

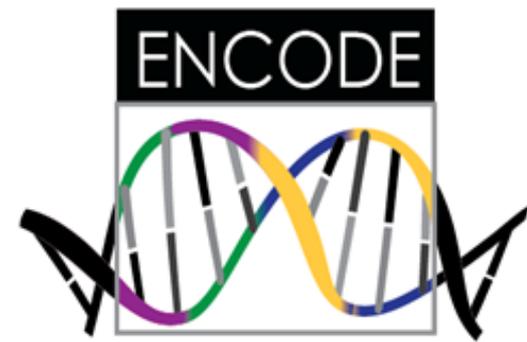
2015 Keystone ENCODE workshop

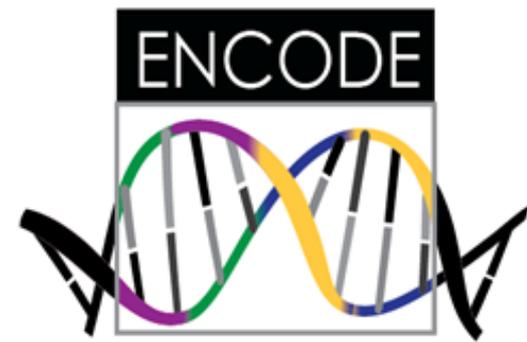
Feng Yue
Assistant Professor
School of Medicine
Penn State University

Workshop Schedule

Five speakers:

- Bing Ren (UCSD)
- Seth Strattan (Stanford Univ.)
- Feng Yue (Penn State)
- Jason Ernst (UCLA)
- Manolis Kellis (MIT)





Outline

1. How to query the ENCODE elements:

<http://promoter.bx.psu.edu/ENCODE>

- gene expression
- TF binding sites
- Histone modifications
- DNase I Hypersensitive Site

2. Linkage between enhancers to target genes:

- correlation (tissue-specificity).
- chromatin interaction

Hi-C browser (<http://3dgenome.org>)



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

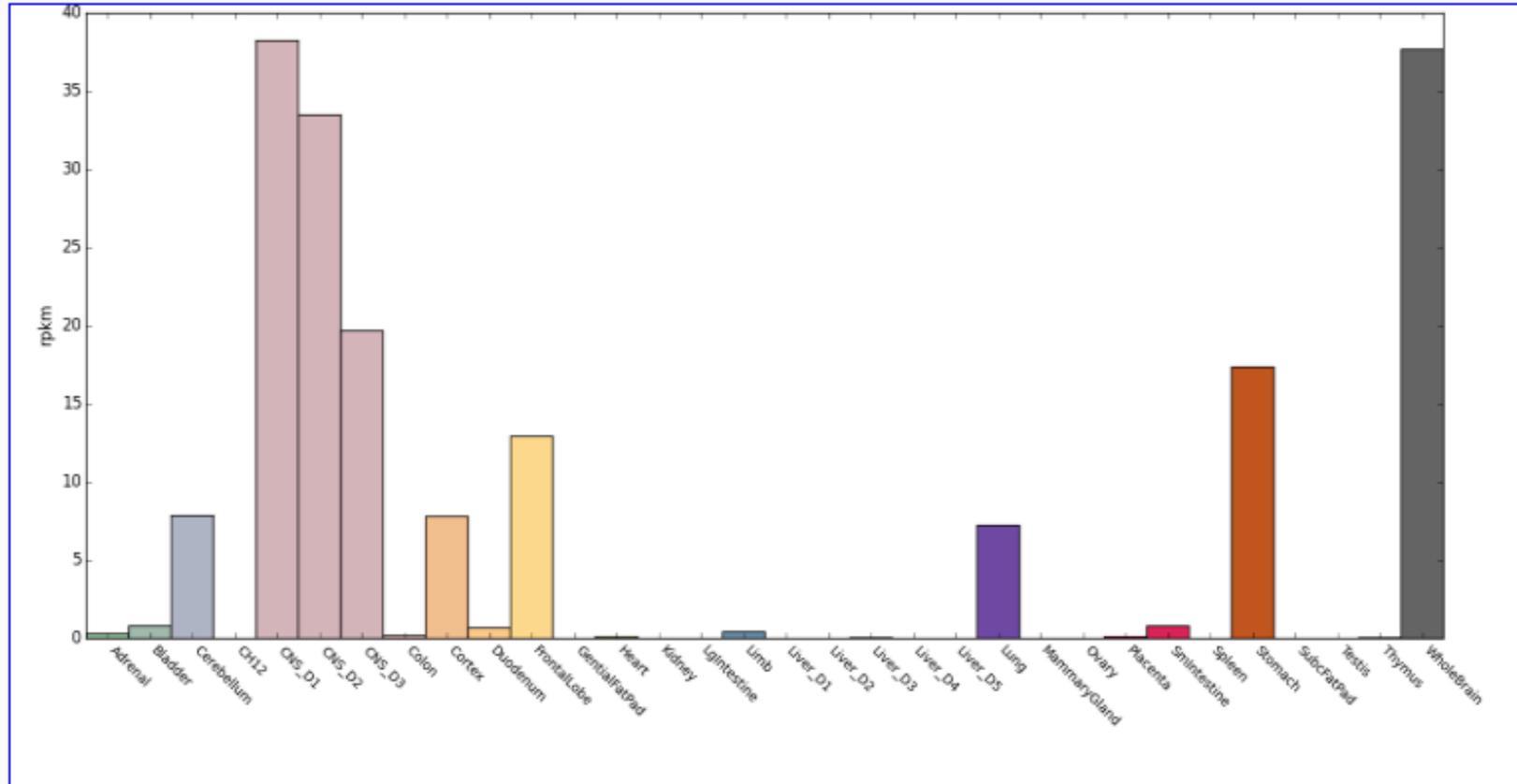
Option 4: search cis-elements LINKED to a gene based on DNaseI HSS specificity

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

Gene expression.

