



ENCODE element browser and 3D genome browser

Feng Yue, PhD
Assistant Professor
Dept. of Biochemistry & Molecular Biology
School of Medicine
Penn State University

Outline



- 1. Search and visualize ENCODE annotations (gene expression, candidate promoters, enhancers ...)**
- 2. Visualize ENCODE chromatin interaction data (Hi-C, ChIA-PET)**

ENCODE element query website

ENCODE

Data

Encyclopedia

Materials & Methods

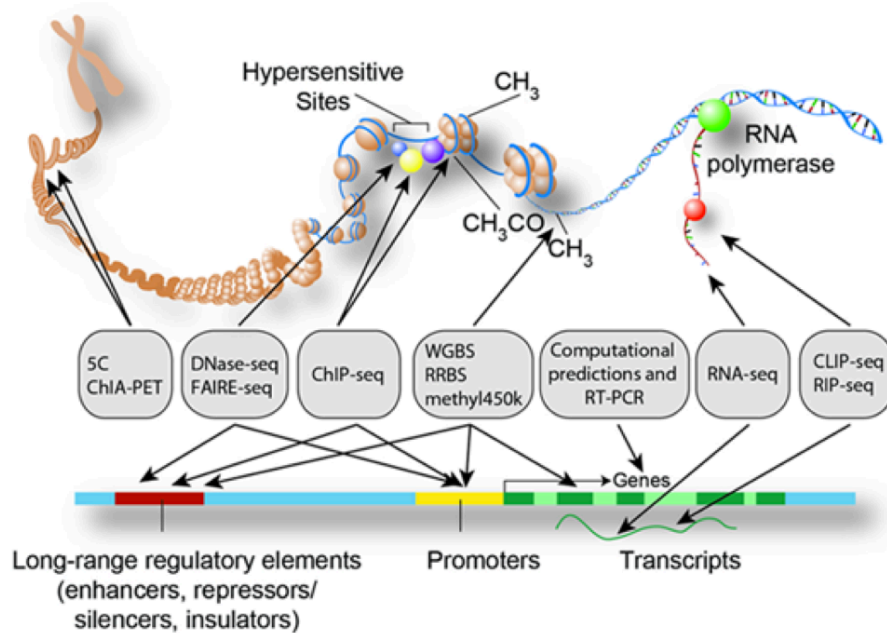
Help

About

Matrix

Search

ENCODE Encyclopedia of DNA Elements



The ENCODE (Encyclopedia of DNA collaboration of research groups functional elements in the human genome and RNA levels, and regulatory elements a gene is active.

Image credits: Darryl Leja (NHGRI), la

ENCODE element query website

ENCODE Data Encyclopedia Materials & Methods Help

Ground Level

- DNase-seq (peaks)
- Hi-C (links, TADs, compartments)
- ChIA-PET (links)
- gene expression
- transcription start sites
- TF ChIP-seq (peaks, motif sites)

Legend: available under development

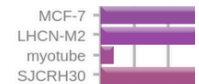
- Middle level annotations integrate multiple types of experimental data and multiple ground level annotations.
- Top level annotations integrate a broad range of experimental data and ground and middle level annotations.

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](#)]

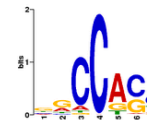


Transcription factor binding (TF ChIP-seq)

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](#)]



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 ...)

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 ...)

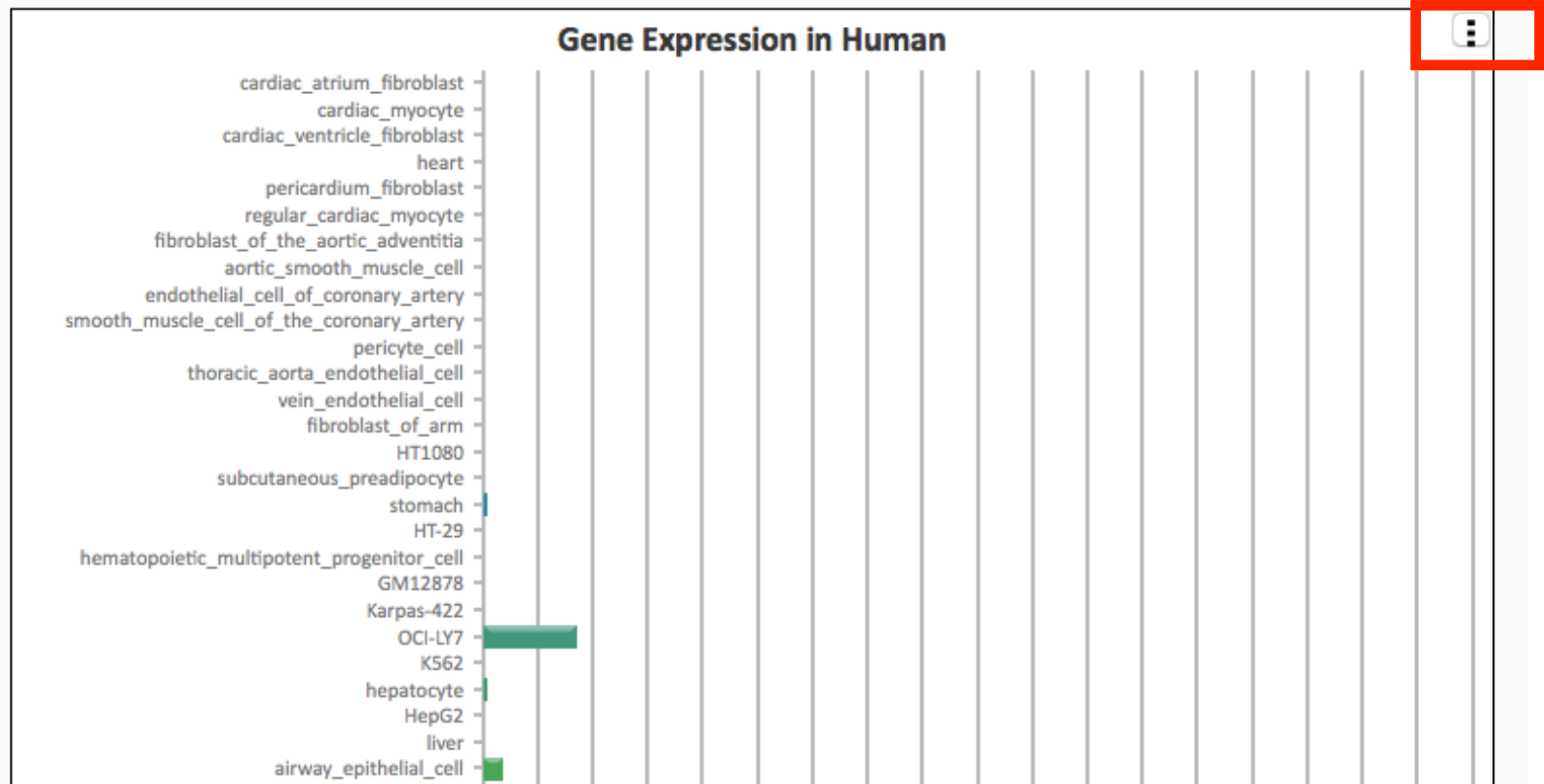
Visualize and save the gene expression profile

