf

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)

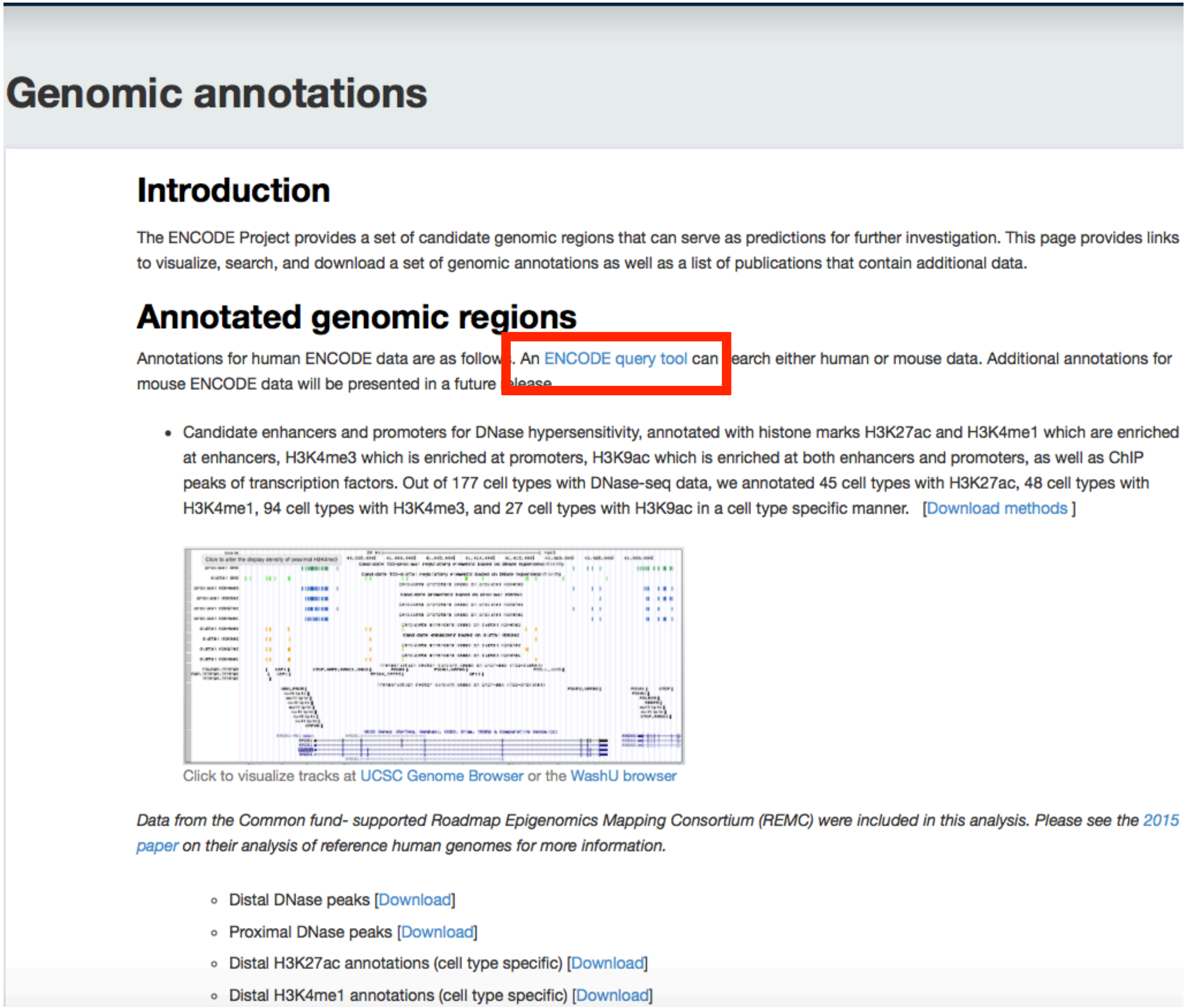
ENCODE annotations



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



Genomic annotations

Introduction

The ENCODE Project provides a set of candidate genomic regions that can serve as predictions for further investigation. This page provides links to visualize, search, and download a set of genomic annotations as well as a list of publications that contain additional data.

Annotated genomic regions

Annotations for human ENCODE data are as follow. An ENCODE query tool can earch either human or mouse data. Additional annotations for mouse ENCODE data will be presented in a future elease.

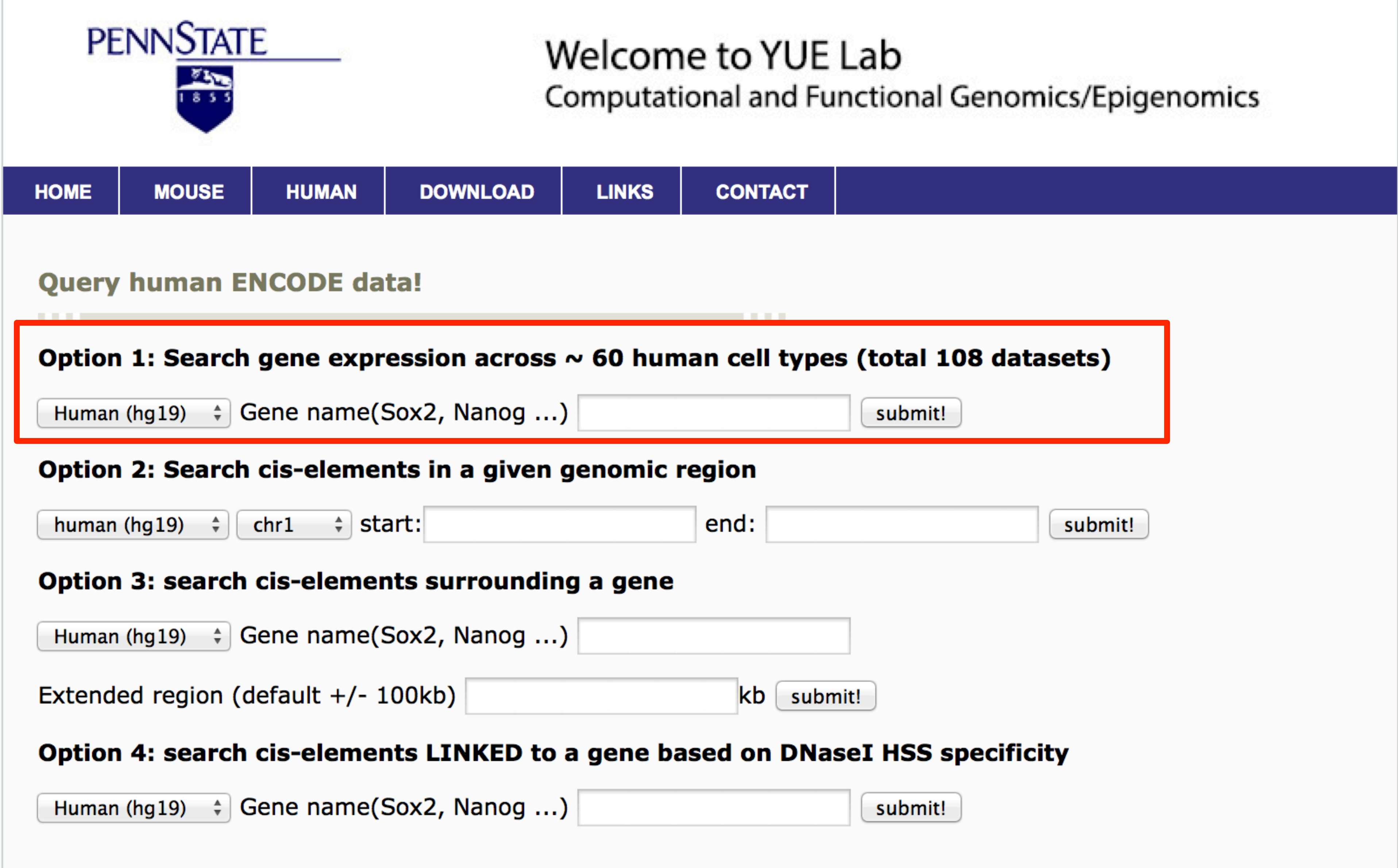
- Candidate enhancers and promoters for DNase hypersensitivity, annotated with histone marks H3K27ac and H3K4me1 which are enriched at enhancers, H3K4me3 which is enriched at promoters, H3K9ac which is enriched at both enhancers and promoters, as well as ChIP peaks of transcription factors. Out of 177 cell types with DNase-seq data, we annotated 45 cell types with H3K27ac, 48 cell types with H3K4me1, 94 cell types with H3K4me3, and 27 cell types with H3K9ac in a cell type specific manner. [Download methods]

Click to visualize tracks at UCSC Genome Browser or the WashU browser

Data from the Common fund- supported Roadmap Epigenomics Mapping Consortium (REMC) were included in this analysis. Please see the 2015 paper on their analysis of reference human genomes for more information.

  - Distal DNase peaks [Download]
  - Proximal DNase peaks [Download]
  - Distal H3K27ac annotations (cell type specific) [Download]
  - Distal H3K4me1 annotations (cell type specific) [Download]

Click human tab, and search for the expression for Sox2



Welcome to YUE Lab
Computational and Functional Genomics/Epigenomics

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) chr1 start: end:

Option 3: search cis-elements surrounding a gene

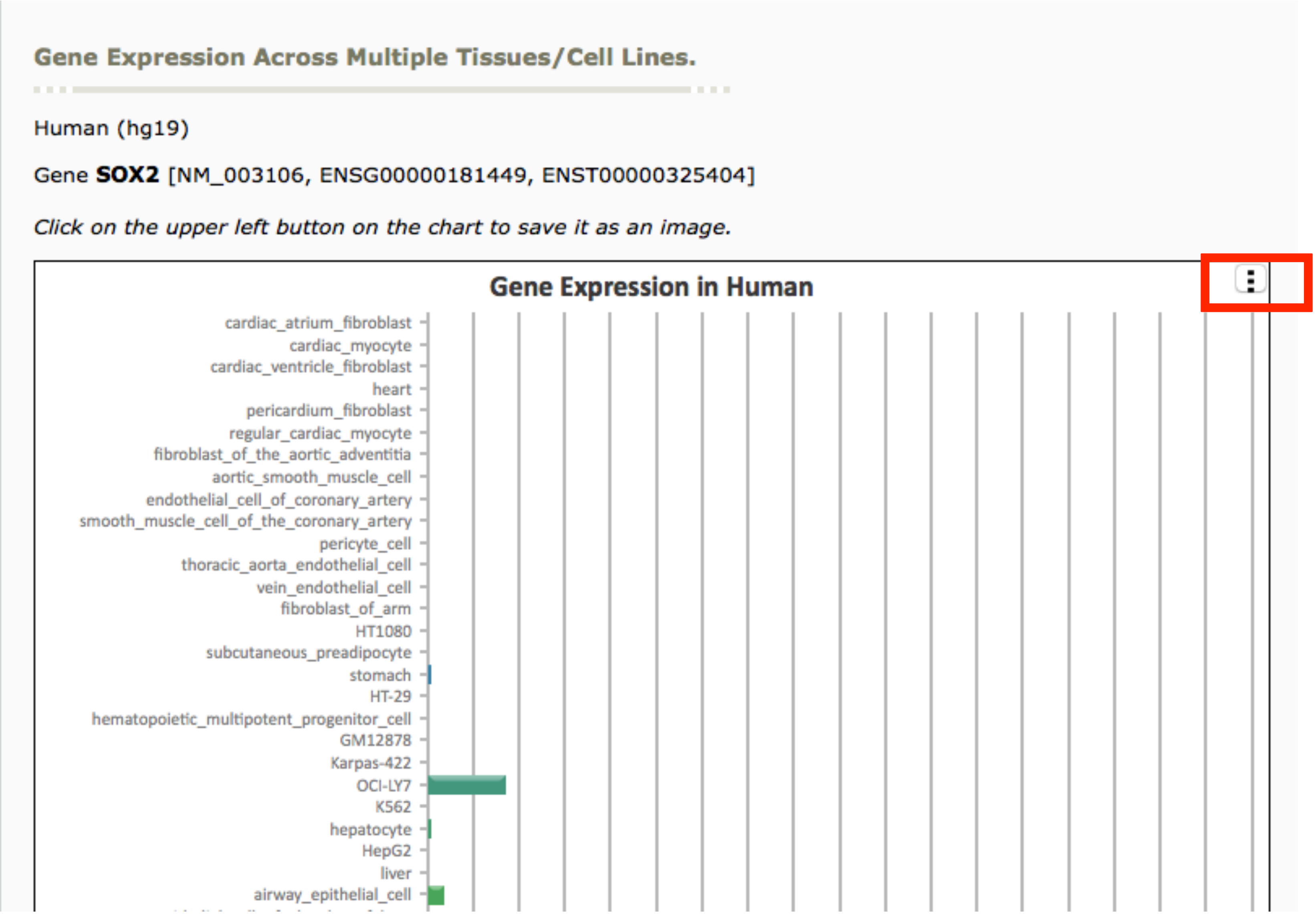
Human (hg19) Gene name(Sox2, Nanog ...)

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 ...)