Mouse (mm9)

Gene **Sox2** (mCG_15247) [NM_011443, ENSMUSG00000074637, ENSMUST00000099151]





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

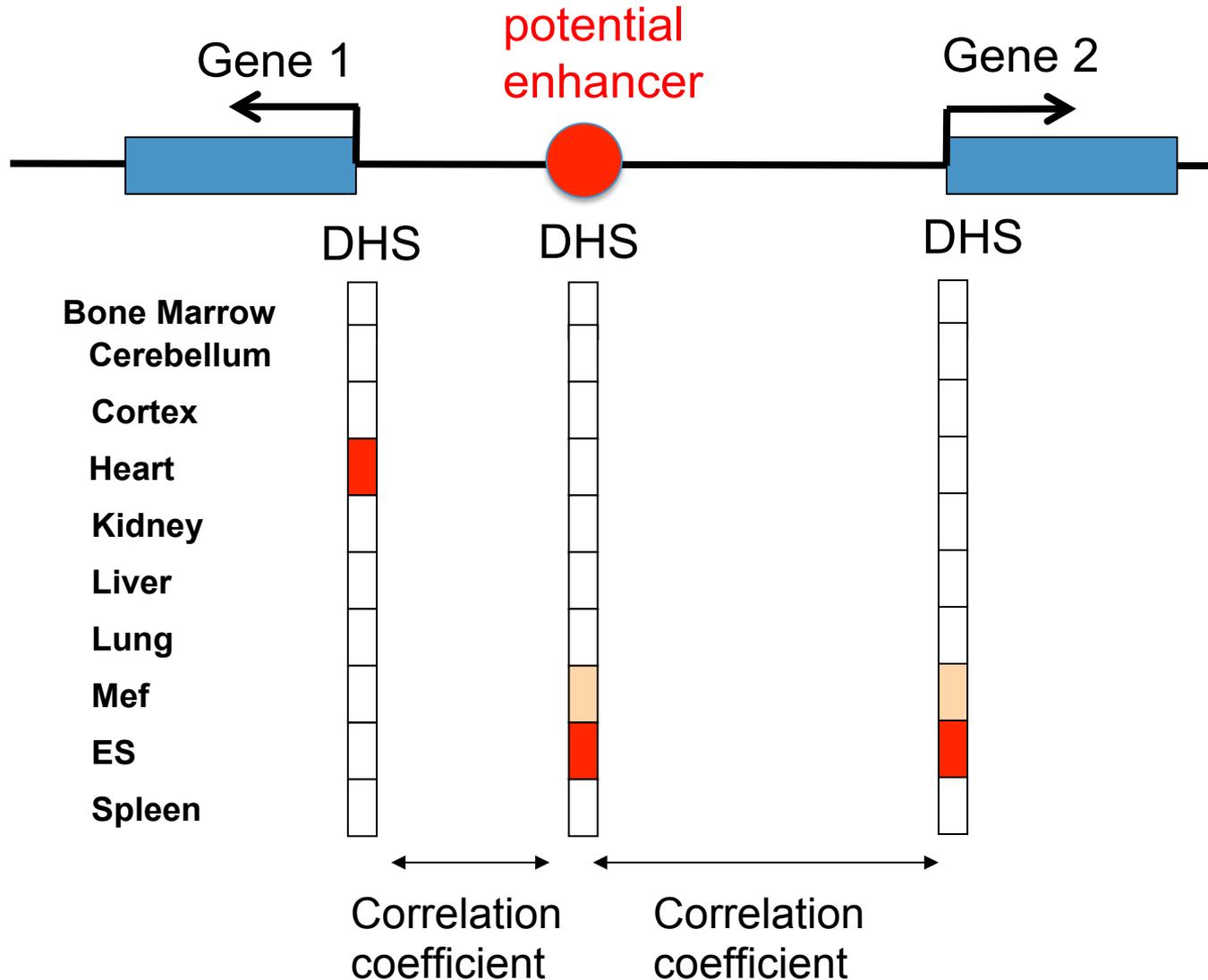
Human (hg19) Gene name(Sox2, Nanog ...)

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](#)

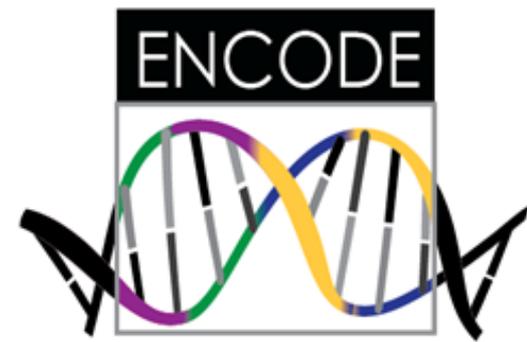
Cis-elements linked to your queried gene.