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

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



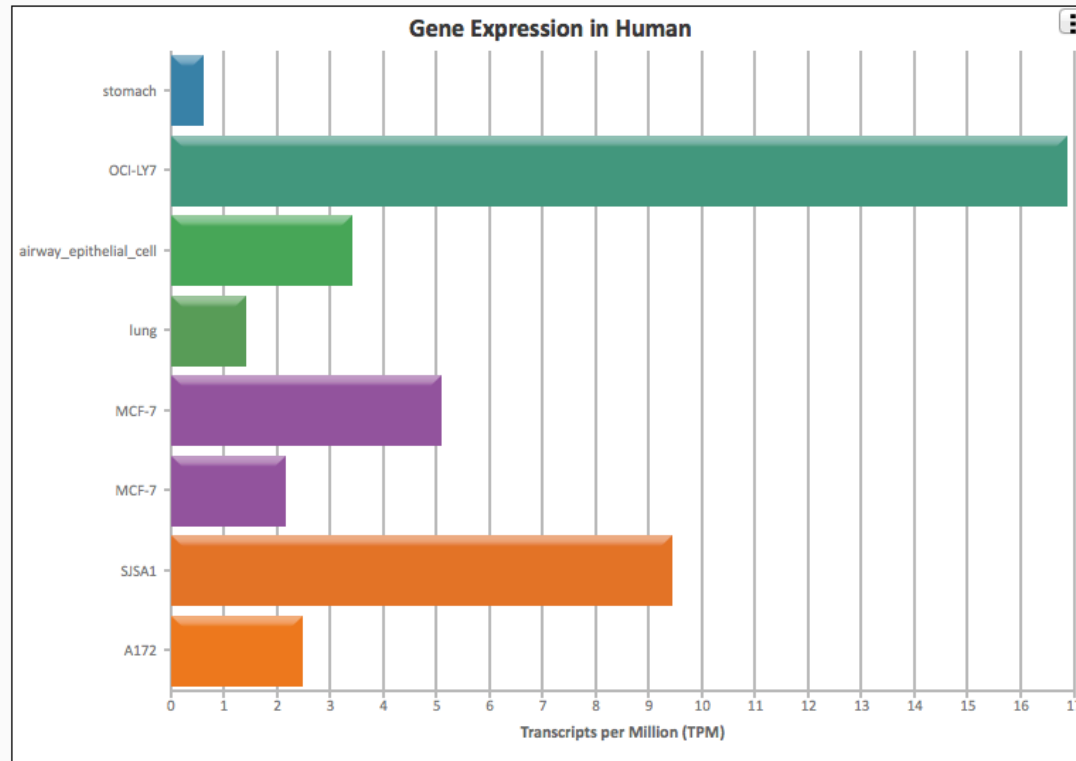
[ABOUT](#) [MOUSE](#) [HUMAN](#) [DOWNLOAD](#) [LINKS](#) [CONTACT](#)

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 ...) submit!

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

human (hg19) chr3 start: 181000000 end: 181200000 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

[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

ABOUT	MOUSE	HUMAN	DOWNLOAD	LINKS	CONTACT
-----------------------	-----------------------	-----------------------	--------------------------	-----------------------	-------------------------

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

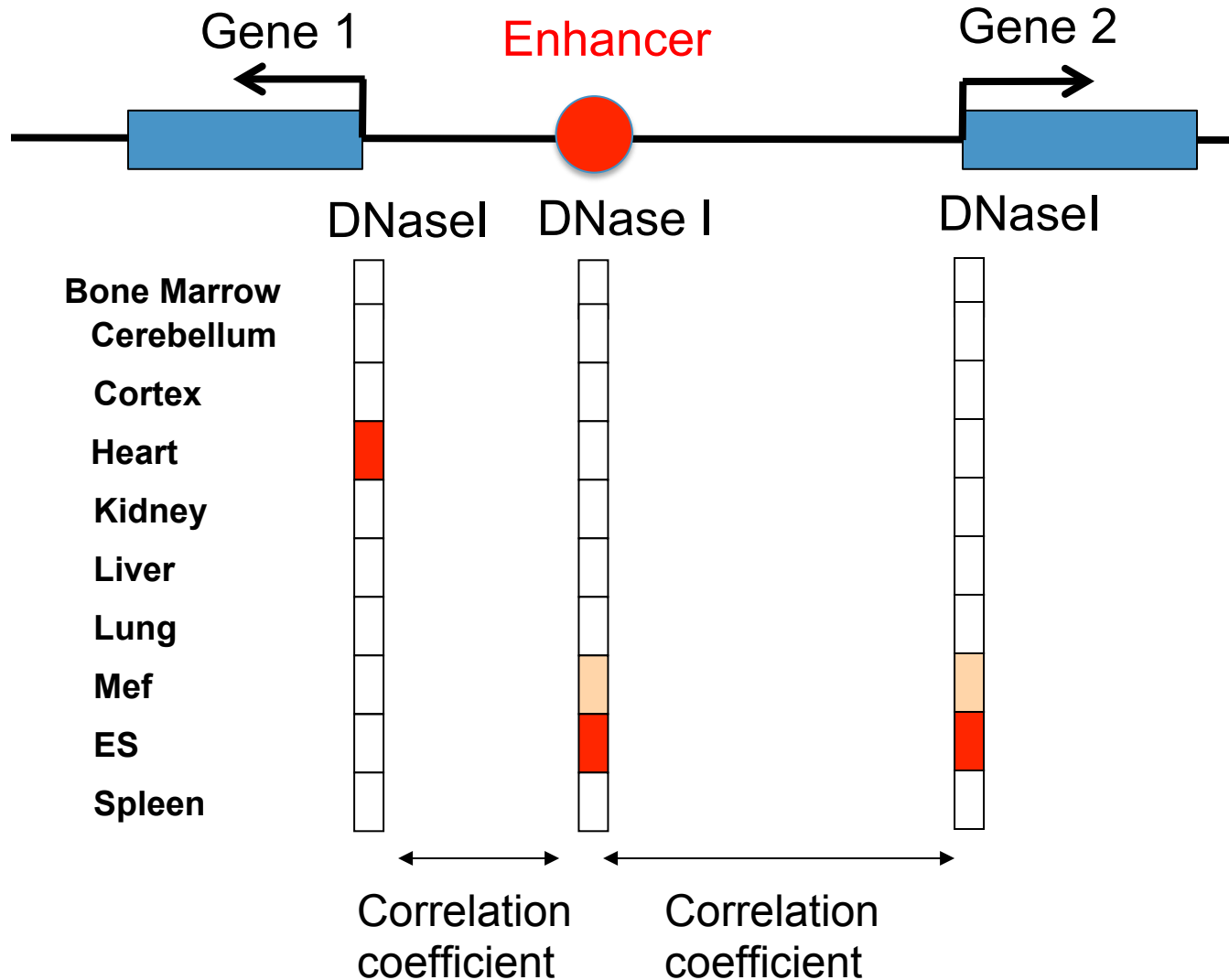
Human (hg19) Gene name(Sox2, Nanog ...)