Visualize and save the gene expression profile



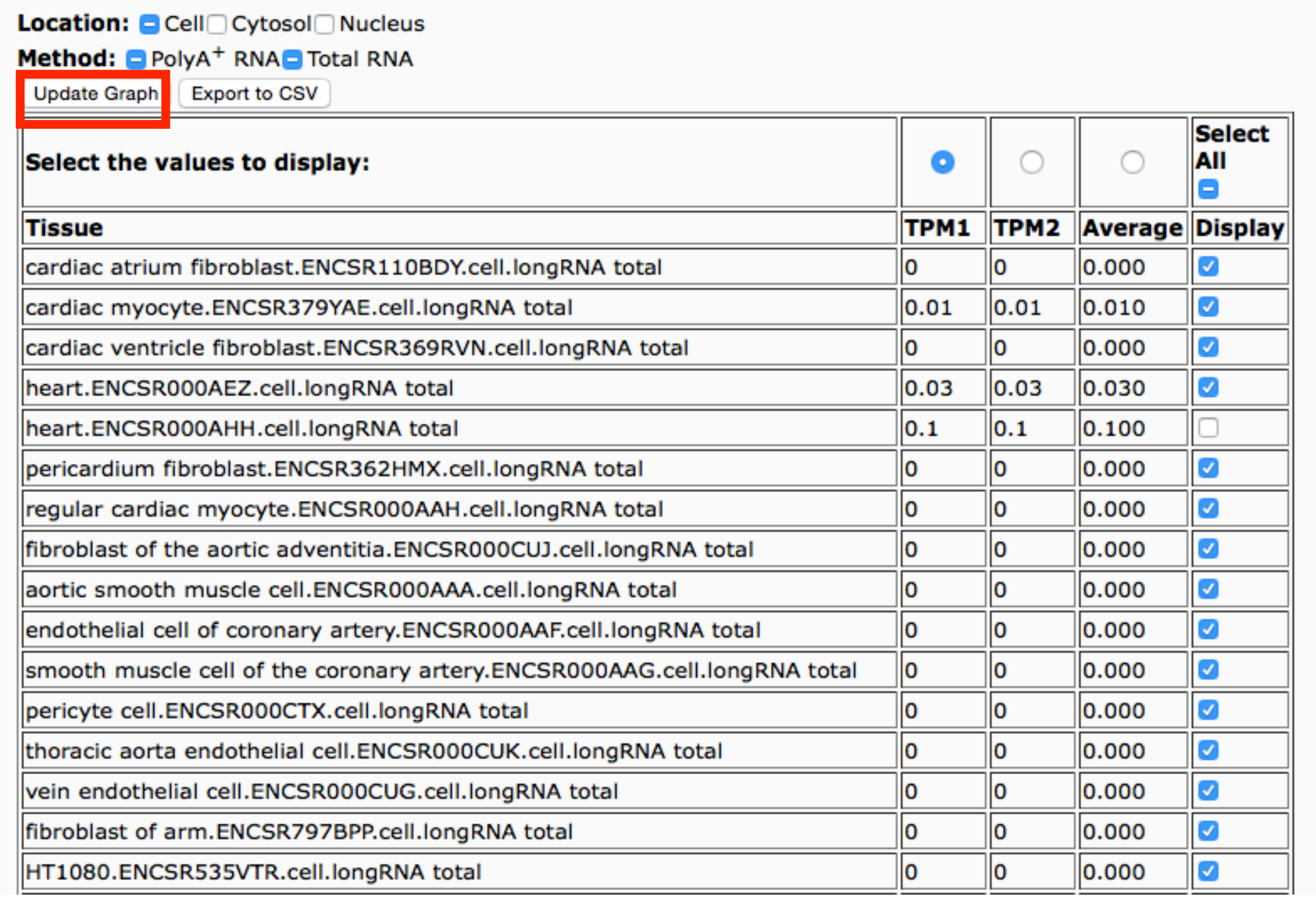
Gene Expression Across Multiple Tissues/Cell Lines.

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!

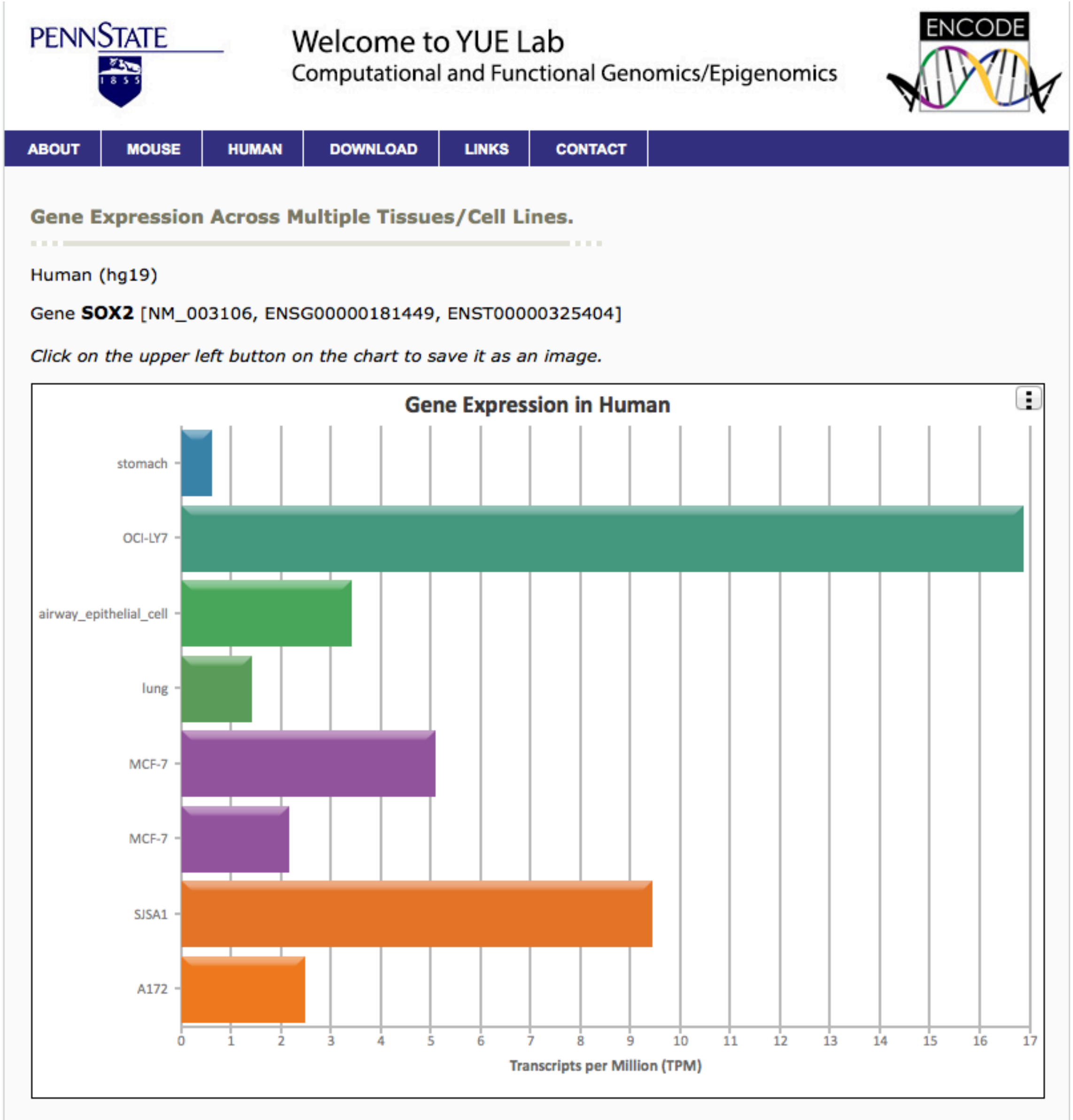


Location: ⊟ Cell ☐ Cytosol ☐ Nucleus

Method: ⊟ PolyA$^+$ RNA ⊟ Total RNA

Select the values to display:	●	○	○	**Select All** ⊟
Tissue	**TPM1**	**TPM2**	**Average**	**Display**
cardiac atrium fibroblast.ENCSR110BDY.cell.longRNA total	0	0	0.000	☑
cardiac myocyte.ENCSR379YAE.cell.longRNA total	0.01	0.01	0.010	☑
cardiac ventricle fibroblast.ENCSR369RVN.cell.longRNA total	0	0	0.000	☑
heart.ENCSR000AEZ.cell.longRNA total	0.03	0.03	0.030	☑
heart.ENCSR000AHH.cell.longRNA total	0.1	0.1	0.100	☐
pericardium fibroblast.ENCSR362HMX.cell.longRNA total	0	0	0.000	☑
regular cardiac myocyte.ENCSR000AAH.cell.longRNA total	0	0	0.000	☑
fibroblast of the aortic adventitia.ENCSR000CUJ.cell.longRNA total	0	0	0.000	☑
aortic smooth muscle cell.ENCSR000AAA.cell.longRNA total	0	0	0.000	☑
endothelial cell of coronary artery.ENCSR000AAF.cell.longRNA total	0	0	0.000	☑
smooth muscle cell of the coronary artery.ENCSR000AAG.cell.longRNA total	0	0	0.000	☑
pericyte cell.ENCSR000CTX.cell.longRNA total	0	0	0.000	☑
thoracic aorta endothelial cell.ENCSR000CUK.cell.longRNA total	0	0	0.000	☑
vein endothelial cell.ENCSR000CUG.cell.longRNA total	0	0	0.000	☑
fibroblast of arm.ENCSR797BPP.cell.longRNA total	0	0	0.000	☑
HT1080.ENCSR535VTR.cell.longRNA total	0	0	0.000	☑

You can choose the cell types to display!



Welcome to YUE Lab

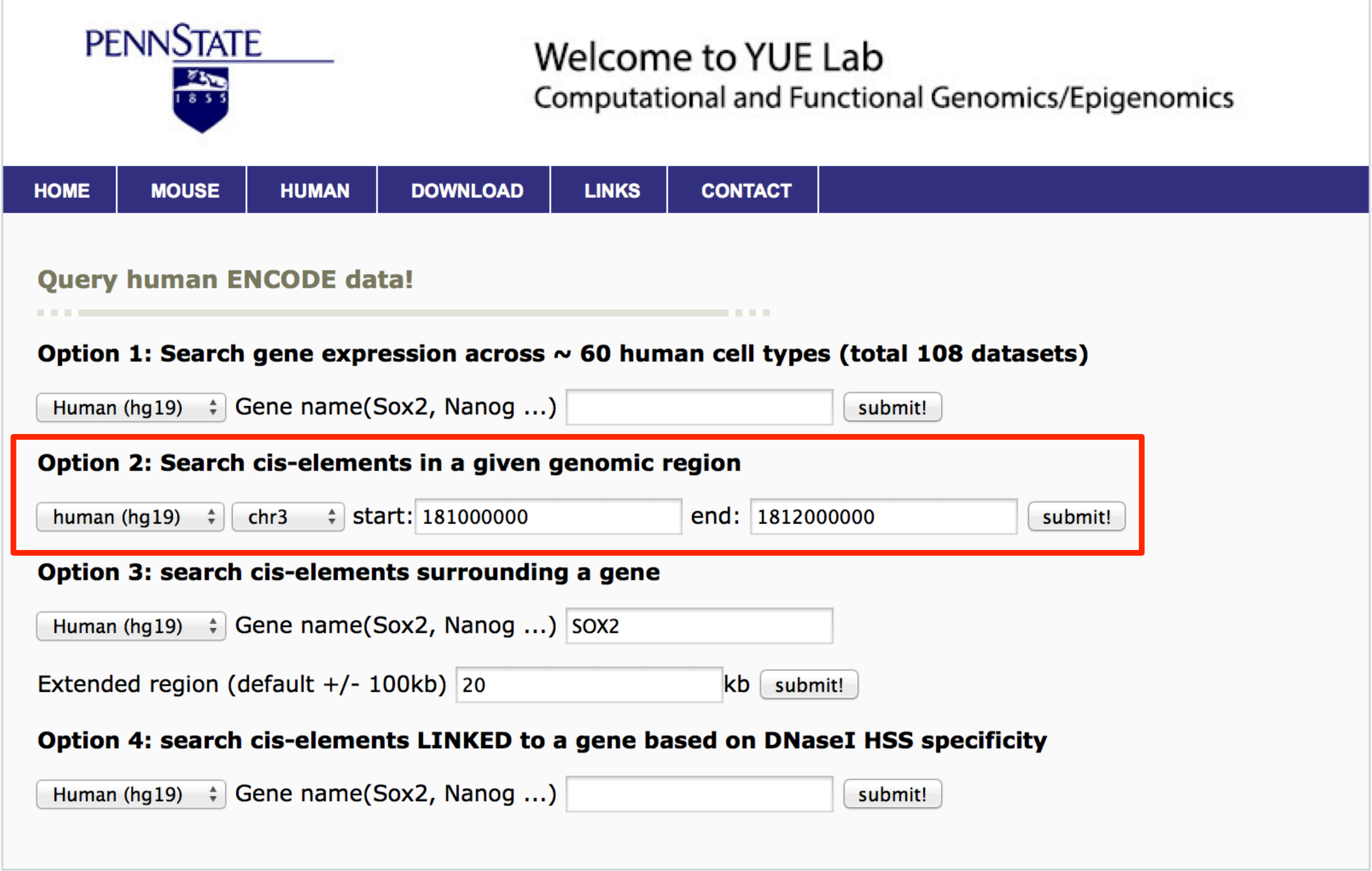
Computational and Functional Genomics/Epigenomics

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

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) chr3 start: 181000000 end: 1812000000 submit!