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

1. How to query the ENCODE elements:
<http://promoter.bx.psu.edu/ENCODE>
2. Linkage between enhancers to target genes by chromatin interactions:
 - Hi-C browser (<http://3dgenome.org>)

Welcome to the Interactive Hi-C Data Browser!

1. Choose the assembly of Hi-C data:

Species Assembly

2. Choose the source of the data:

Browse Available Hi-C Data

Tissue Resolution

OR

"C" Your Data

Data URL Resolution (in kb)

Please specify the URL of the data, e.g. <http://promoter.bx.psu.edu/hi-c/example/GM12878.25kb>

Note that the data must be in binary format. Ensure that the server supports HTTP and that the permission is set appropriately.

3. Choose the regions to show Hi-C interactions:

Option 1: Search by Gene Name	Option 2: Search by Location
<p>Gene name (Sox2, Nanog ...)</p> <input type="text"/> <p><input type="button" value="Show Interaction"/></p>	<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 align="center">***Optional***</p> <p>UCSC Genome Browser Session ID <input type="text"/></p> <p align="center">The browser URL or session hgsid, i.e. 423049851_sXOHIP1DqBDOniURwFqjfbBzqzja</p>	

First impression

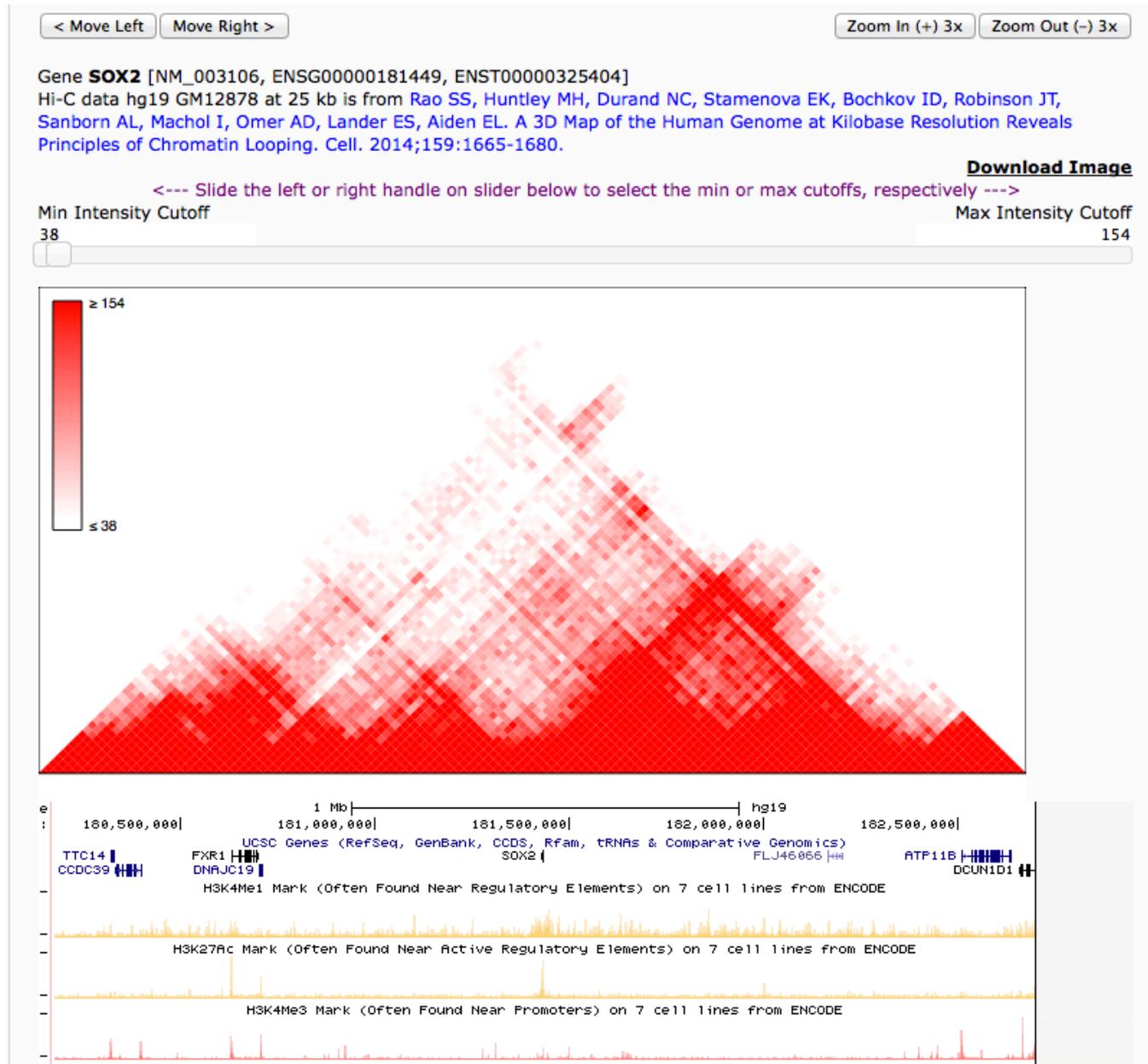
Navigation: ->

Data source: ->

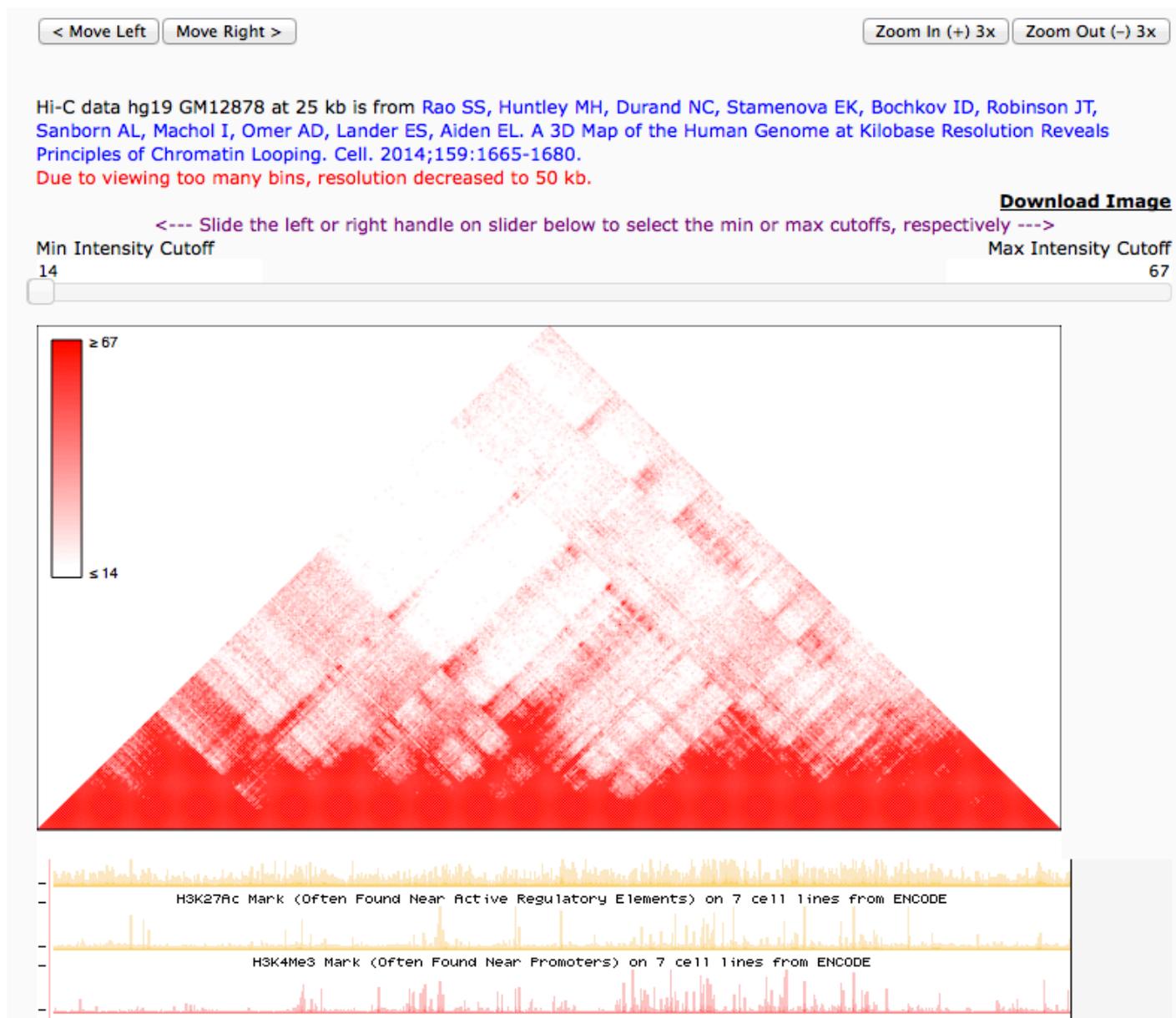
Dynamic
color bar: ->

Hi-C Heatmap ->

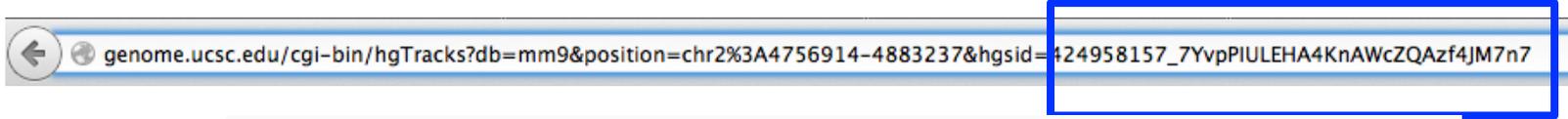
Imbedded UCSC
Genome browser->



Navigation (zooming out 3x)



Using your own UCSC Genome browser session



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

"C" Your Data
Data URL Resolution (in kb)
Please specify the URL of the data, e.g. <http://promoter.bx.psu.edu/hi-c/example/GM12878.25kb>
Note that the data must be in binary format. Ensure that the server supports HTTP and that the permissions are set appropriately.

