

ENCODE Element Browser & the 3D Genome Browser

Yanli Wang

Feng Yue Laboratory

MD/PhD Student

Bioinformatics and Genomics Program

**inspired
together**



THE HUCK INSTITUTES
OF THE LIFE SCIENCES

PENNSTATE HERSEY
 College of Medicine

Overview

- ENCODE Element Browser:
 - Gene expression across multiple cell types
 - (Putative) cis-regulatory elements
- 3D Genome Browser:
 - Hi-C visualization
 - ChIA-PET visualization
- Goals of these resources:
 - Querying the most relevant ENCODE data
 - Visualizing complex data
 - Providing additional layer of evidence for the target gene of cis-regulatory elements
- **Requires Javascript; Recommends HTML5,**

Chrome 40+

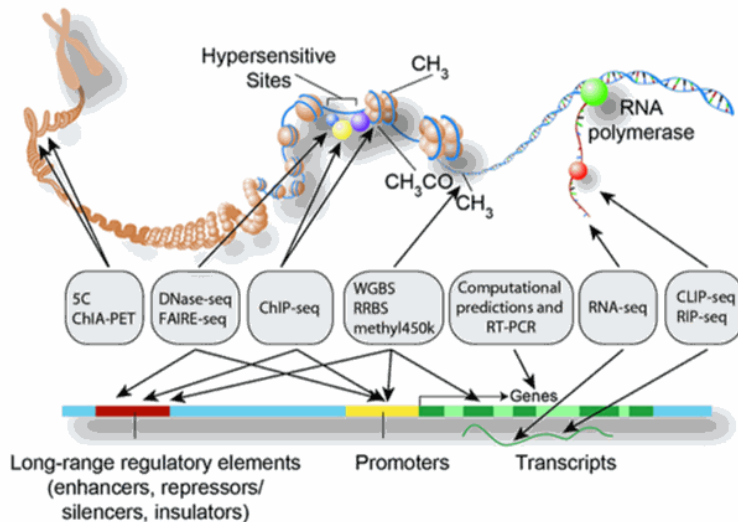
THE HUCK INSTITUTES
OF THE LIFE SCIENCES

Part I: The ENCODE Element Browser

**inspired
together**



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)

together

ENCODE Element Browser: Gene expression (Option 1)

- Click on **HUMAN**
- Under **Option 1: Search Gene Expression**, enter *IKZF1*, click **Submit**. (Accepted inputs include gene symbol, refseq, Uniprot)



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 +/- 20kb) kb submit!

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

human (hg19) ▼ Gene name(Sox2, Nanog ...) submit!

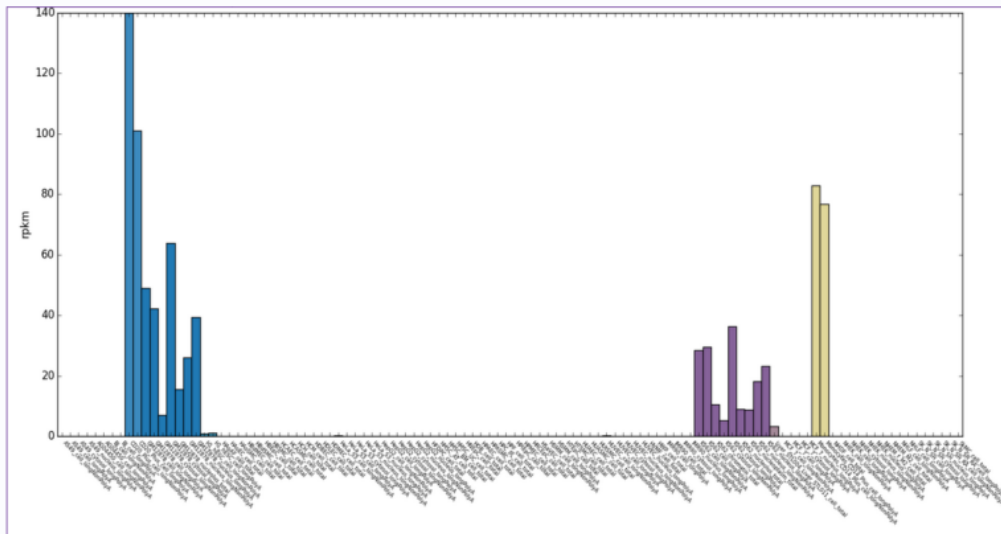
ENCODE Element Browser: Gene expression (Option 1) (human)

ABOUT MOUSE HUMAN DOWNLOAD LINKS CONTACT

Gene Expression Across Multiple Tissues/Cell Lines.