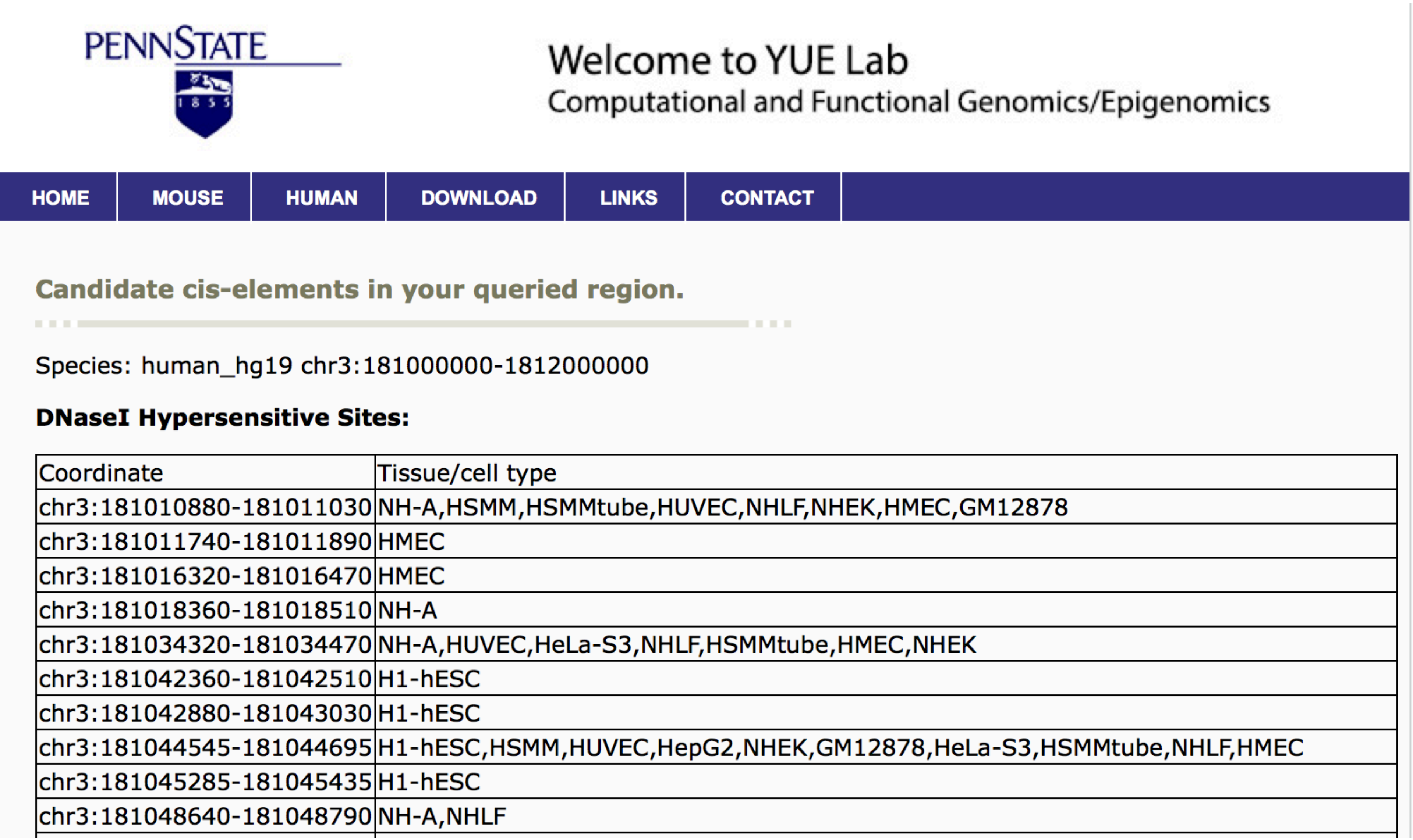
Option 3: search cis-elements surrounding a gene

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

Extended region (default +/- 100kb) 20 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

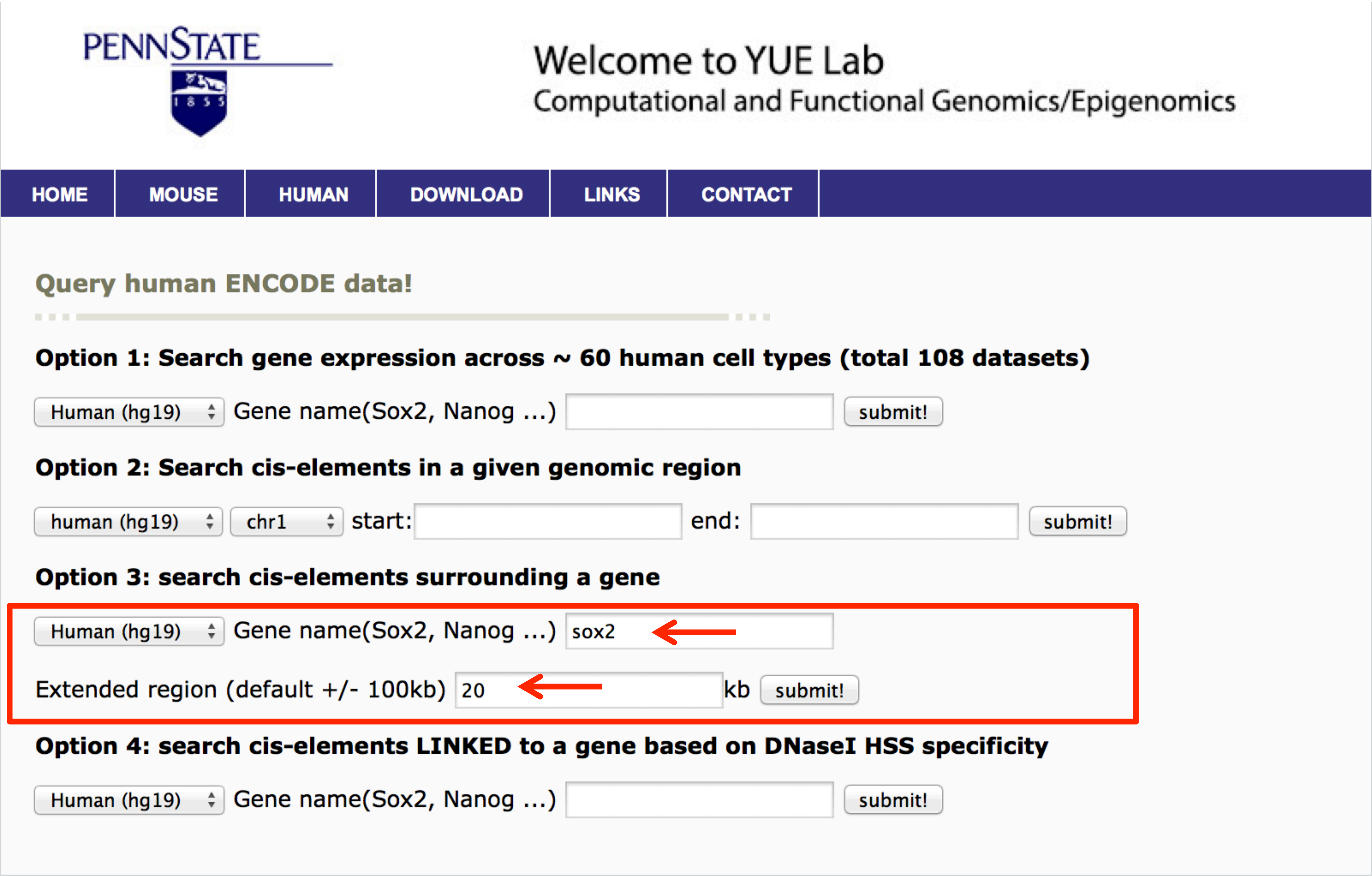
Computational and Functional Genomics/Epigenomics

Candidate cis-elements in your queried region.

Species: human_hg19 chr3:181000000-1812000000

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

Query human ENCODE data!

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

Gene name(Sox2, Nanog ...)

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

start:

end:

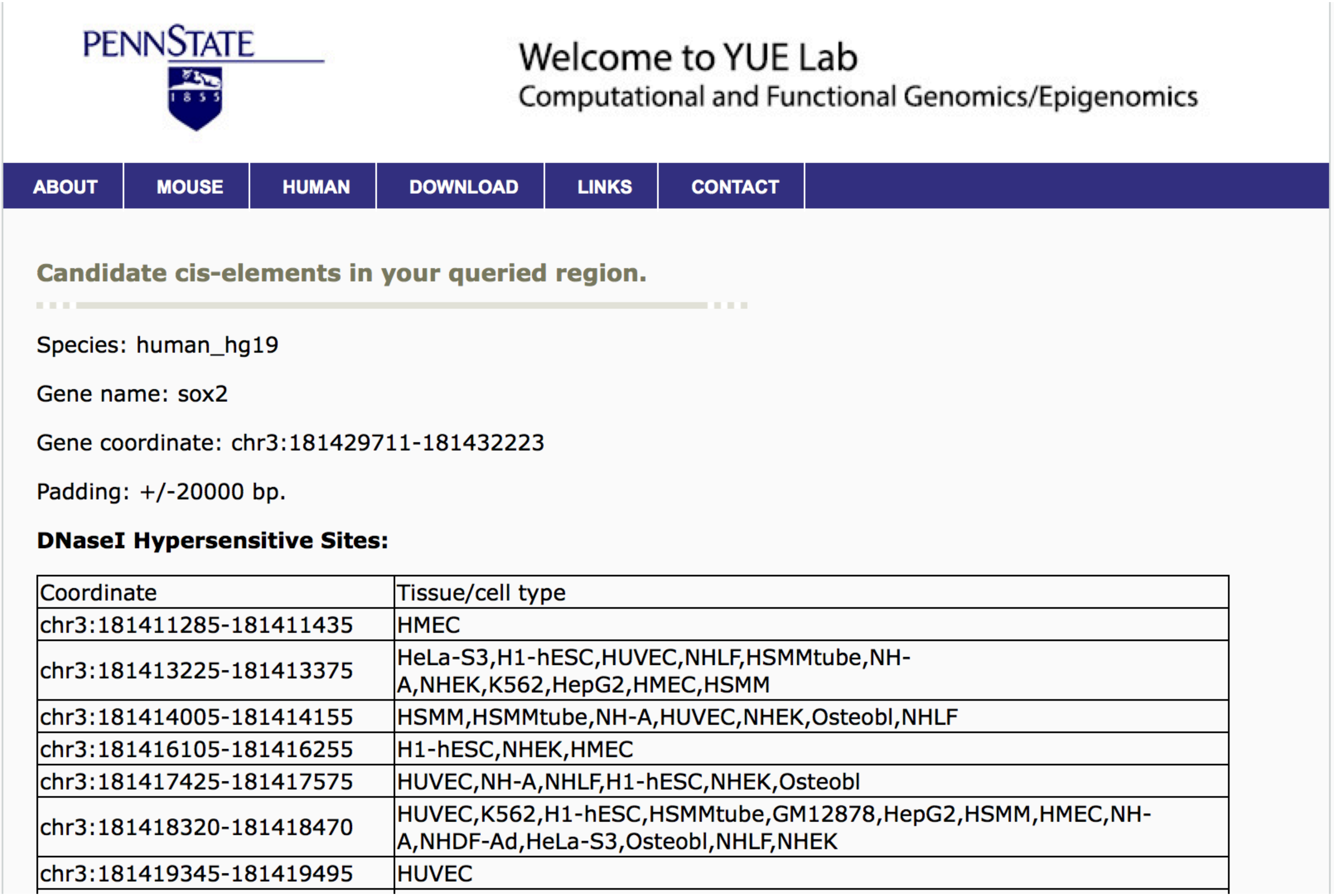
Option 3: search cis-elements surrounding a gene

Gene name(Sox2, Nanog ...) sox2

Extended region (default +/- 100kb) 20 kb

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

Gene name(Sox2, Nanog ...)



Welcome to YUE Lab

Computational and Functional Genomics/Epigenomics

Candidate cis-elements in your queried region.