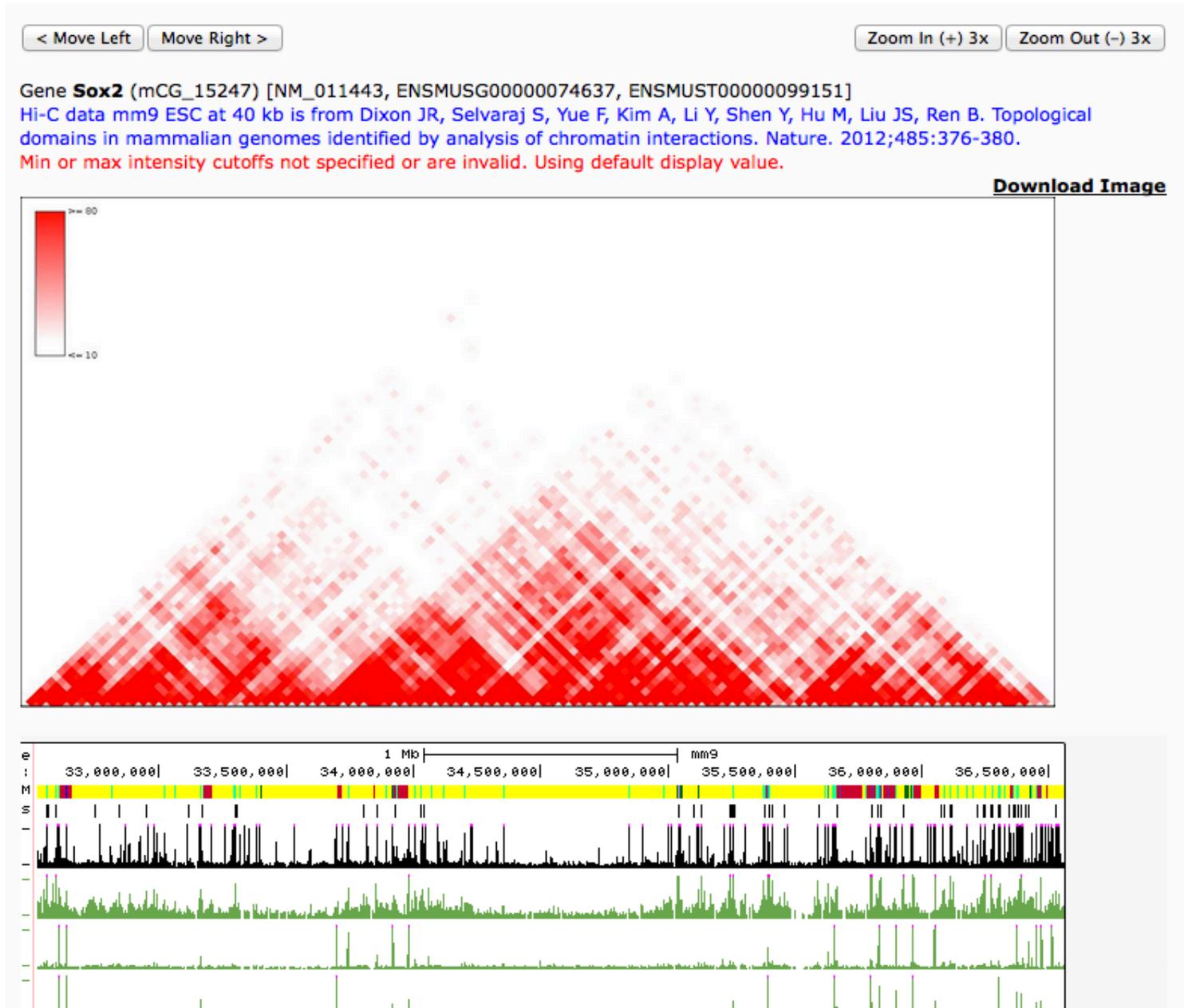
Step 3. Choose the region to show Hi-C interactions:

Option 1: Search by Gene Name	Option 2: Search by Location
Gene name (Sox2, Nanog ...) <input type="text"/> <input type="button" value="Show Interaction"/>	Chromosome <input type="text" value="chr1"/> Start <input type="text" value="172600000"/> End <input type="text" value="190275000"/> <input type="button" value="Show Interaction"/>

Optional
UCSC Genome Browser Session ID
The browser URL or session hgsid, i.e. 423049851_sXOHIP1DqBDOniURwFqjfbBzqzja



Using your own UCSC Genome browser session



Using your own Hi-C Data

1. Binary data – we will provide a script
http://promoter.bx.psu.edu/hi-c/database_new.php
2. Data in a internet accessible place.
3. Make sure the genome assembly and bin size is correct.
4. Note: it's slower during the internet traffic.

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

Welcome to the Interactive Hi-C Data Browser!

...