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



Welcome to YUE Lab

Computational and Functional Genomics/Epigenomics

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

Cis-elements linked to your queried gene.

Species: mouse_mm9

Gene name: Sox2

Cis-element lined by DNaseI Hypersensitive Sites Linkage:

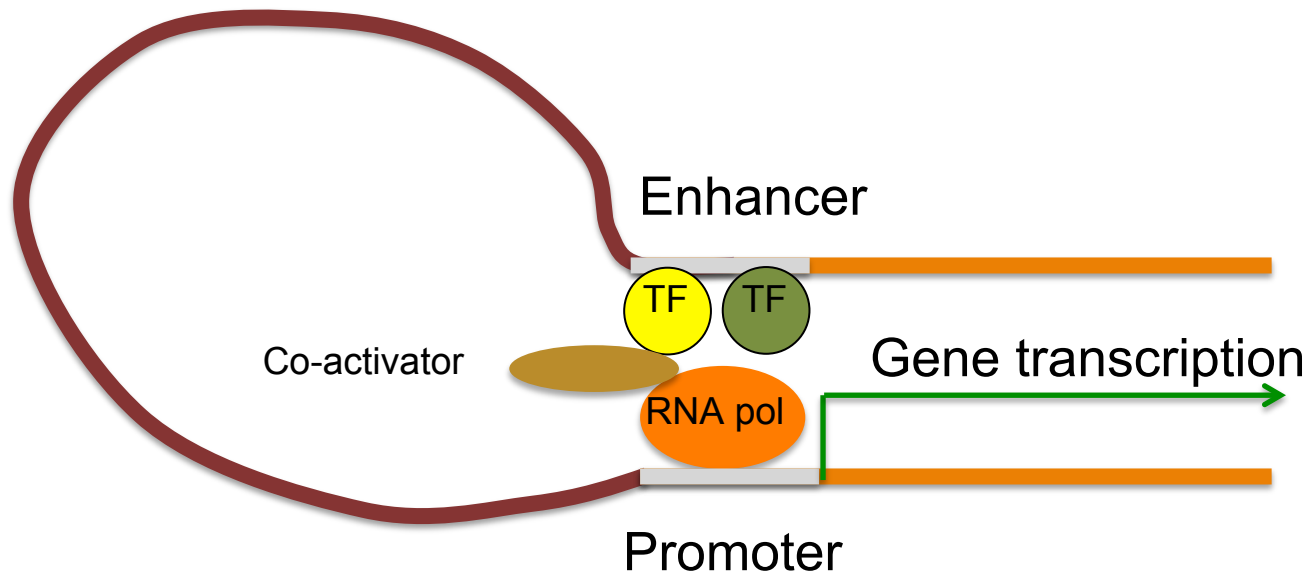
Proximal DHS(TSS)	start	end	Gene	Distal DHS	start	end	correlation
chr3	34548800	34548950	Sox2	chr3	34049080	34049230	0.746471
chr3	34548800	34548950	Sox2	chr3	34051020	34051170	0.751643
chr3	34548800	34548950	Sox2	chr3	34096080	34096230	0.851654
chr3	34548800	34548950	Sox2	chr3	34097280	34097430	0.700074
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