Species: human_hg19

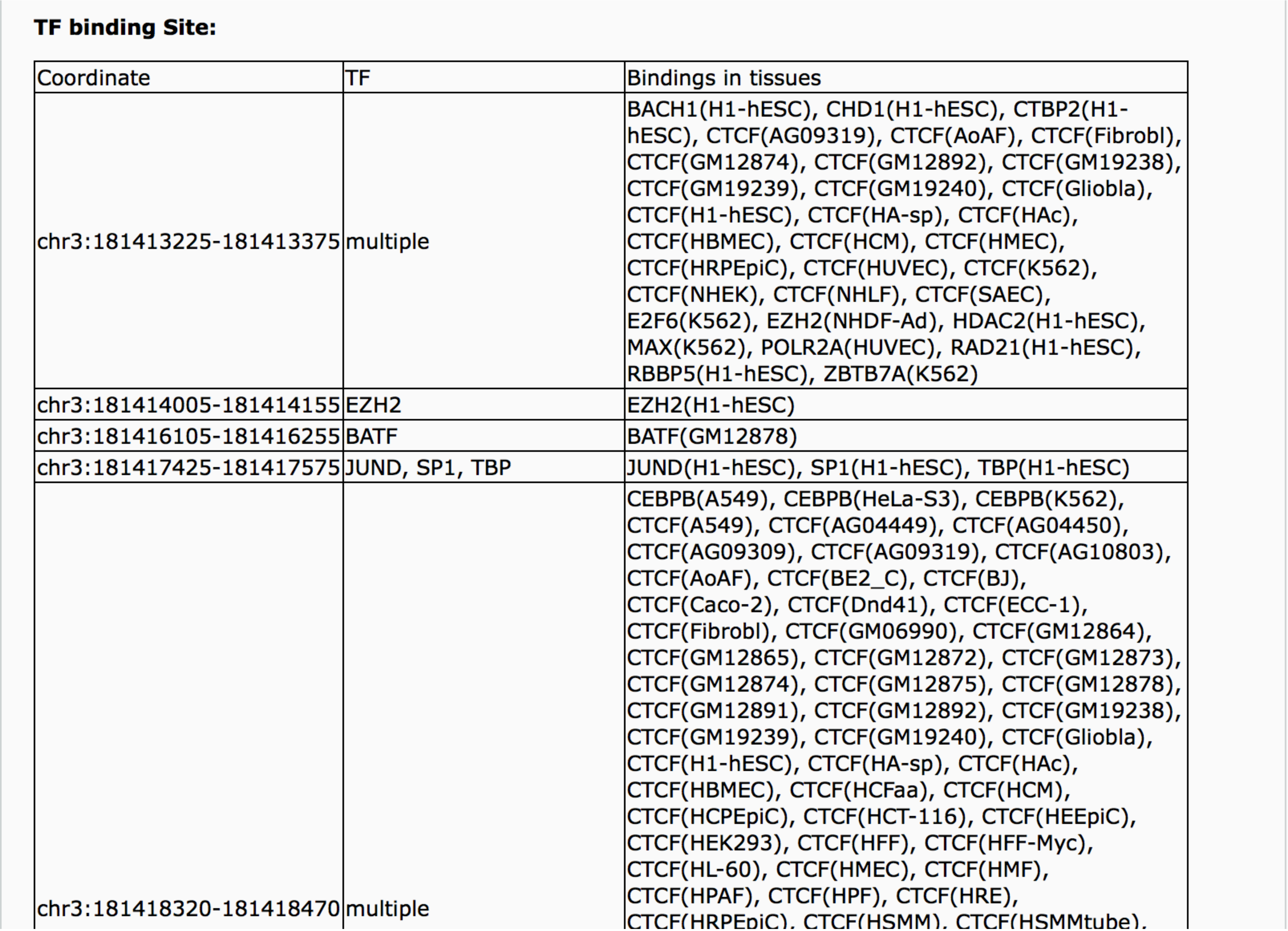
Gene name: sox2

Gene coordinate: chr3:181429711-181432223

Padding: +/-20000 bp.

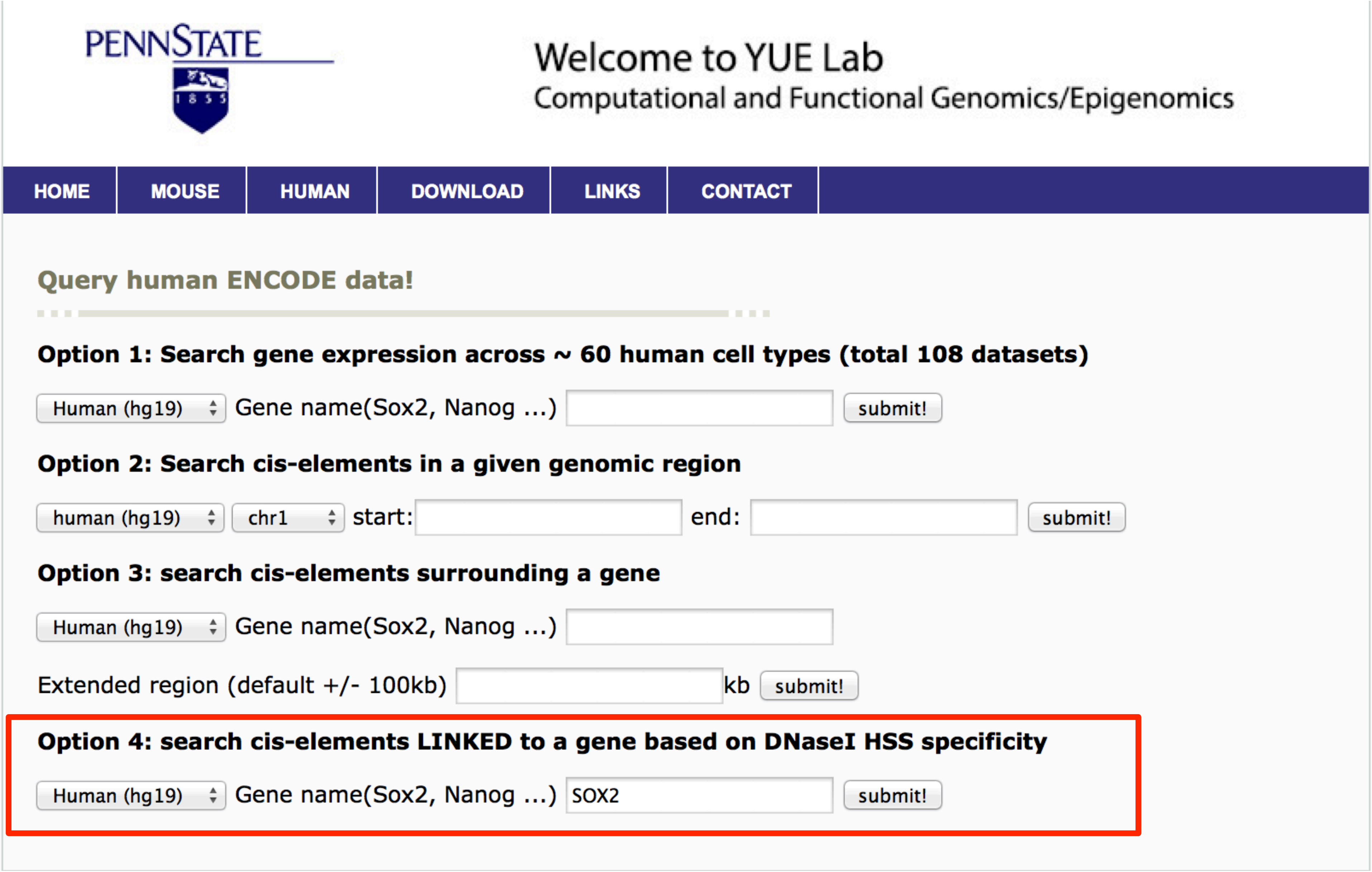
DNaseI Hypersensitive Sites:

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

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) chr1 start: end:

Option 3: search cis-elements surrounding a gene

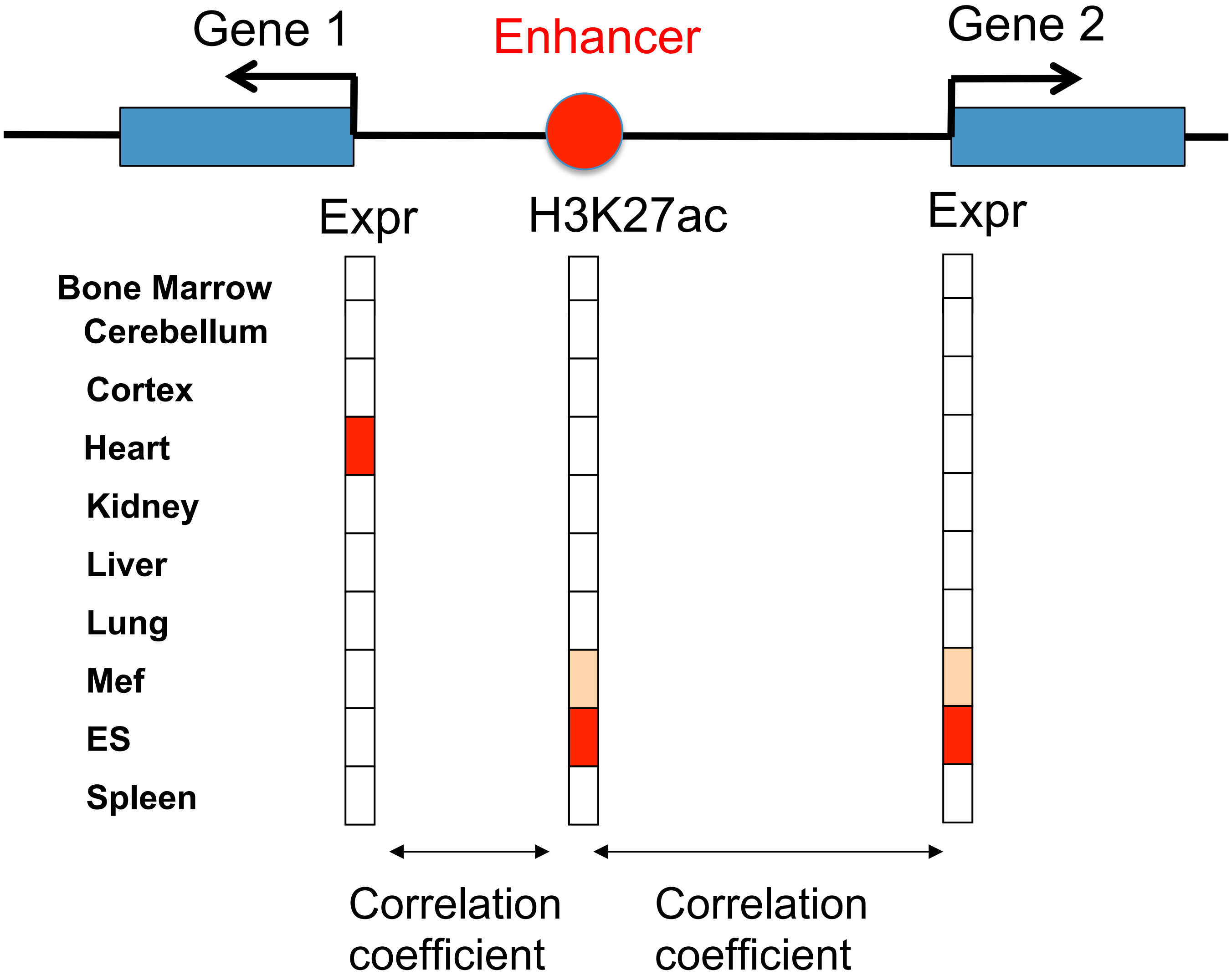
Human (hg19) Gene name(Sox2, Nanog ...)