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

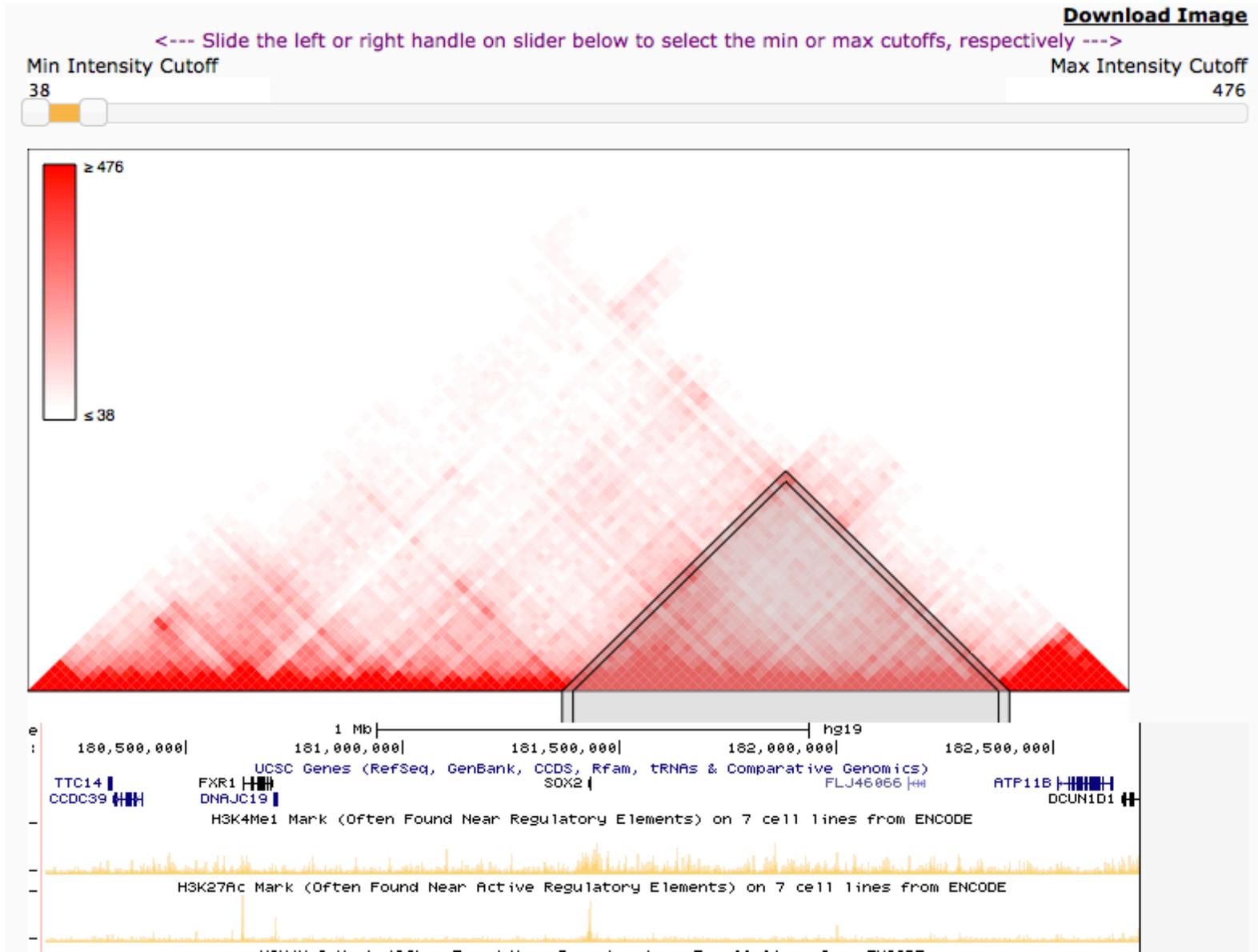
"C" Your Data
Data URL Resolution (in
kb)

Please specify the URL of the data, e.g. <http://promoter.bx.psu.edu/hi-c/example/GM12878.25kb>
Note that the data must be in binary format. Ensure that the server supports HTTP and that the permissions are set appropriately.

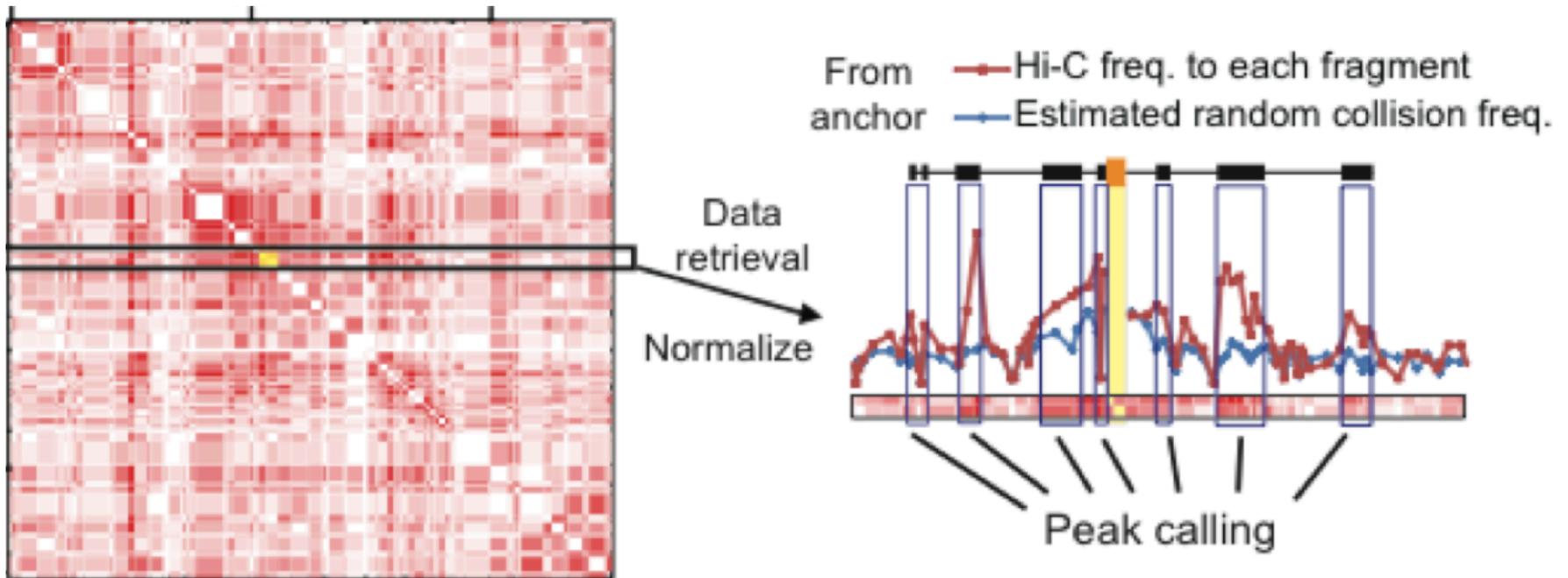
Step 3. Choose the region to show Hi-C interactions:



Click on image to display the interacting regions



Virtual 4C to reveal potential distal regulatory elements



Gene centric view of virtual 4C



Welcome to YUE Lab
Computational and Functional Genomics/Epigenomics

HOME

HI-C INTERACTIONS

VIRTUAL 4C

DOWNLOAD

TUTORIAL

CITATION

CONTACT

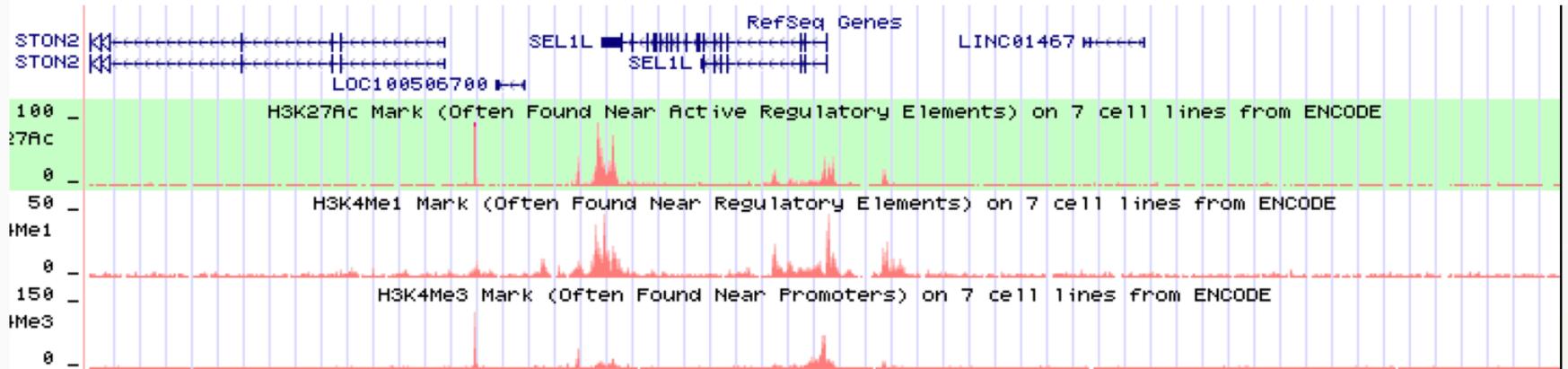
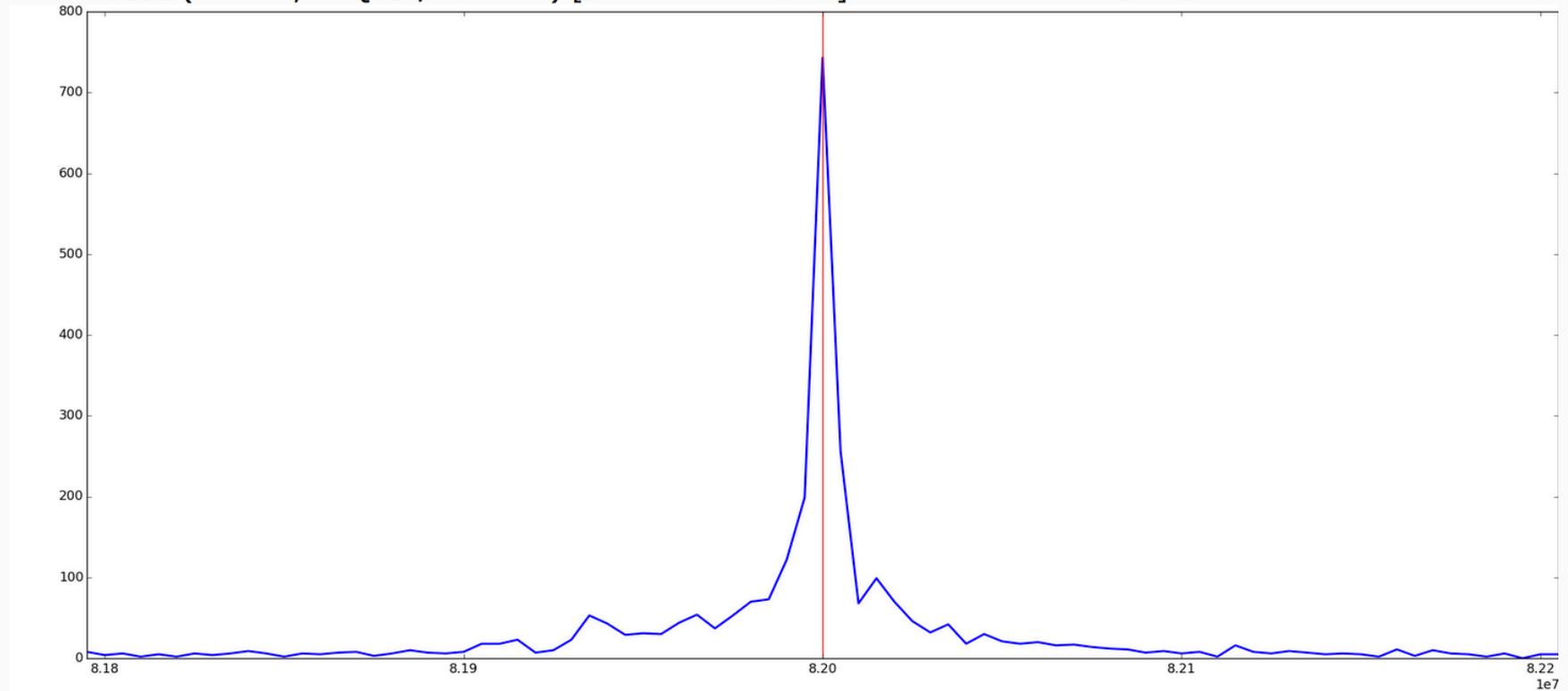
Virtual 4C created from Hi-C data