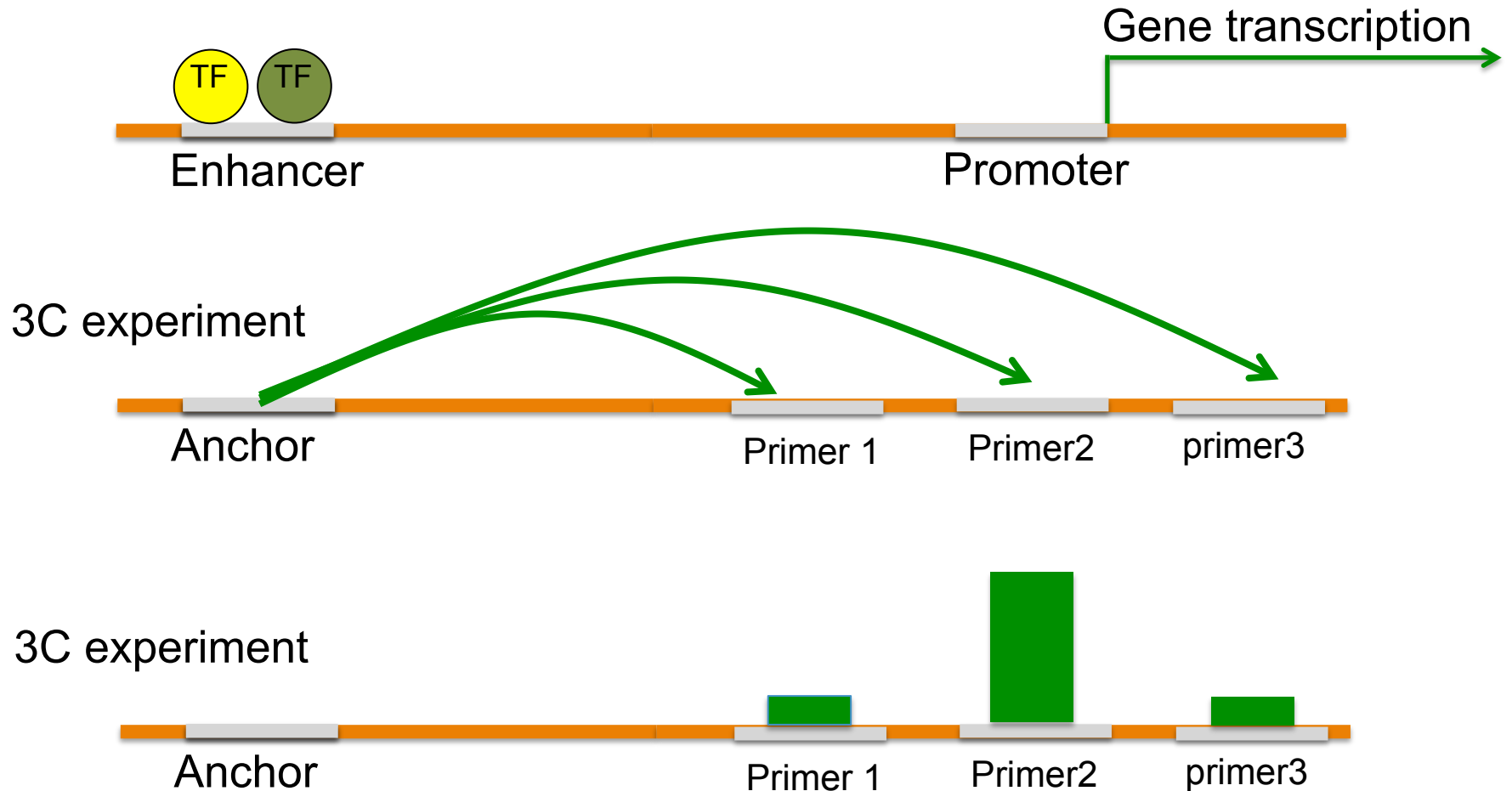


1. Search and download through ENCODE portal
2. Search and visualize ENCODE annotations (gene expression, promoters, enhancers ...)
3. Visualize ENCODE chromatin interaction data (Hi-C, ChIA-PET)

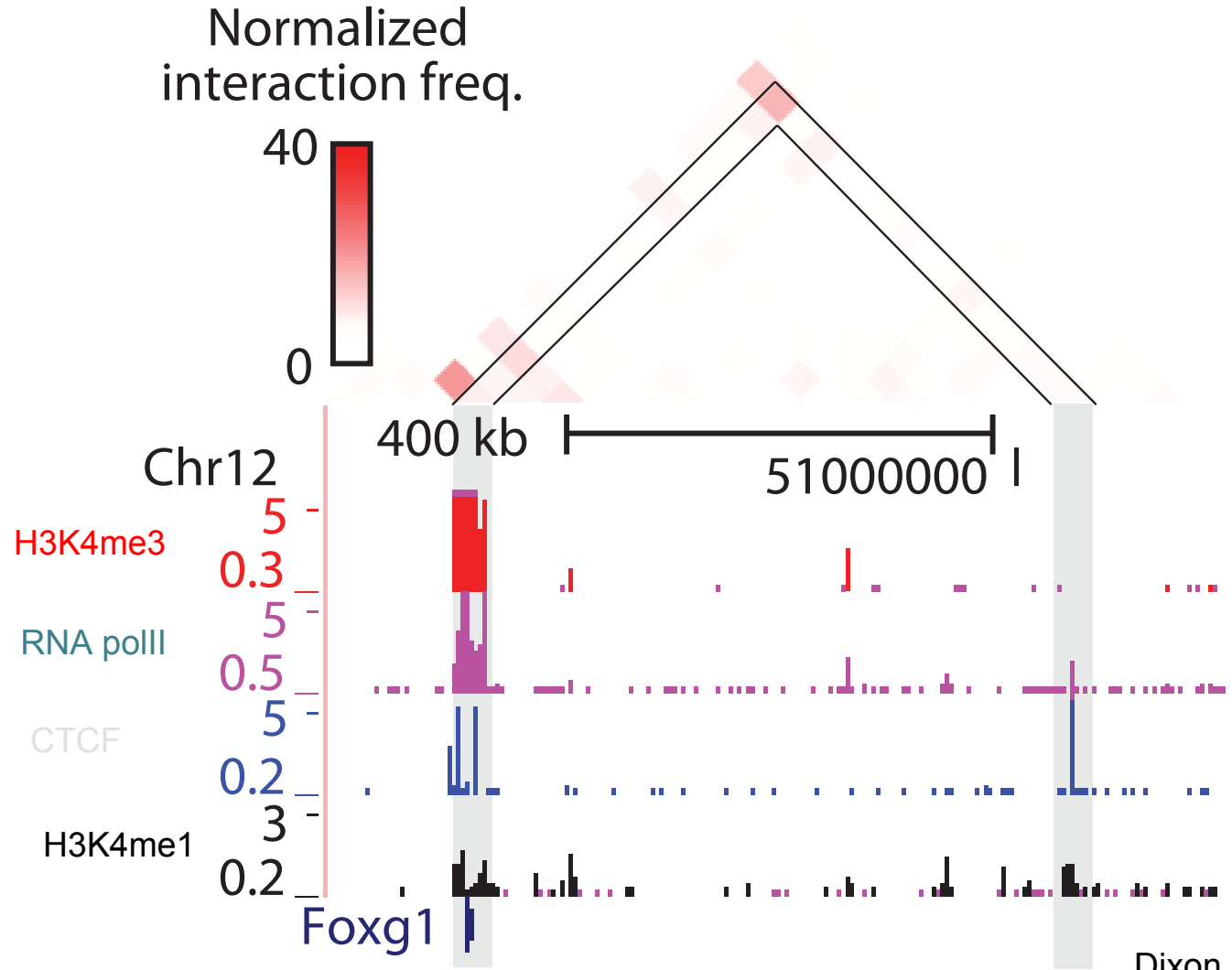
Transcriptional regulation by cis-regulatory elements



Chromatin conformation capture (3C) experiment



Hi-C measures genome-wide chromatin interaction



Hi-C measures genome-wide chromatin interaction

ENCODE

Data

Encyclopedia

Materials & Methods

Help

visualize sequence motifs and other information [[Factorbook](#)]

Histone mark enrichment (ChIP-seq)

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

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

Open chromatin (DNase-seq)

DNase I hypersensitive sites (also known as DNase-seq peaks) computed from ~300 human and mouse experiments.

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

Topologically associating domains (TADs) and compartments (Hi-C)

TADs and A and B compartments computed from 12 human cell lines.

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

Promoter-enhancer links (ChIA-PET)

Links between promoters and distal regulatory elements such as enhancers computed from 8 ChIA-PET experiments.