Human (hg19)

Gene **IKZF1** (IKAROS, LYF1, ZNFN1A1) [NM_001220771, NM_006060, NM_001220765, NM_001220776, NM_001220772, NM_001220766, NM_001220775, NM_001220767, NM_001220770, NM_001220769, NM_001220768, NM_001220774, NM_001220773, ENSG00000185811, ENST00000413698, ENST00000331340, ENST00000343574, ENST00000346667, ENST00000349824, ENST00000357364]



A549_cell_longNonPolyA	0
A549_cell_longPolyA	0
A549_cytosol_longPolyA	0
A549_nucleus_longPolyA	0
AG04450_cell_longNonPolyA	0
AG04450_cell_longPolyA	0
BJ_cell_longNonPolyA	0.018953
BJ_cell_longPolyA	0
CD20_Plus_cell_longNonPolyA	139.906
CD20_Plus_cell_longPolyA	101.079
GM12878_cell_longNonPolyA	49.0226
GM12878_cell_longPolyA	42.328
GM12878_cytosol_longNonPolyA	6.90498
GM12878_cytosol_longPolyA	63.8007
GM12878_nucleolus_total	15.5367
GM12878_nucleus_longNonPolyA	26.14

Gene IDs

e Expression (RPKM)
Bar Graph

Gene Expression
(RKPM) List

Candidate *cis*-regulatory element finder (Option 2 & Option 3)

- Queries for candidate *cis*-regulatory element regions with **DNase Hypersensitive Sites (DHSs)** and **Transcription Factor Binding Sites (TFBS)**, from Stamatoyannopoulos and Crawford Labs, as compiled by Weng Lab.
- Fast and easy way to determine *cis*-regulatory elements and their tissue specificity.

**inspired
together**



THE HUCK INSTITUTES
OF THE LIFE SCIENCES

PENNSTATE HERSEY
 College of Medicine

ENCODE Element Browser: Candidate *cis*-regulatory element finder (Option 2)

- Under **Option 2: search cis-elements in a given genomic region**
 - Select *chr7*
 - Enter *50,300,000* for start and *50,305,000* for end

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) ▼ chr7  start: 50300000  end: 50305000  submit!

Option 3: search cis-elements surrounding a gene

human (hg19) ▼ Gene name(Sox2, Nanog ...)

Extended region (default +/- 20kb) kb submit!

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

human (hg19) ▼ Gene name(Sox2, Nanog ...) submit!

ENCODE Element Browser: Candidate *cis*-regulatory element finder (Option 2)

DNaseI Hypersensitive Sites:

Coordinate	Tissue/cell type
chr7:50300040-50300190	Th1 Wb54553204, Treg Wb83319432
chr7:50300280-50300430	Th1 Wb33676984, Th1 Wb54553204, Th17
chr7:50300860-50301010	CD3 Primary Cells, CD8 Primary Cells, K562, SK-N-SH RA, Th1
chr7:50301060-50301210	CD8 Primary Cells, K562, Mobilized CD56 Primary Cells, Th1, Th1 Wb33676984, Th2 Wb33676984

TF binding Site:

Coordinate	TF	tissue
chr7:50301060-50301210	GATA2, GATA3, MAFF	GATA2(SH-SY5Y), GATA3(SH-SY5Y), MAFF(K562)
chr7:50301320-50301470	GATA2, GATA3, MAFF, RCOR1	GATA2(SH-SY5Y), GATA3(SH-SY5Y), MAFF(K562), RCOR1(K562)
chr7:50301600-50301750	GATA3	GATA3(SH-SY5Y)
chr7:50301880-50302030	ELF1	ELF1(K562)
chr7:50302280-50302430	IKZF1, MTA3, STAT3	IKZF1(GM12878), MTA3(GM12878), STAT3(GM12878)



ENCODE Element Browser: Candidate *cis*-regulatory element finder (Option 3)

- Under **Option 3: search cis-elements surrounding a gene**, enter ***IKZF1*** for gene and **1 kb** for window.

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 +/- 20kb) kb submit!

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

human (hg19) ▼ Gene name(Sox2, Nanog ...) submit!



ENCODE Element Browser: Candidate *cis*-regulatory element finder (Option 3)

Candidate *cis*-elements near your gene of interest.