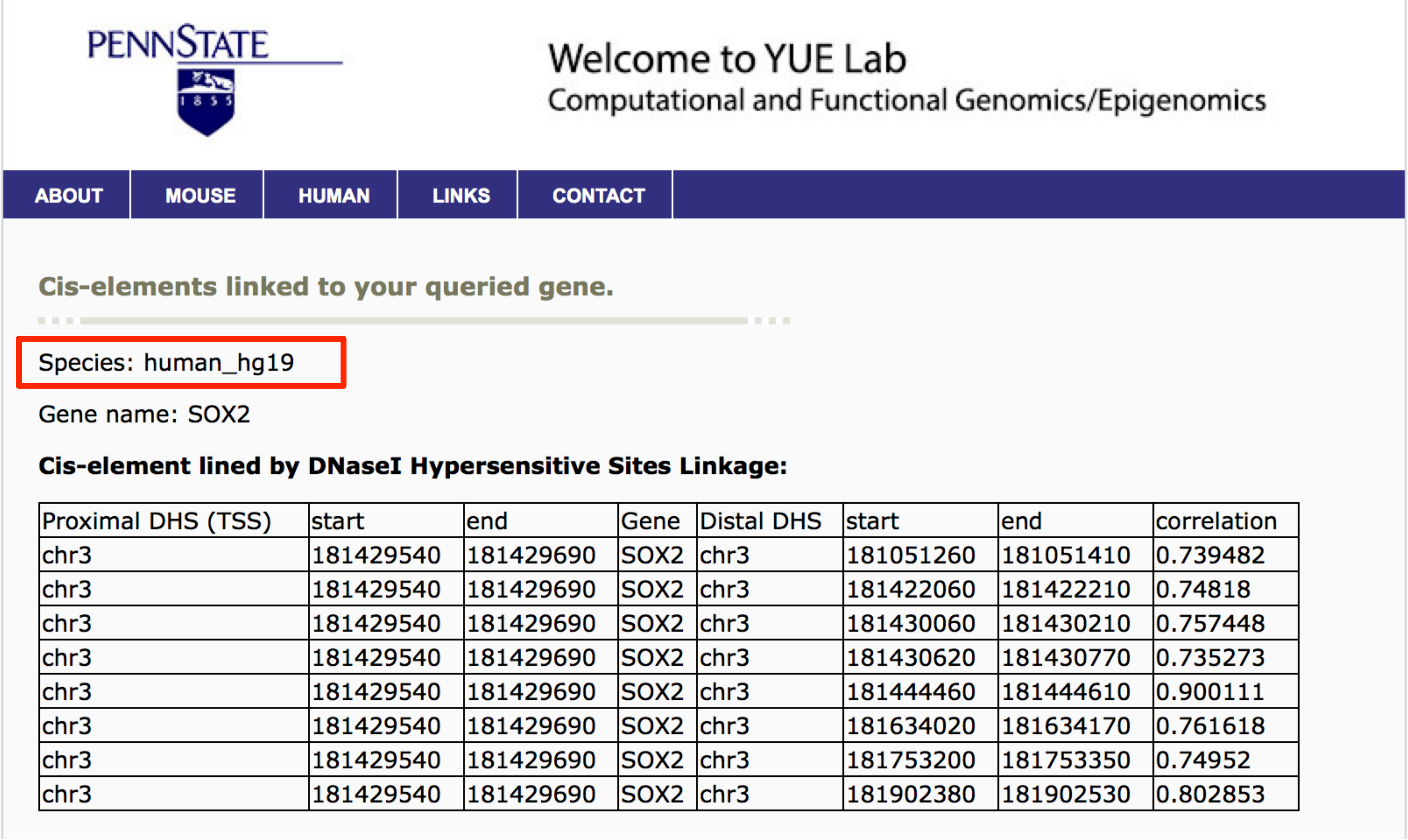
Extended region (default +/- 100kb) kb

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

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

Linking enhancers to target promoters





Welcome to YUE Lab

Computational and Functional Genomics/Epigenomics

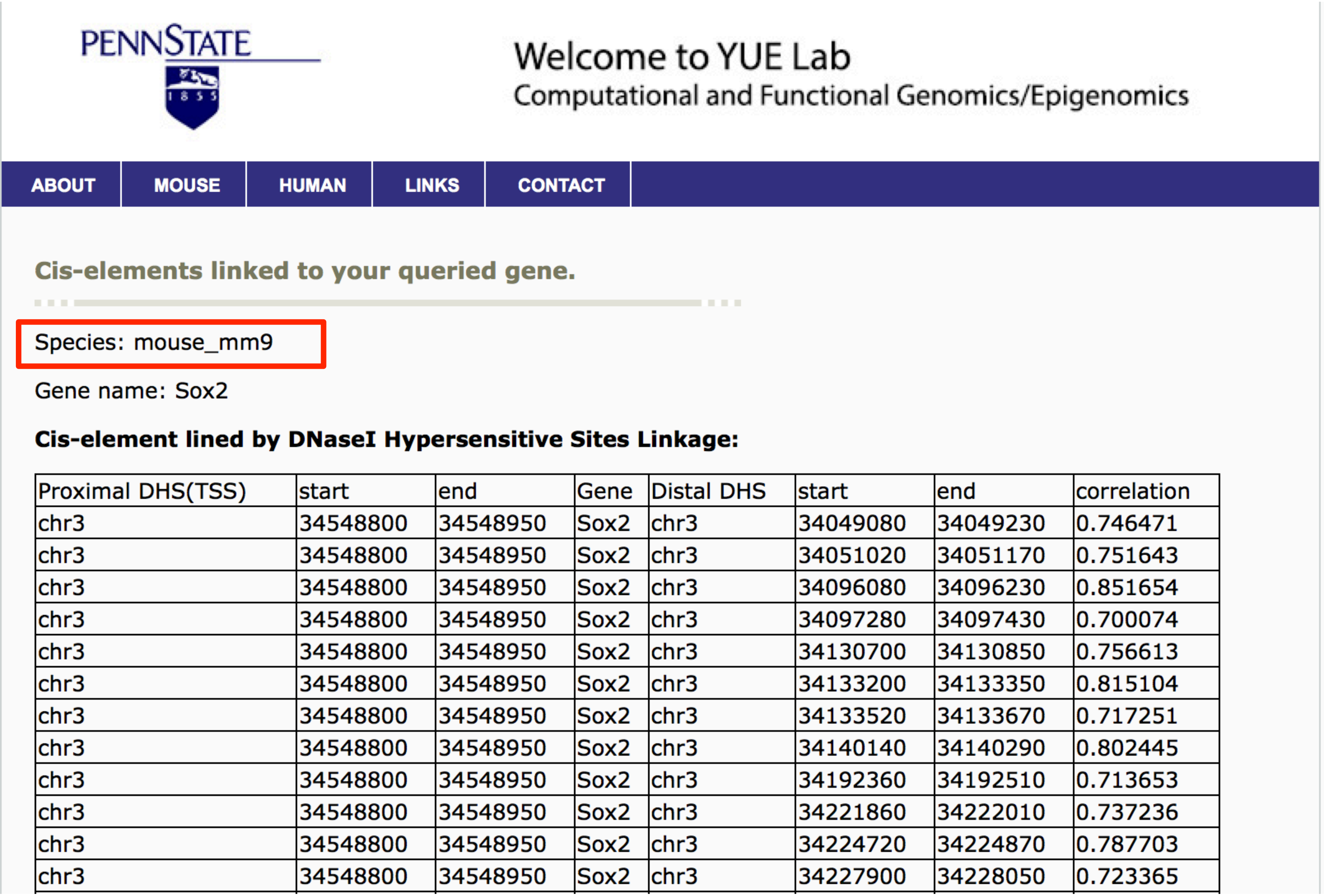
Cis-elements linked to your queried gene.

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



Welcome to YUE Lab

Computational and Functional Genomics/Epigenomics

Cis-elements linked to your queried gene.

Species: mouse_mm9

Gene name: Sox2

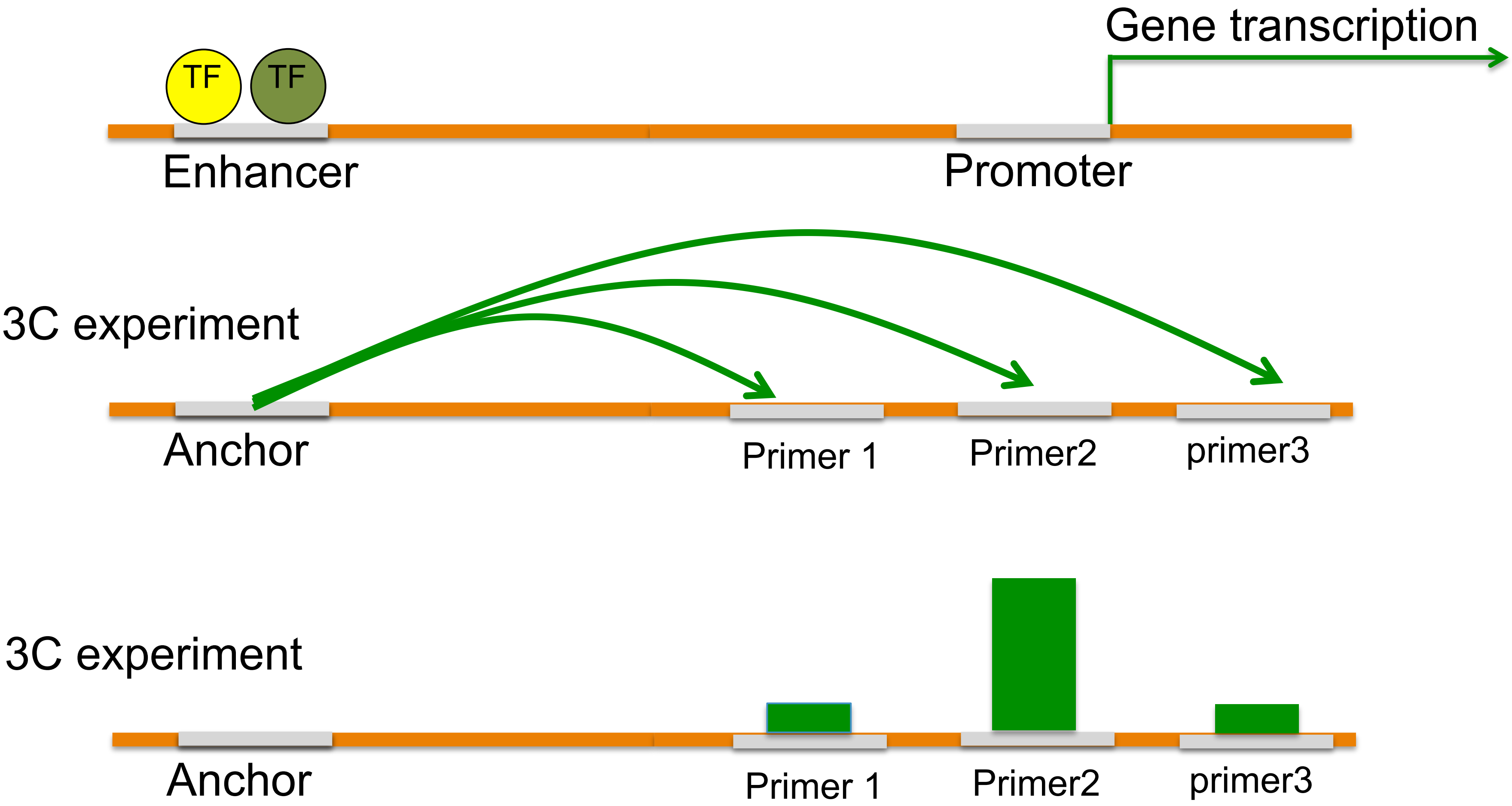
Cis-element lined by DNaseI Hypersensitive Sites Linkage:

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)