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

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

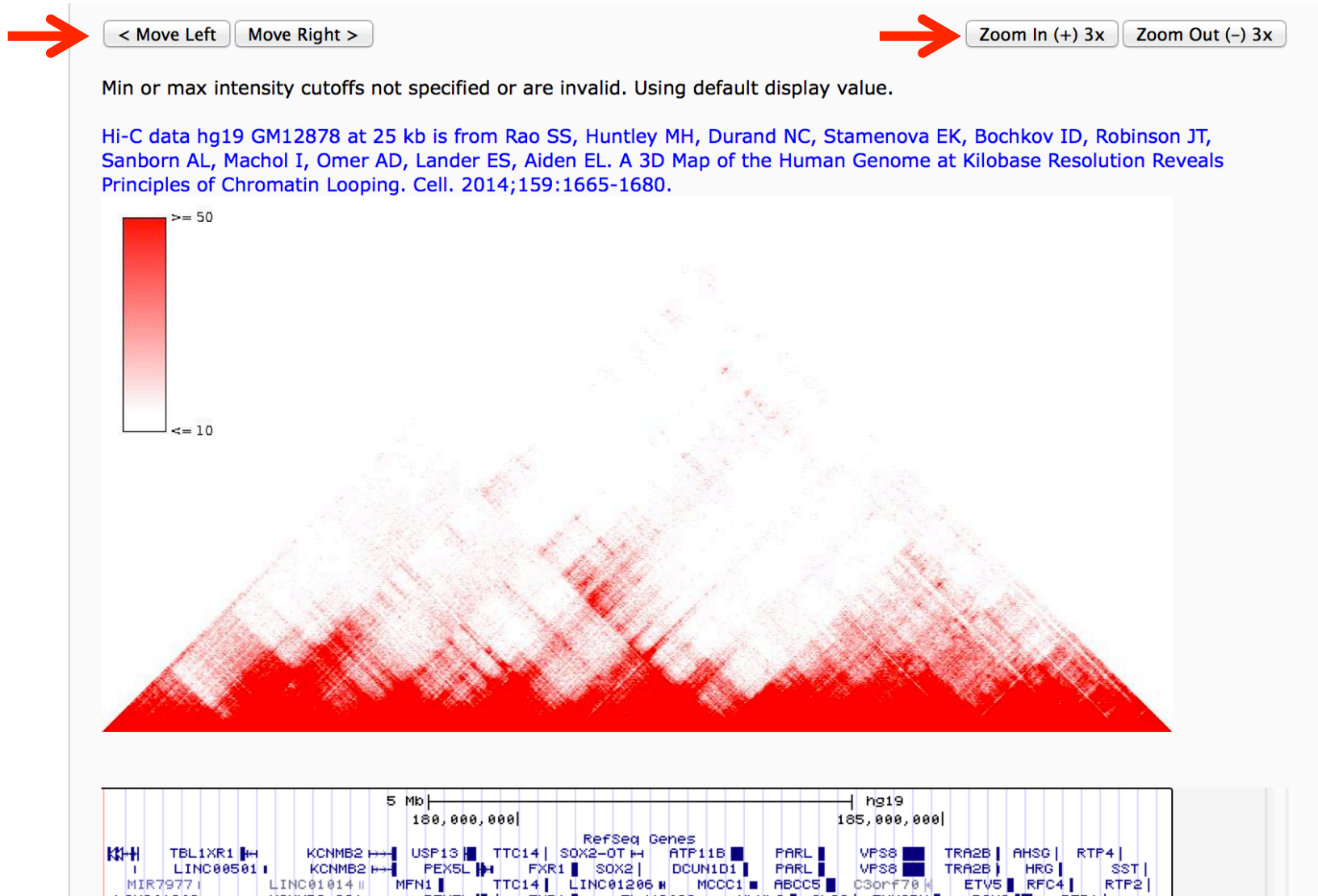
Welcome to Hi-C Data Browser!

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">SOX2SOX2-OTSOX20SOX21SOX22	<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

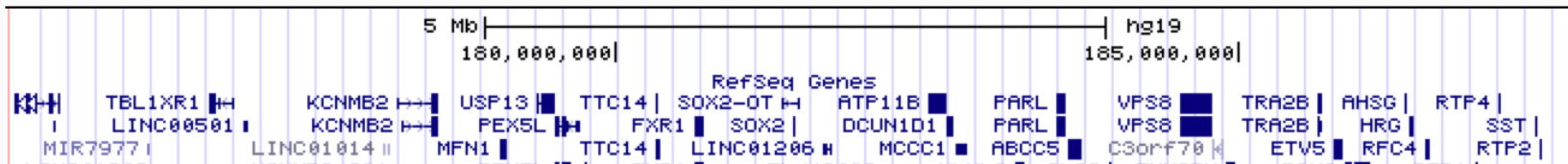
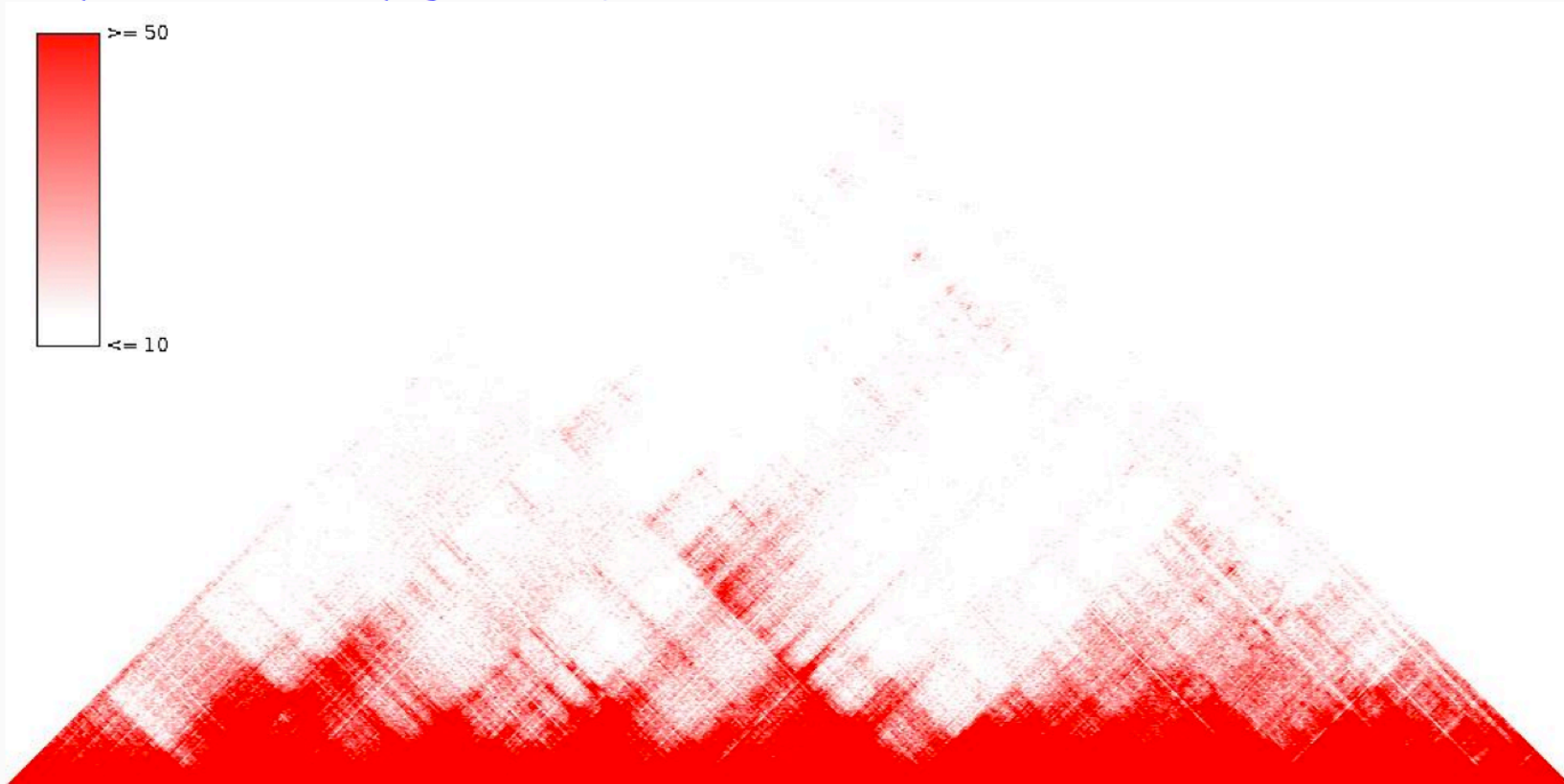
Move Right >