Human (hg19)

Gene **IKZF1** (IKAROS, LYF1, ZNFN1A1) [NM_001220771, NM_006060, NM_001220765, NM_001220776, NM_001220772, NM_001220766, NM_001220775, NM_001220767, NM_001220770, NM_001220769, NM_001220768, NM_001220774, NM_001220773, ENSG00000185811, ENST00000413698, ENST00000331340, ENST00000343574, ENST00000346667, ENST00000349824, ENST00000357364]

Gene coordinate: chr7:50344264-50472798

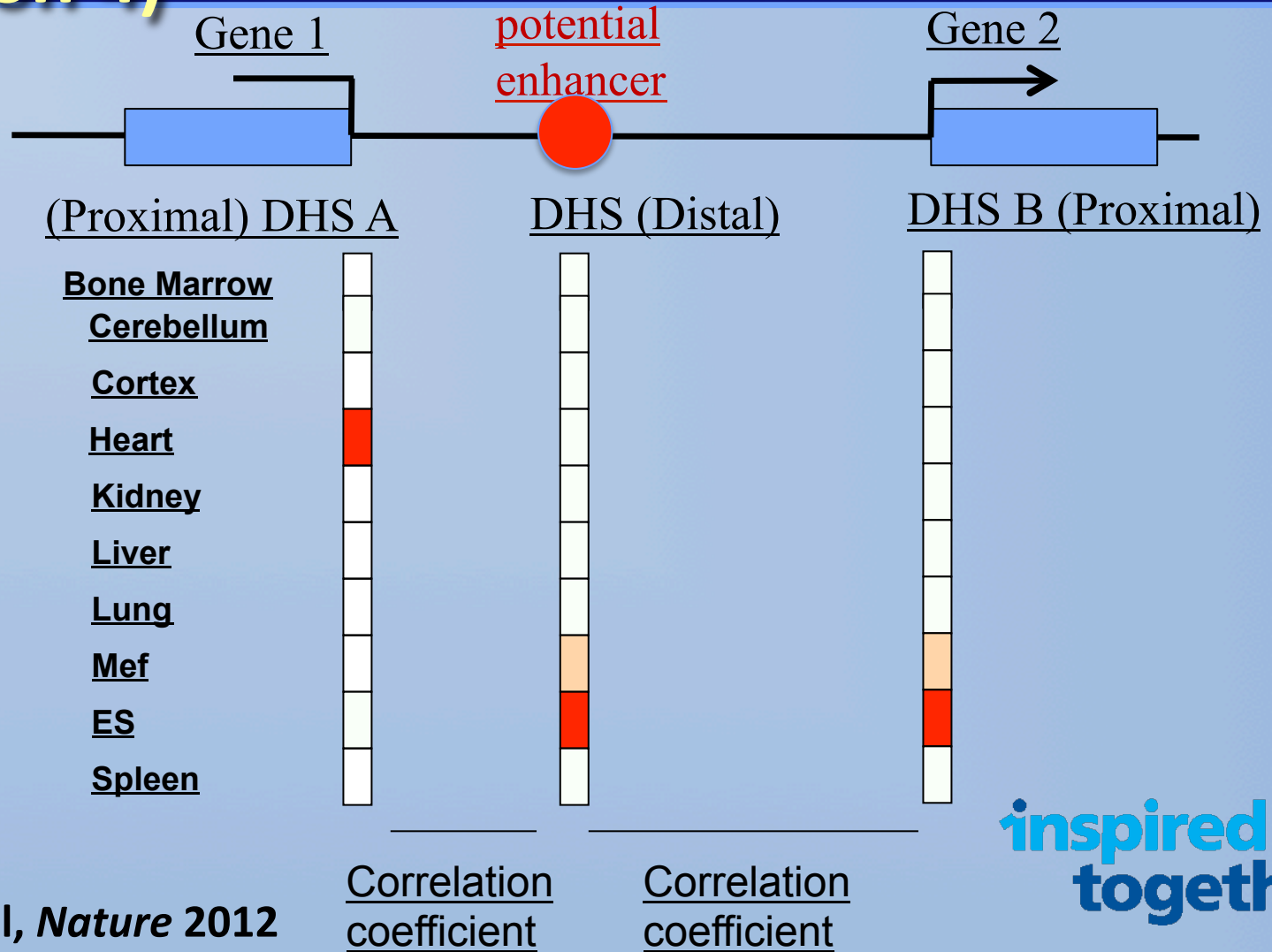
Extension: +/-1000 bp.