Chromatin conformation capture (3C) experiment



Gene transcription

TF

TF

Enhancer

Promoter

3C experiment

Anchor

Primer 1

Primer2

primer3

3C experiment

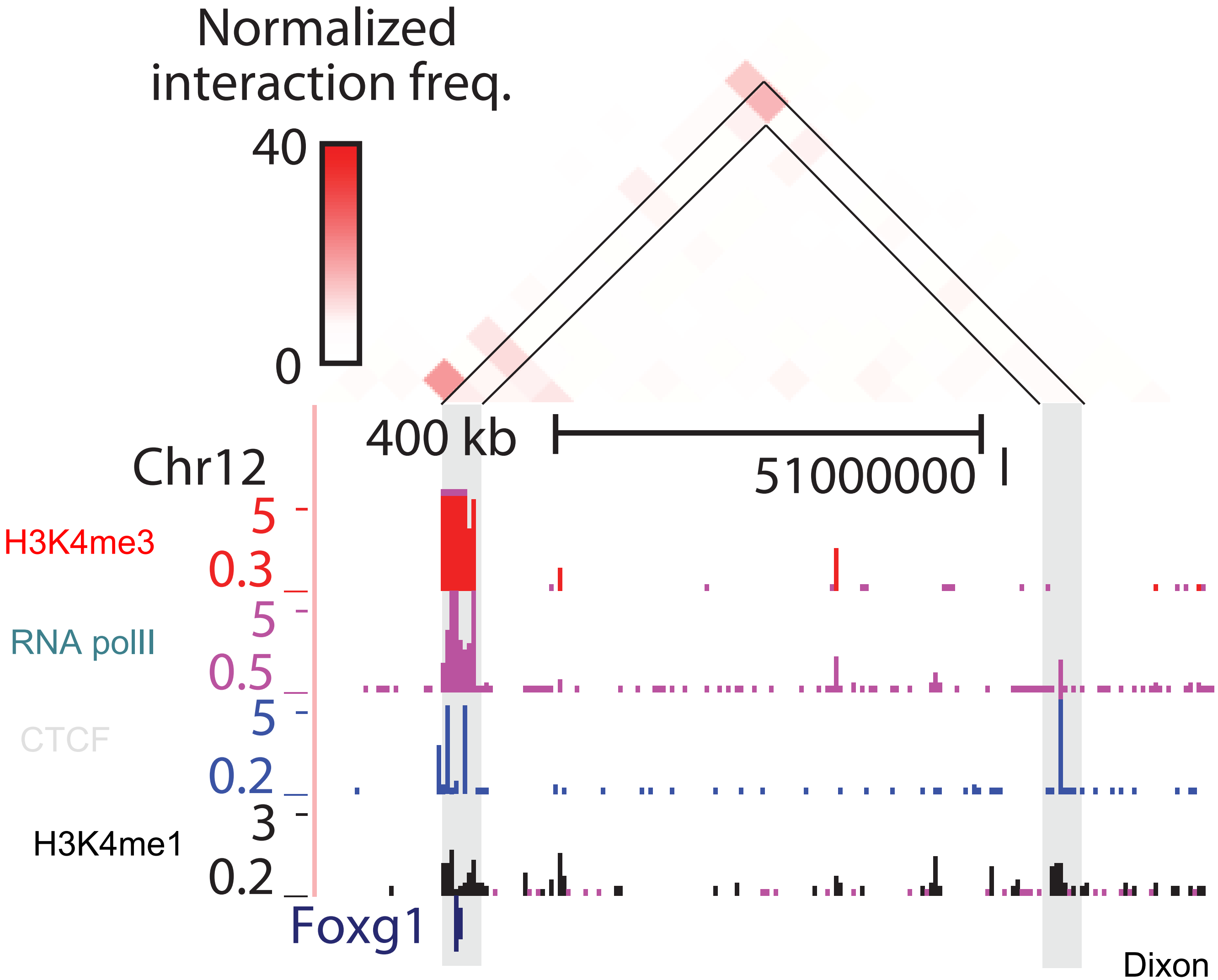
Anchor

Primer 1

Primer2

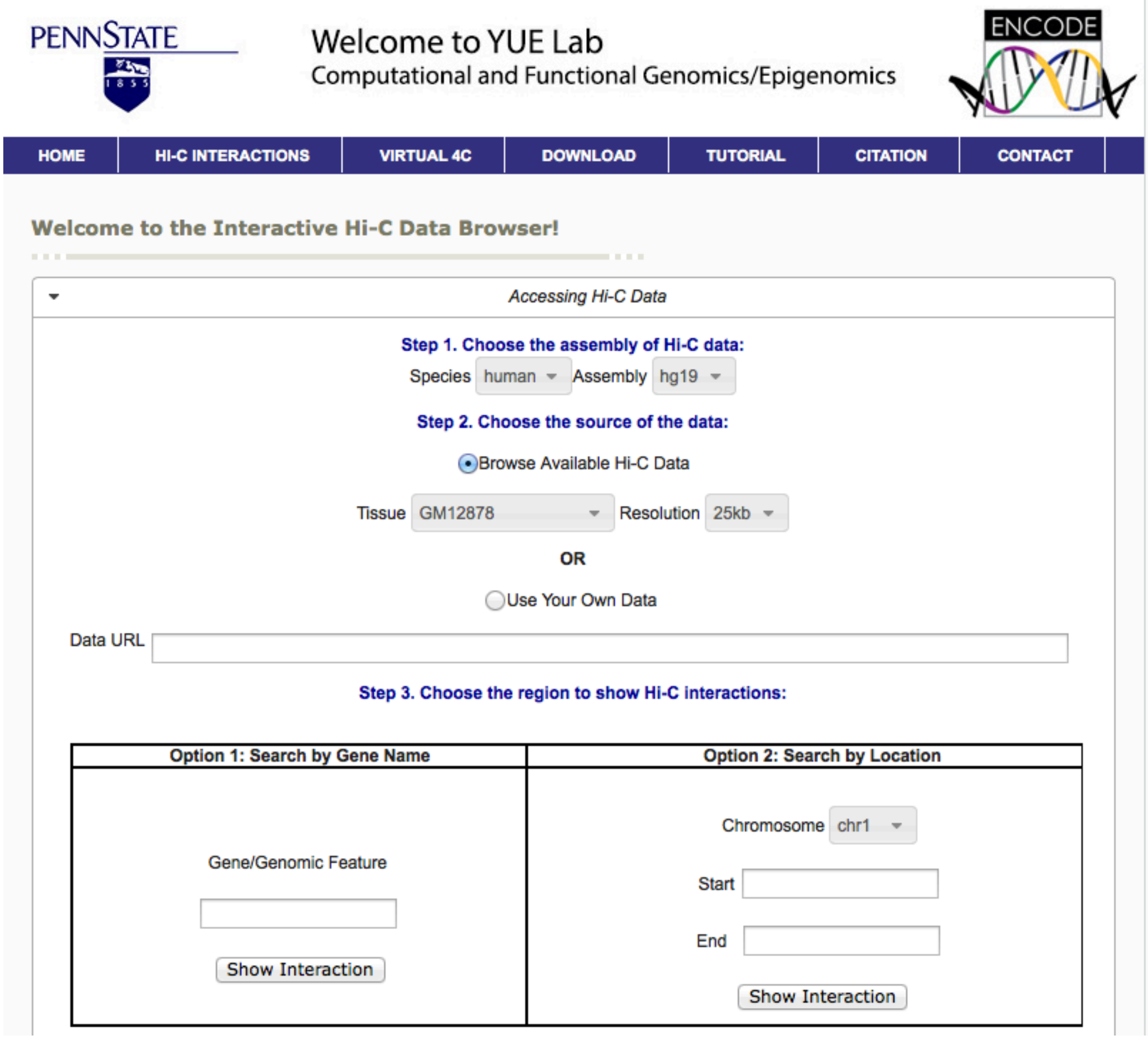
primer3

Hi-C measures genome-wide chromatin interaction



Dixon *et al.* nature, 2012

http://3dgenome.org



Welcome to YUE Lab
Computational and Functional Genomics/Epigenomics

Welcome to the Interactive Hi-C Data Browser!

Accessing Hi-C Data

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

Species human Assembly hg19

Step 2. Choose the source of the data:

● Browse Available Hi-C Data

Tissue GM12878 Resolution 25kb

OR

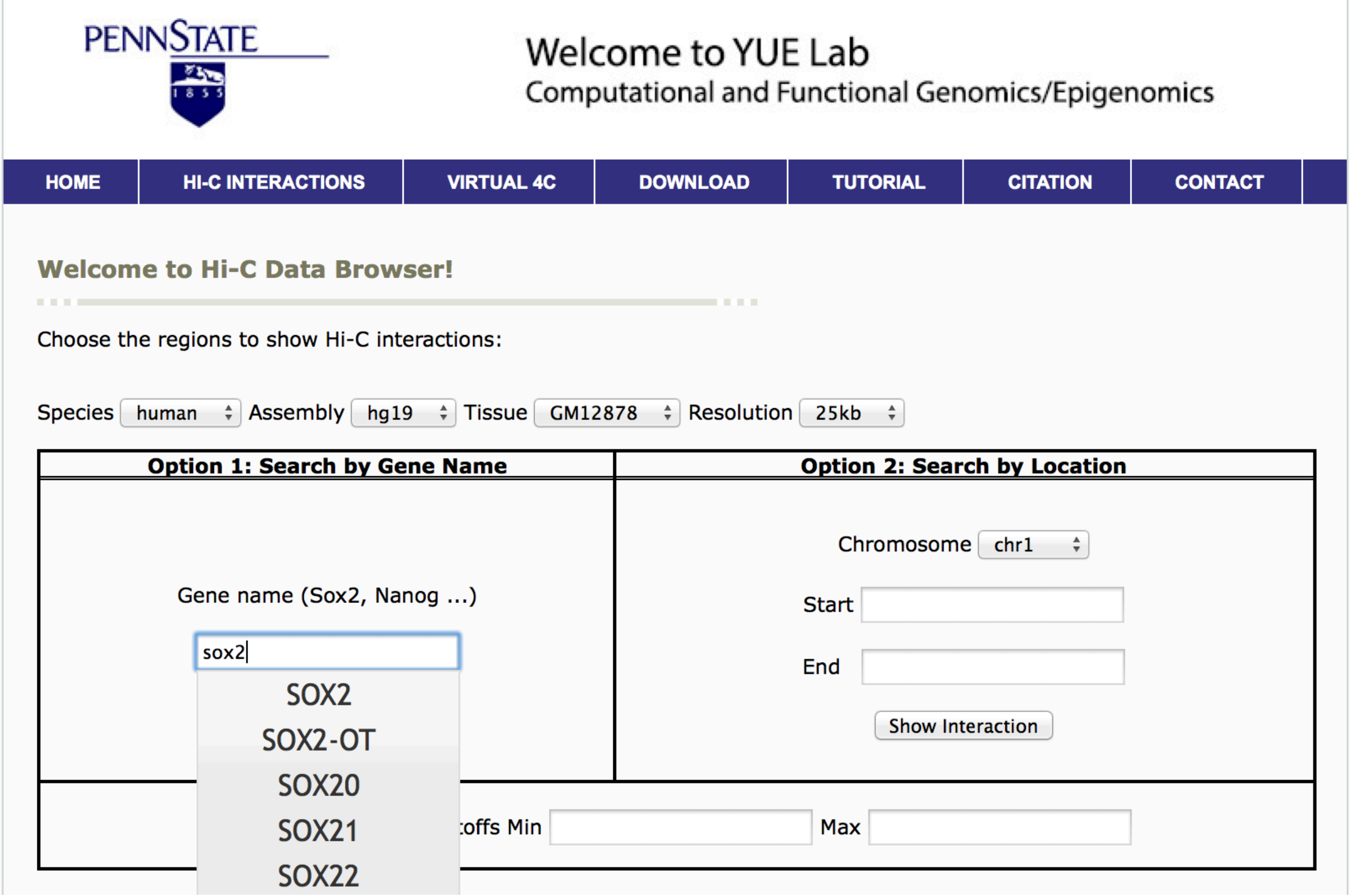
○ Use Your Own Data

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	Chromosome chr1 Start End

Users can query by gene name or genomic loci



Welcome to YUE Lab

Computational and Functional Genomics/Epigenomics

Welcome to Hi-C Data Browser!

Choose the regions to show Hi-C interactions:

Species: human | Assembly: hg19 | Tissue: GM12878 | Resolution: 25kb

Option 1: Search by Gene Name

Gene name (Sox2, Nanog ...)

sox2

Option 2: Search by Location

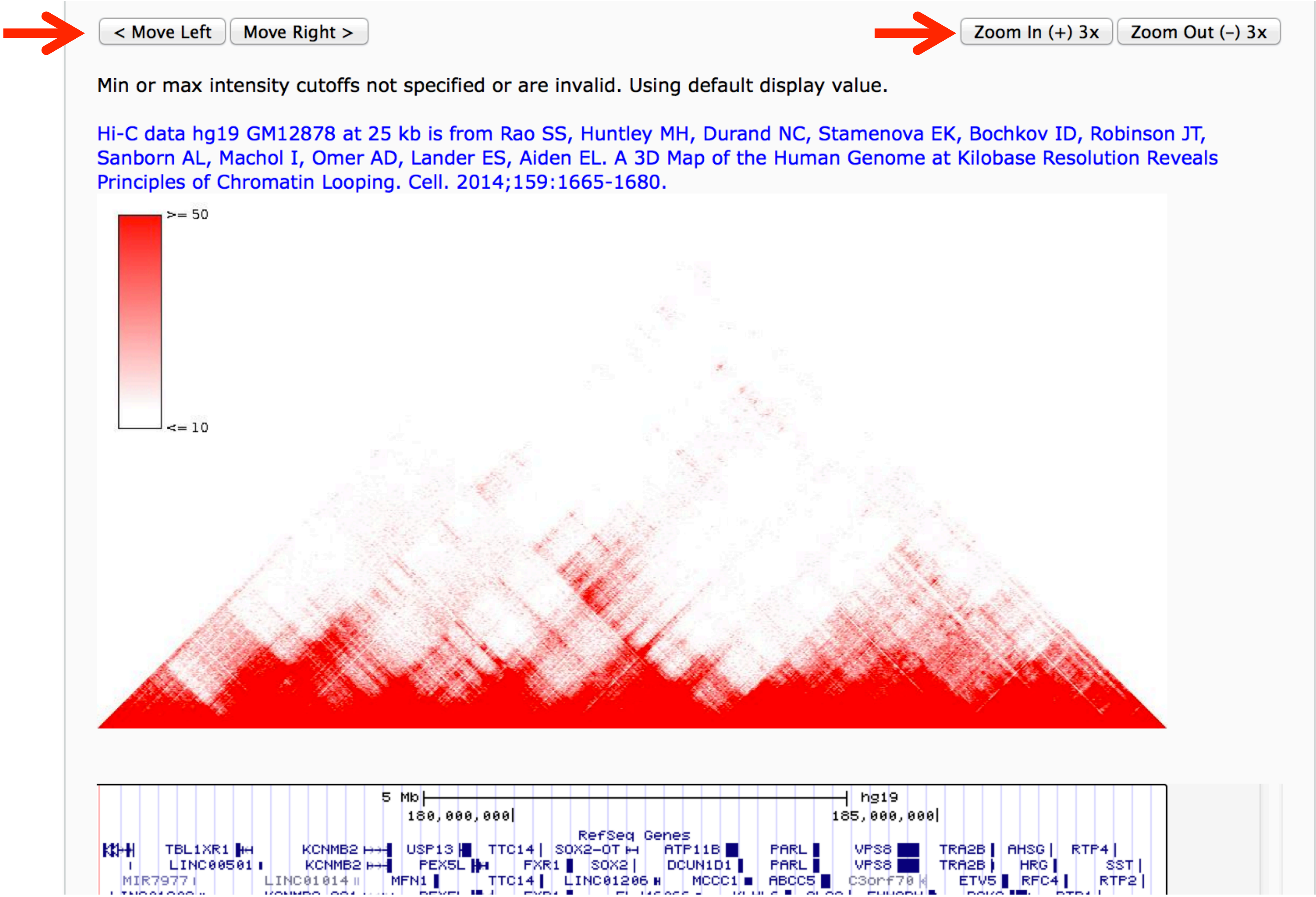
Chromosome: chr1

Start:

End:

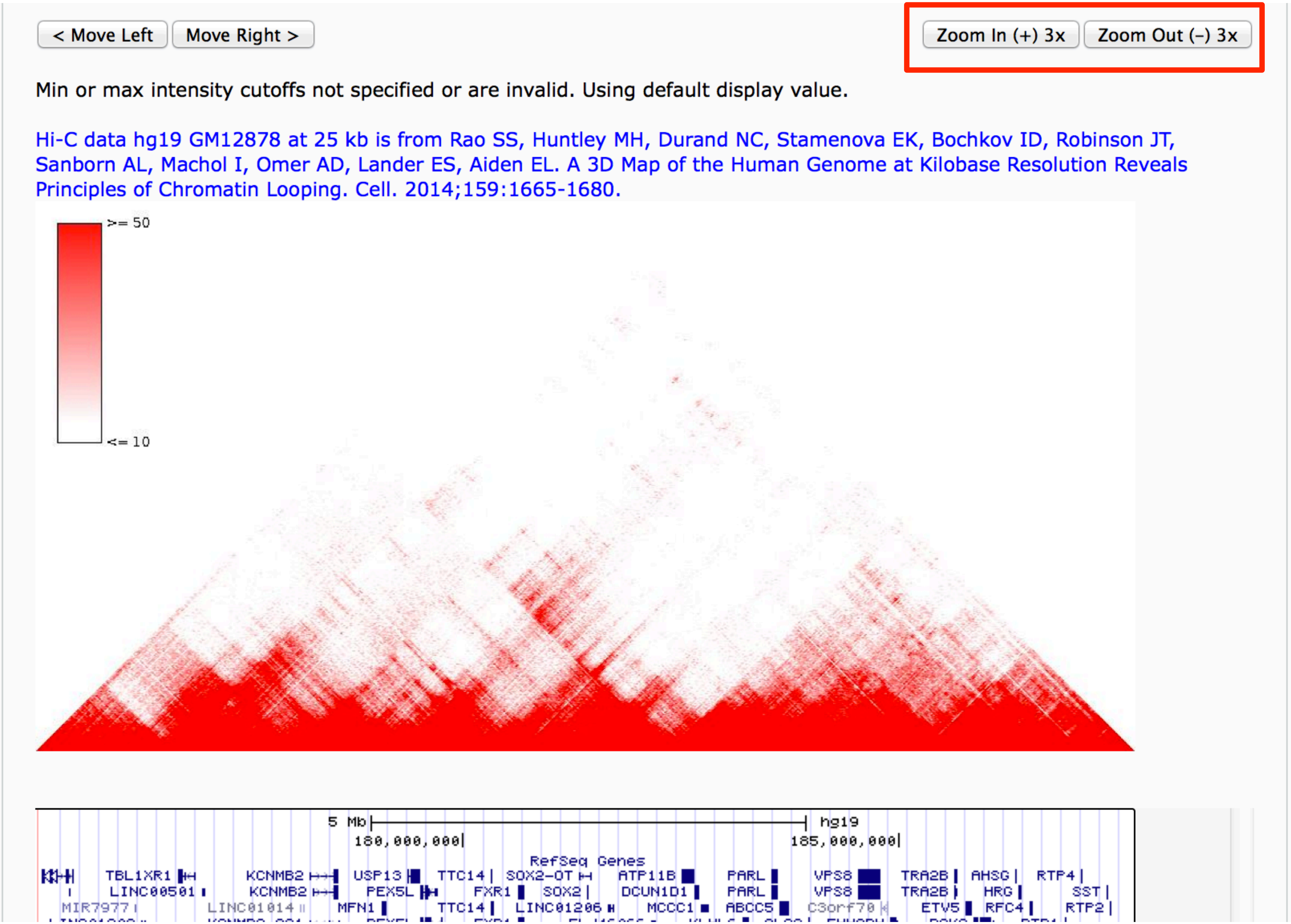
offs Min: Max:

Navigation bars to explore Hi-C data



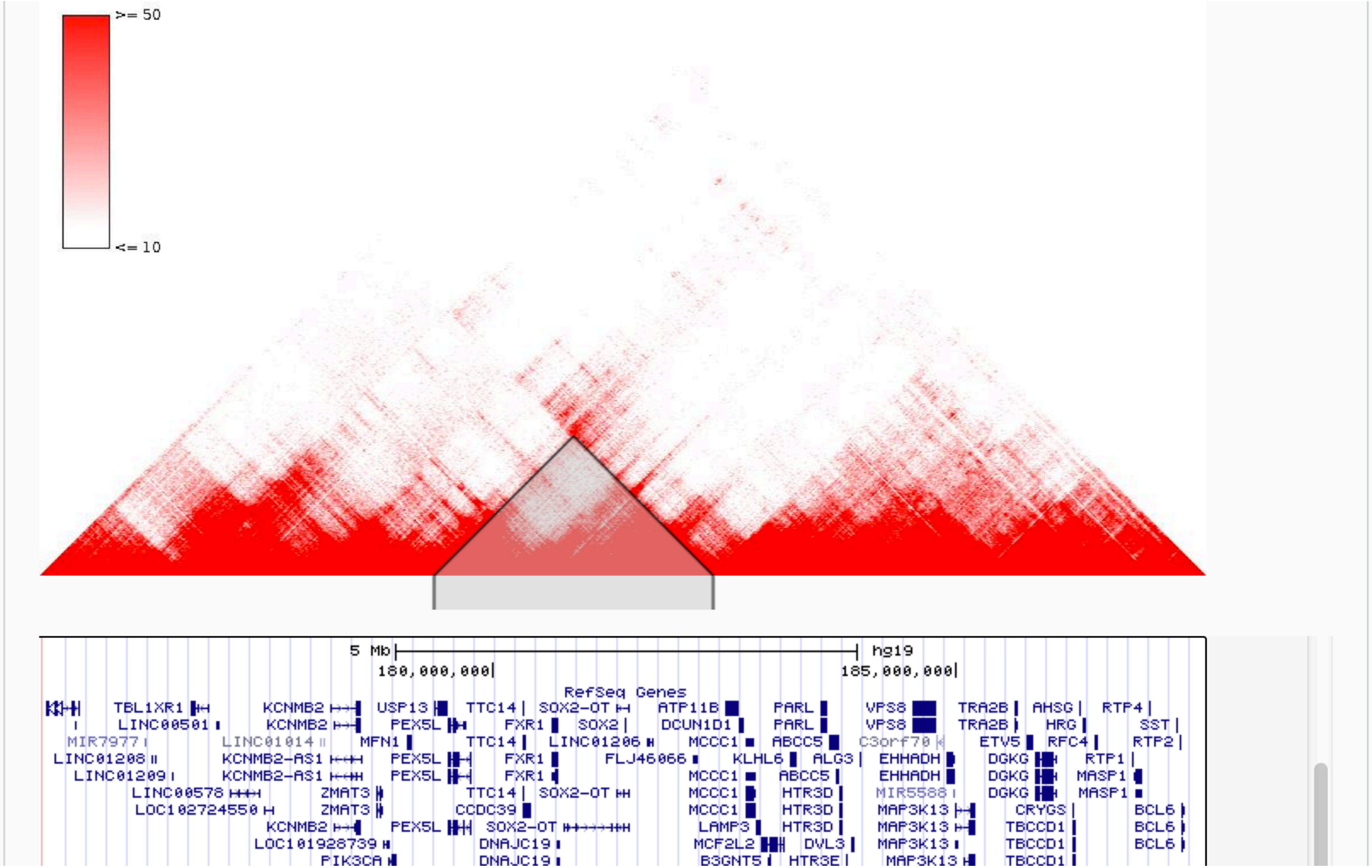
Min or max intensity cutoffs not specified or are invalid. Using default display value.

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.

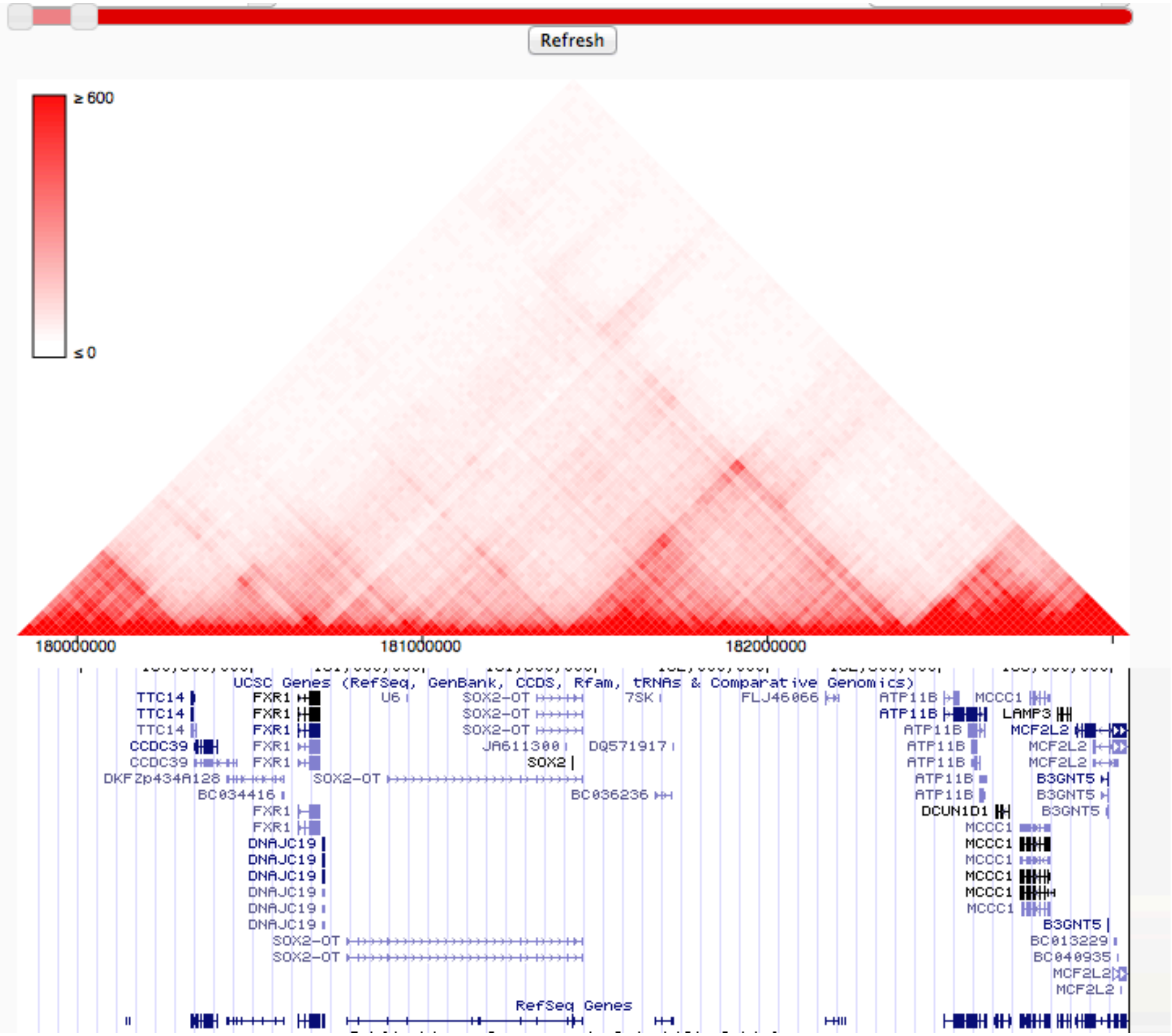


Min or max intensity cutoffs not specified or are invalid. Using default display value.

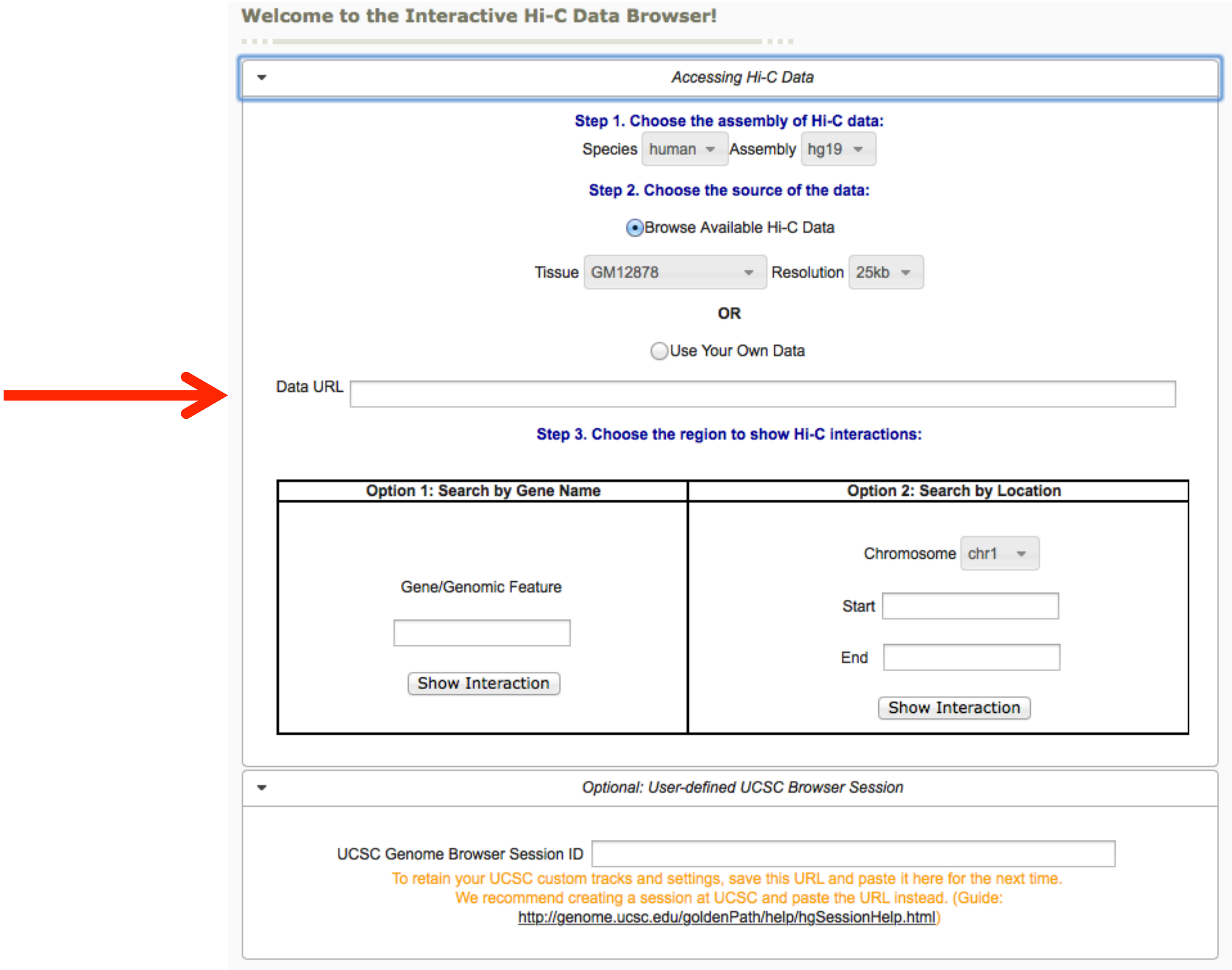
Hi-C data hg19 GM12878 at 25 kb is from Rao SS, Huntley MH, Durand NC, Stamenova EK, Bochkov ID, Robinson JT, Sanborn AL, Machol I, Omer AD, Lander ES, Aiden EL. A 3D Map of the Human Genome at Kilobase Resolution Reveals Principles of Chromatin Looping. Cell. 2014;159:1665-1680.