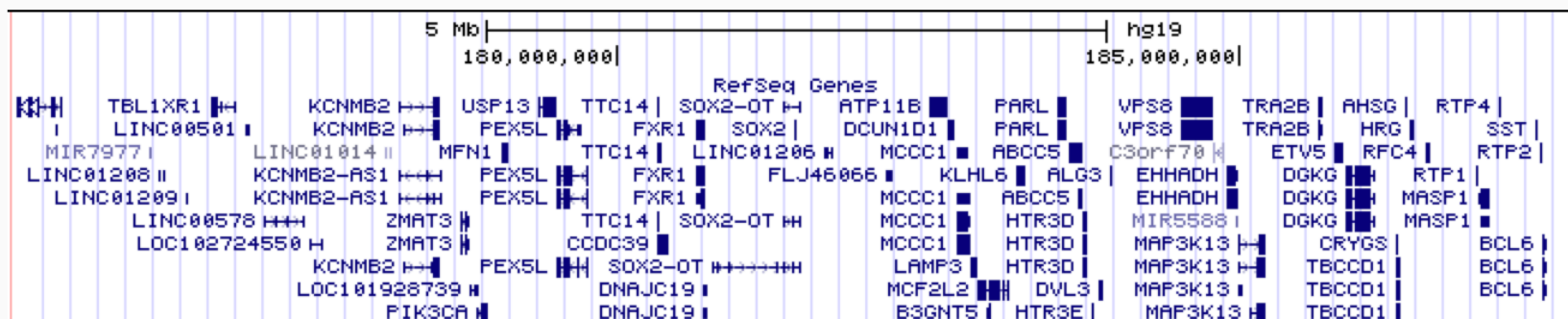
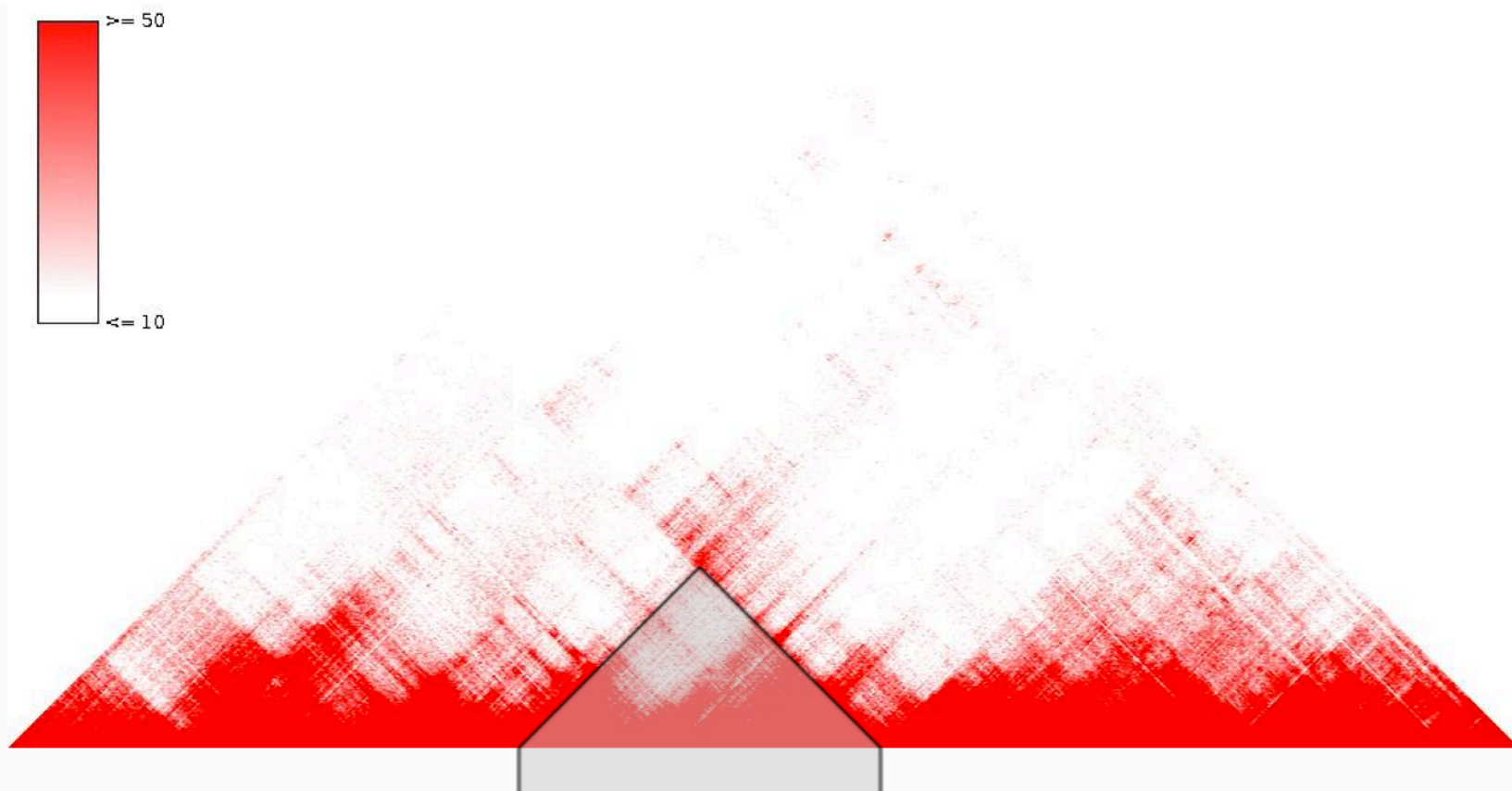
Zoom In (+) 3x