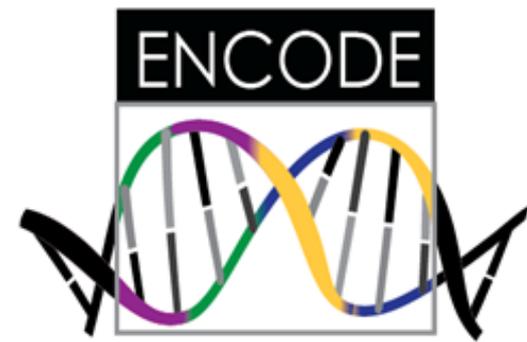
Species Assembly Tissue

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

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.

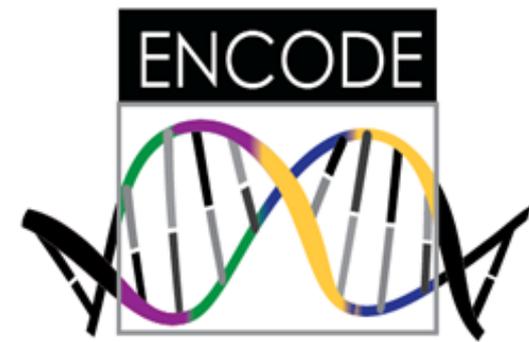




Summary

1. We have built two websites to:
 - Query the ENCODE elements.
 - Query the published/unpublished Hi-C data.
2. Comments or suggestions?

Acknowledgement



ENCODE consortium

ENCODE DCC

ENCODE Outreach Group

ENCODE Encyclopedia Group

Yue Lab:

Dr. Tingting Liu

Dr. Hongbo Yang

Yanli Wang

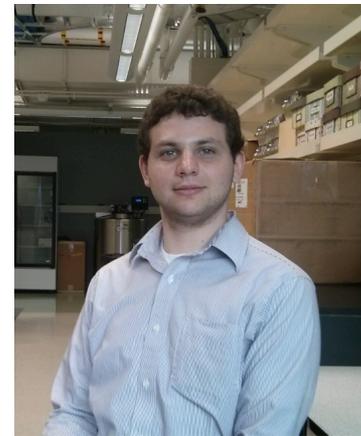
Tarik Salameh

Tyler Derr

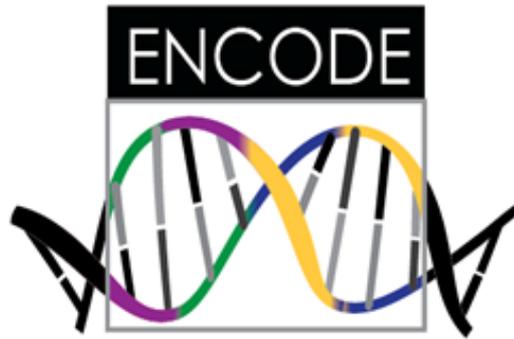
Gal Yaroslavsky



Yanli Wang



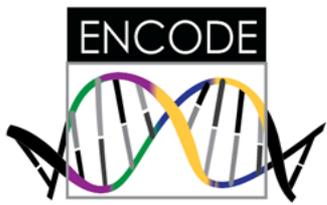
Gal Yaroslavsky



ENCODE User's Meeting

June 29 – July 1, 2015
Bolger Center, Potomac, MD

encode2015.org



ENCODE Users Meeting

Bolger Center, Potomac, MD, June 29 – July 1, 2015

- Goals: Teach users to navigate, analyze, use, and integrate ENCODE data
- Use ENCODE data:
 - 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
- Scientific talks in the morning
 - Invited speakers and selected from submitted abstract
- 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