>= 50
<= 10
5 Mb
hg19
180,000,000
185,000,000
RefSeq Genes
TBL1XR1
KCNMB2
USP13
TTC14
SOX2-OT
ATP11B
PARL
VPS8
TRA2B
AHSG
RTP4
LINC00501
KCNMB2
PEX5L
FXR1
SOX2
DCUN1D1
PARL
VPS8
TRA2B
HRG
SST
MIR7977
LINC01014
MFN1
TTC14
LINC01206
MCCC1
ABCC5
C3orf70
ETV5
RFC4
RTP2
LINC01208
KCNMB2-AS1
PEX5L
FXR1
FLJ46066
KLHL6
ALG3
EHHADH
DGKG
RTP1
LINC01209
KCNMB2-AS1
PEX5L
FXR1
MCCC1
ABCC5
EHHADH
DGKG
MASP1
LINC00578
ZMAT3
TTC14
SOX2-OT
MCCC1
HTR3D
MIR5588
DGKG
MASP1
LOC102724550
ZMAT3
CCDC39
MCCC1
HTR3D
MAP3K13
CRYGS
BCL6
KCNMB2
PEX5L
SOX2-OT
LAMP3
HTR3D
MAP3K13
TBCCD1
BCL6
LOC101928739
DNAJC19
MCF2L2
DVL3
MAP3K13
TBCCD1
BCL6
PIK3CA
DNAJC19
B3GNT5
HTR3E
MAP3K13
TBCCD1

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 human Assembly hg19

Step 2. Choose the source of the data:

● Browse Available Hi-C Data

Tissue GM12878 Resolution 25kb

OR

○ Use Your Own Data

Data URL

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

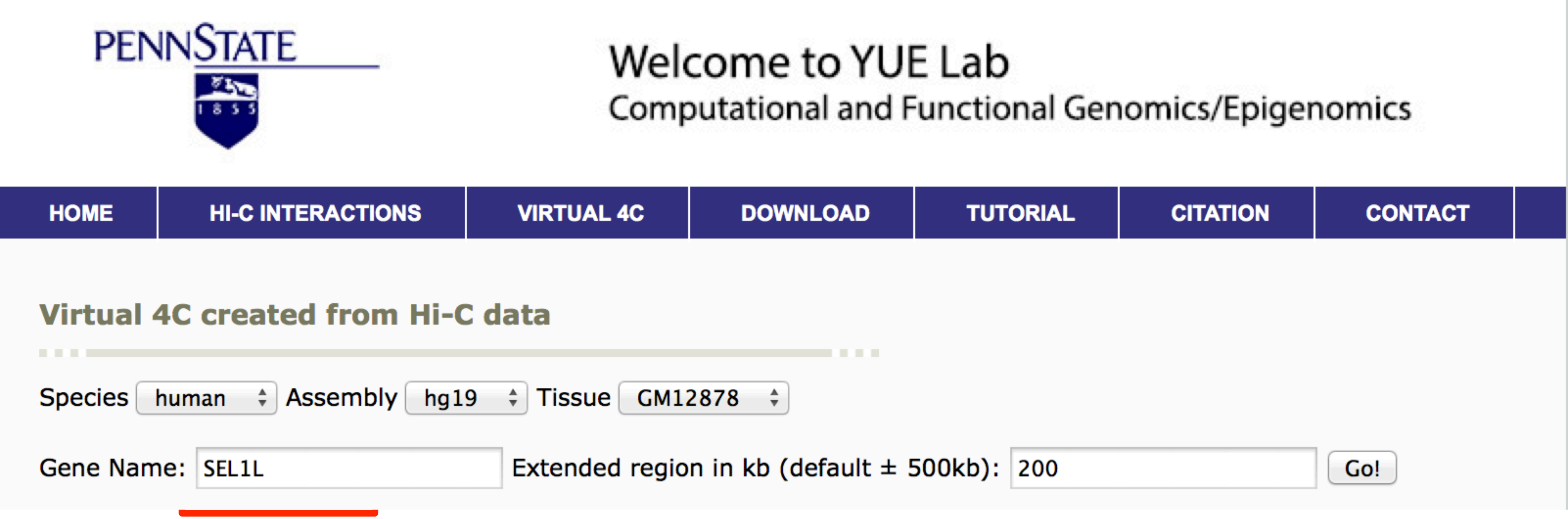
Option 1: Search by Gene Name	Option 2: Search by Location
Gene/Genomic Feature Show Interaction	Chromosome chr1 Start End Show Interaction

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



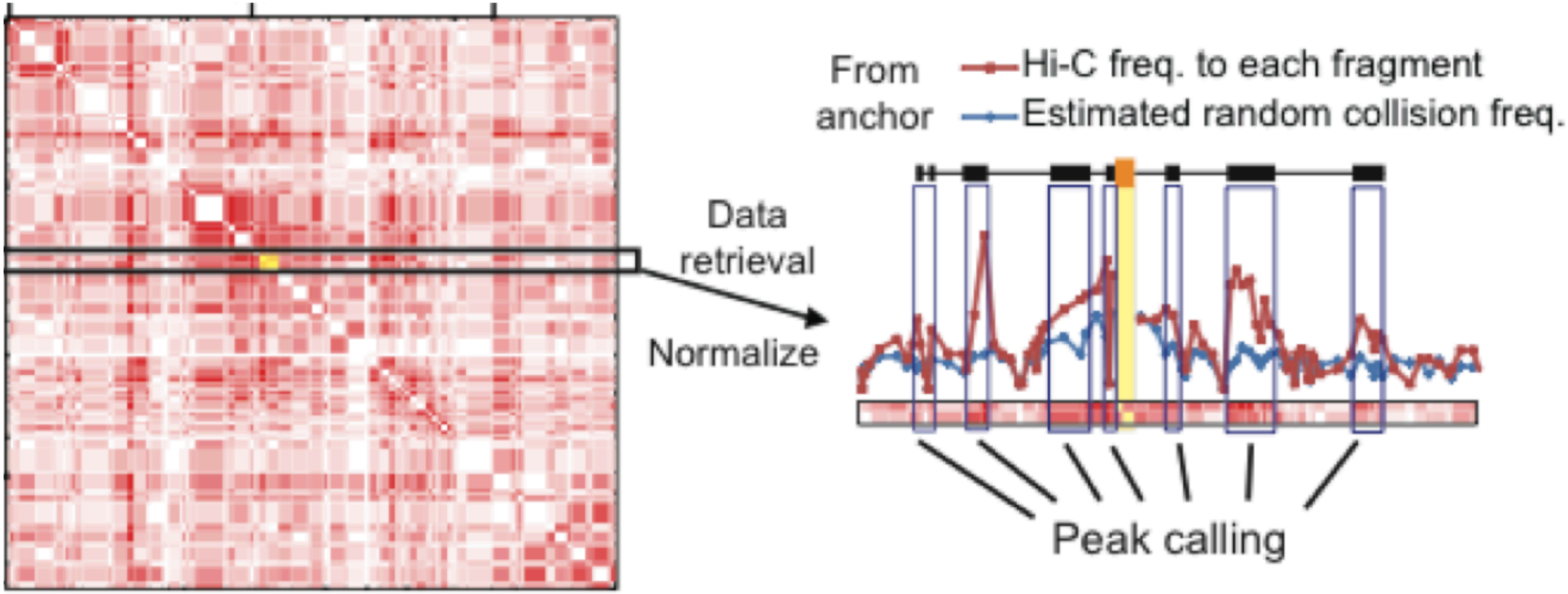
Welcome to YUE Lab
Computational and Functional Genomics/Epigenomics

Virtual 4C created from Hi-C data

Species: human
Assembly: hg19
Tissue: GM12878

Gene Name: SEL1L
Extended region in kb (default ± 500kb): 200

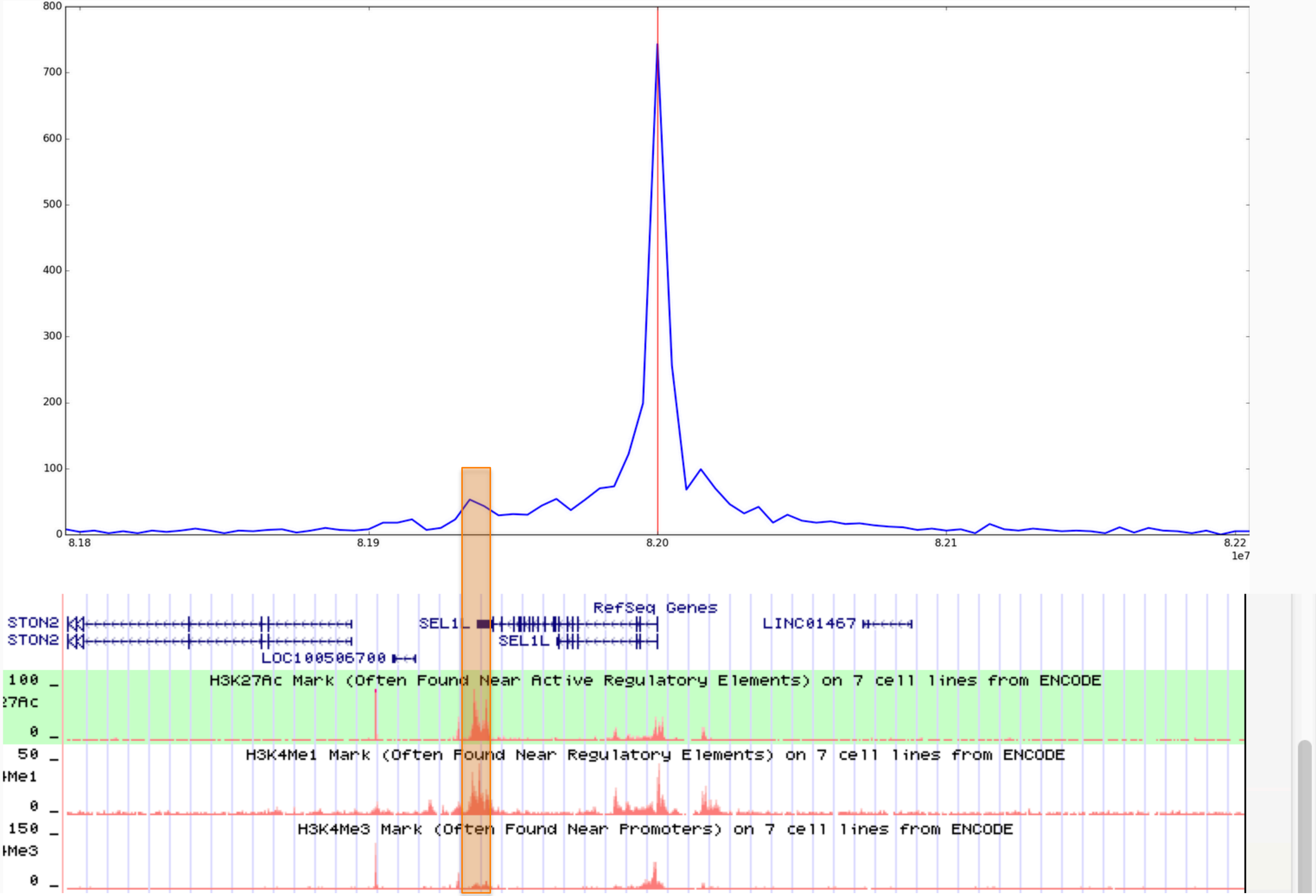
Virtual 4C is derived from Hi-C matrix



Jin et al, Nature, 2013

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.



Summary

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)

Acknowledgement

ENCODE consortium

ENCODE DCC

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

…

ENCODE EDAC

- Zhiping Weng

…

Dr. Bing Ren

Yue Lab

- Yanli Wang
- Lijun Zhang
- GalYaroslavsky
- Tyler Derr

and many others …

ENCODE Users Meeting

Stanford University, June 8 – June 10, 2016

- Goals: Teach users to navigate, analyze, use, and integrate ENCODE data

- Hands-on workshops in the afternoon
  - Access data thru the portal
  - Run uniform processing pipelines
  - Learn how to use web-based & command-line analysis tools

- Scientific talks in the morning
  - Interpret human variation and personal genomes
  - Interpret cancer genomes
  - Connect genes to their controlling regulatory elements
  - Identify likely cell types and pathways underlying non-coding disease associations

Updates at http://www.encode2016.org