Zoom Out (-) 3x

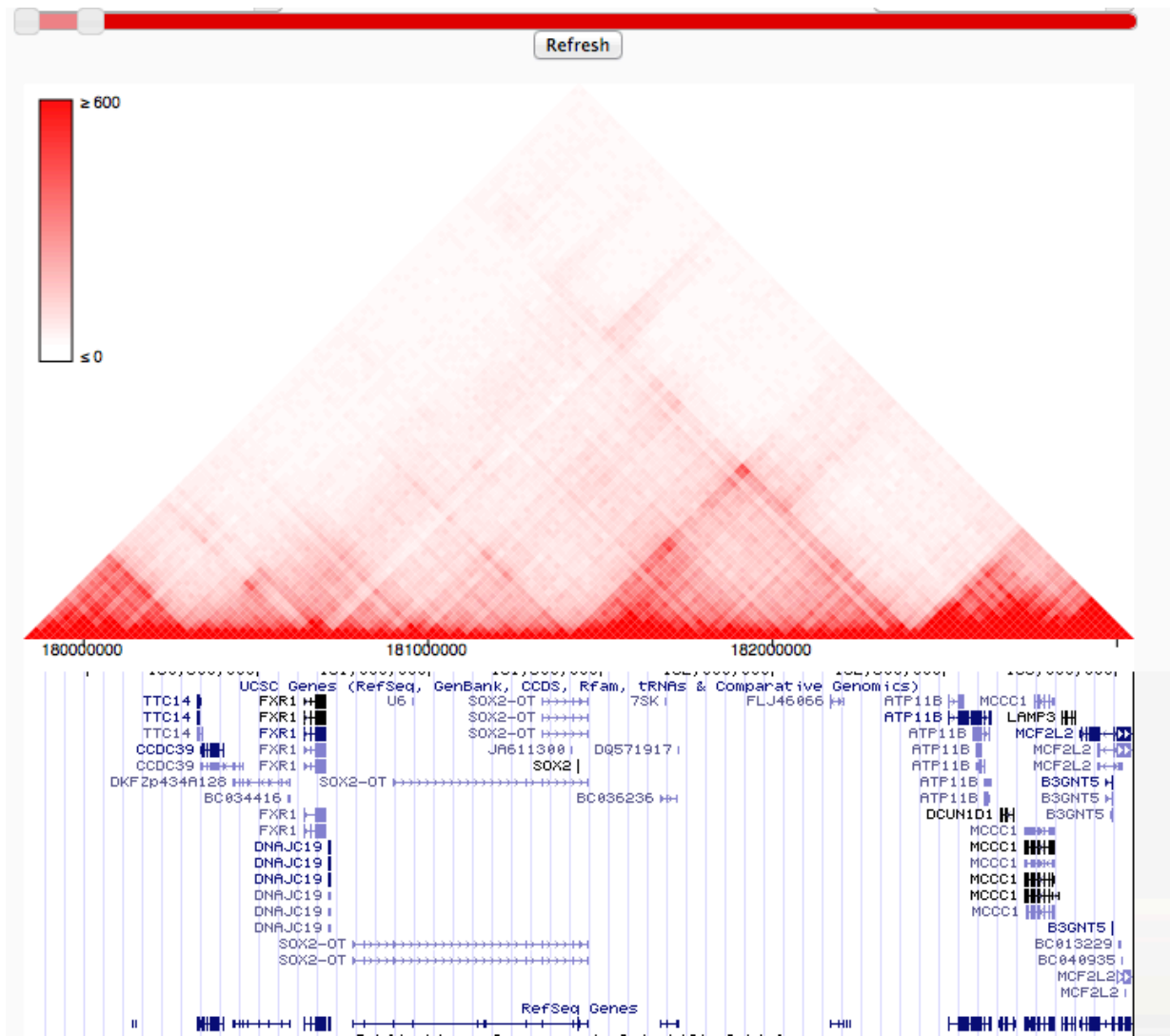
Min or max intensity cutoffs not specified or are invalid. Using default display value.

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

Upload Successful

assembly: hg19

resolution: 25000-bp

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

Visualize your own Hi-C data

Welcome to the Interactive Hi-C Data Browser!

Accessing Hi-C Data

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>)

< Move Left Move Right >

Zoom In (+) 3x

Zoom Out (-) 3x

Hi-C data accessed from <http://promoter.bx.psu.edu/public/GM12878.25kb.v1.2.btr>
assembly: hg19
resolution: 25000-bp

Location:

chr3:179825000-183050000

Name: SOX2

RefSeq: NM_003106

Ensembl (Gene): ENSG00000181449

Ensembl (Transcript): ENST00000325404

[Download Image](#)

0

600

Refresh

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

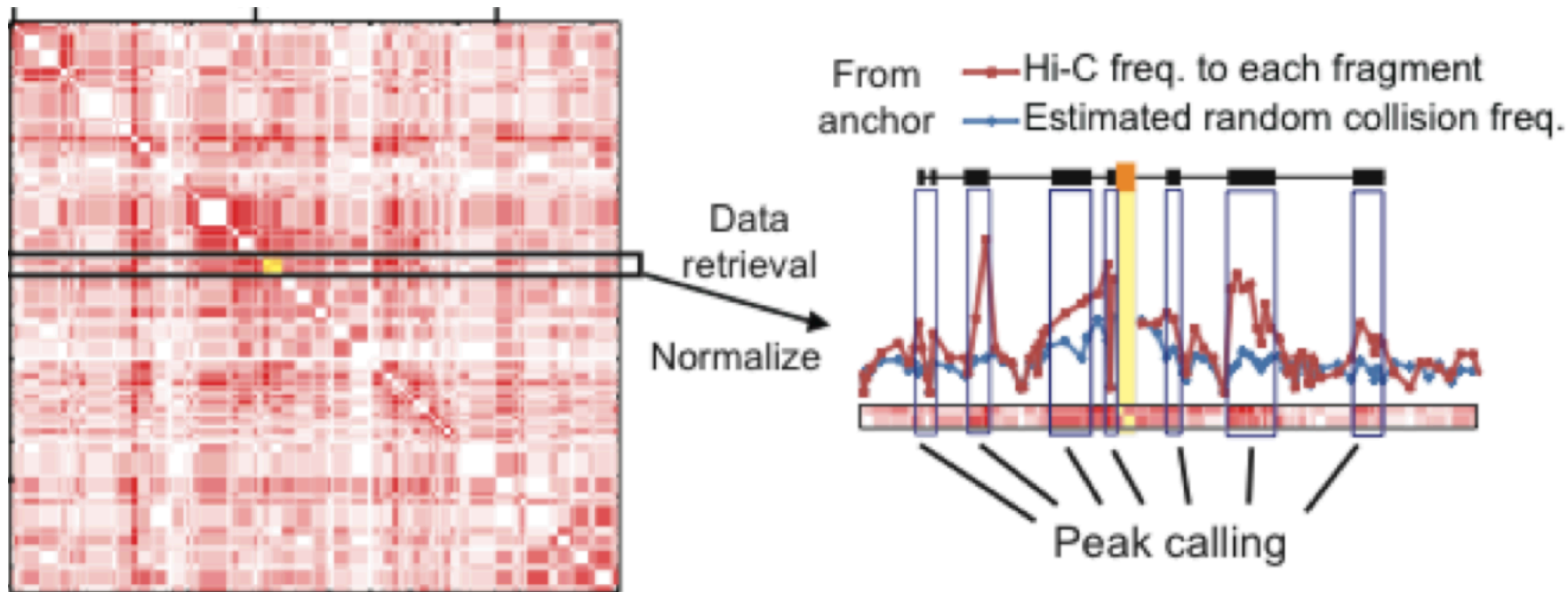
CONTACT

Virtual 4C created from Hi-C data

Species Assembly Tissue

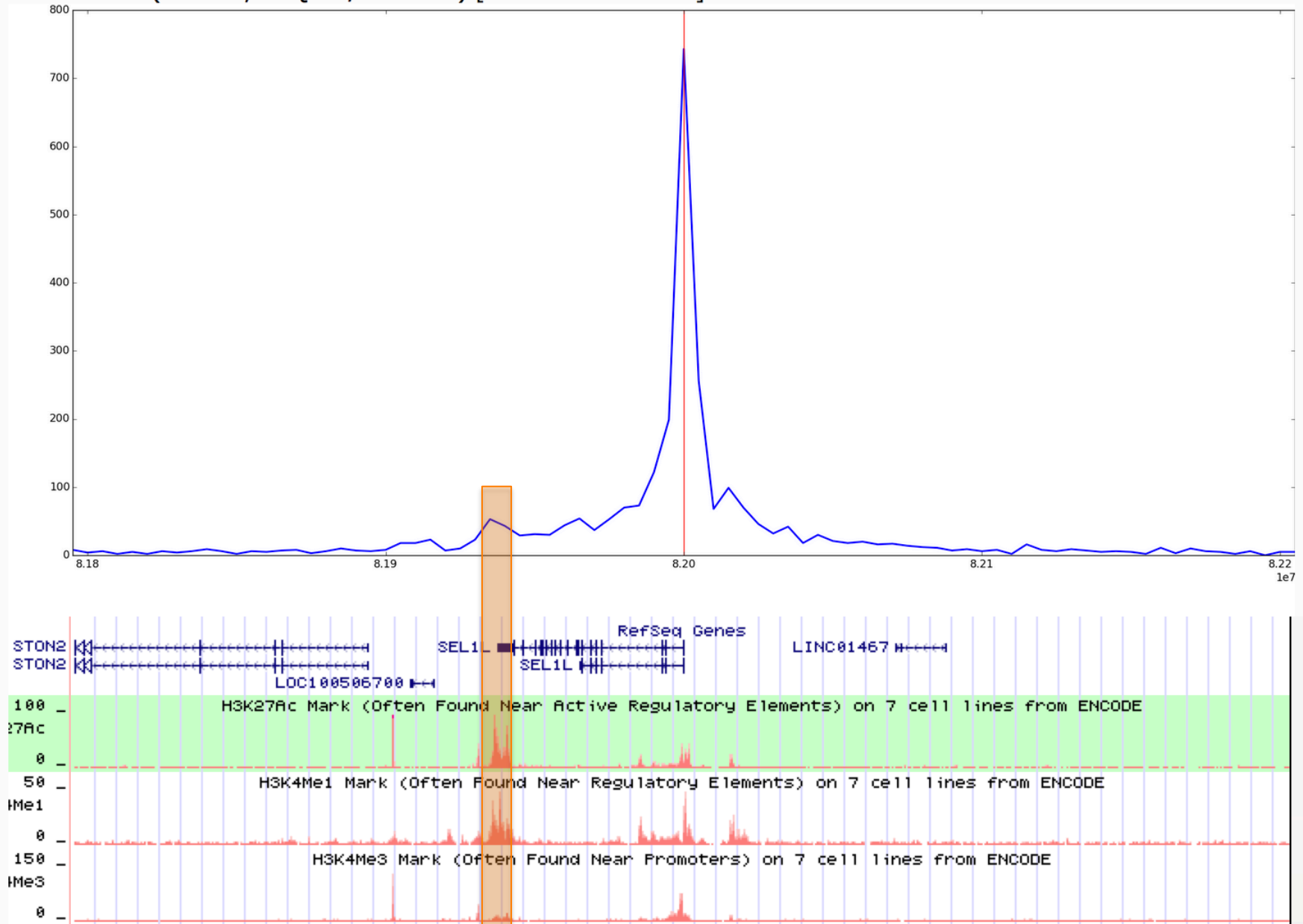
Gene Name: Extended region in kb (default \pm 500kb):

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.



Summary



- 1. Search and visualize ENCODE annotations (gene expression, promoters, enhancers ...)**
- 2. 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 ...