DNaseI Hypersensitive Sites:

Coordinate	Tissue/cell type
chr7:50343280-50343430	CD19 Primary Cells, CD20+ RO01778, CD4 Primary Cells, Fetal Brain, Fetal Intestine Large, Fetal Kidney, Fetal Kidney Right, Fetal Muscle Lower Limb Skeletal, Fetal Ovary, Fetal Placenta, Fetal Renal Cortex, Fetal Stomach, Fetal Thymus, Jurkat, Mobilized CD56 Primary Cells, Monocytes-CD14+ RO01746, Th2
chr7:50343580-50343730	Adult Th1, CD14, CD14 Primary Cells, CD19 Primary Cells, CD20+ RO01778, CD3 Primary Cells, CD34+ Mobilized, CD56 Primary Cells, CMK, Fetal Brain, Fetal Kidney, Fetal Kidney Right, Fetal Lung Right, Fetal Muscle Arm, Fetal Muscle Back, Fetal Muscle Leg, Fetal Muscle Trunk, Fetal Placenta, Fetal Renal Cortex, Fetal Renal Cortex Left, Fetal Renal Cortex Right, Fetal Renal Pelvis, Fetal Renal Pelvis Left, Fetal Renal Pelvis Right, GM06990, H1 Derived Neuronal Progenitor Cultured Cells, H1-hESC, HMEC, HepG2, K562, Mobilized CD3 Primary Cells, Monocytes-CD14+ RO01746, NB4



ENCODE Element Browser: DHS-Linkage (Option 4)



Thurmond et al, *Nature* 2012

ENCODE Element Browser: DHS-Linkage (Option 4)

- Under **Option 4**: search cis-elements **LINKED** to a gene based on DNaseI HSS specificity, enter *IKZF1* for gene.

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 +/- 20kb) kb submit!

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

human (hg19) ▼ Gene name(Sox2, Nanog ...) submit!

ENCODE Element Browser: DHS-Linkage (Option 4)

Cis-elements linked to your queried gene.

Human (hg19)

Gene **IKZF1** (IKAROS, LYF1, ZNFN1A1) [NM_001220771, NM_006060, NM_001220765, NM_001220776, NM_001220772, NM_001220766, NM_001220775, NM_001220767, NM_001220770, NM_001220769, NM_001220768, NM_001220774, NM_001220773, ENSG00000185811, ENST00000413698, ENST00000331340, ENST00000343574, ENST00000346667, ENST00000349824, ENST00000357364]

Cis-element lined by DNaseI Hypersensitive Sites Linkage:

Proximal DHS (TSS)	start	end	Gene	Distal DHS	start	end	correlation
chr7	50343520	50343670	IKZF1	chr7	49848340	49848490	0.730285
chr7	50343520	50343670	IKZF1	chr7	49879300	49879450	0.906691
chr7	50343520	50343670	IKZF1	chr7	49892860	49893010	0.834069
chr7	50343520	50343670	IKZF1	chr7	50052300	50052450	0.8736
chr7	50343520	50343670	IKZF1	chr7	50109340	50109490	0.808862
chr7	50343520	50343670	IKZF1	chr7	50132200	50132350	0.756843
chr7	50343520	50343670	IKZF1	chr7	50152320	50152470	0.781215
chr7	50343520	50343670	IKZF1	chr7	50207300	50207450	0.967203
chr7	50343520	50343670	IKZF1	chr7	50222440	50222590	0.887968

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

TF binding Site: <

Coordinate	TF	tissue
chr7:50207200-50207350	EBF1, ELF1	EBF1(GM12878), ELF1(GM12878)
chr7:50207380-50207530	EBF1, ELF1	EBF1(GM12878), ELF1(GM12878)

Part II: The 3D Genome Browser

**inspired
together**



Overview (<http://3dgenome.org>)

- Hi-C Genome Browser
 - Hi-C visualization
- Virtual 4C
 - Virtual 4C (derived from Hi-C Data)
 - ChIA-PET visualization
- Main Features
 - Easily browse some of the high-quality, published Hi-C data
 - **Contextualized** with customizable UCSC Browser
 - Browse your **own Hi-C data (BUTLR format)**

**inspired
together**



The Hi-C Genome Browser (SOX2)

[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

☐ 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 name (Sox2, Nanog ...) SOX2 Show Interaction	Chromosome chr1 Start End Show Interaction

Optional: User-defined UCSC Browser Session

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

Search by Gene
or Location

together

UCSC Browser
Session

< Move Left Move Right >

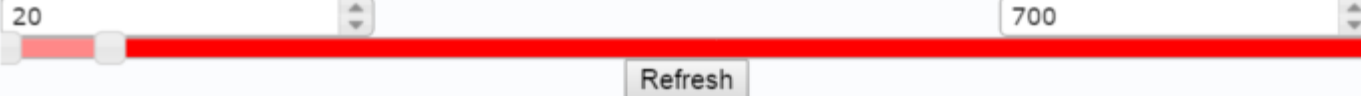
Zoom In (+) 3x Zoom Out (-) 0.3x

[Navigation Bar](#)

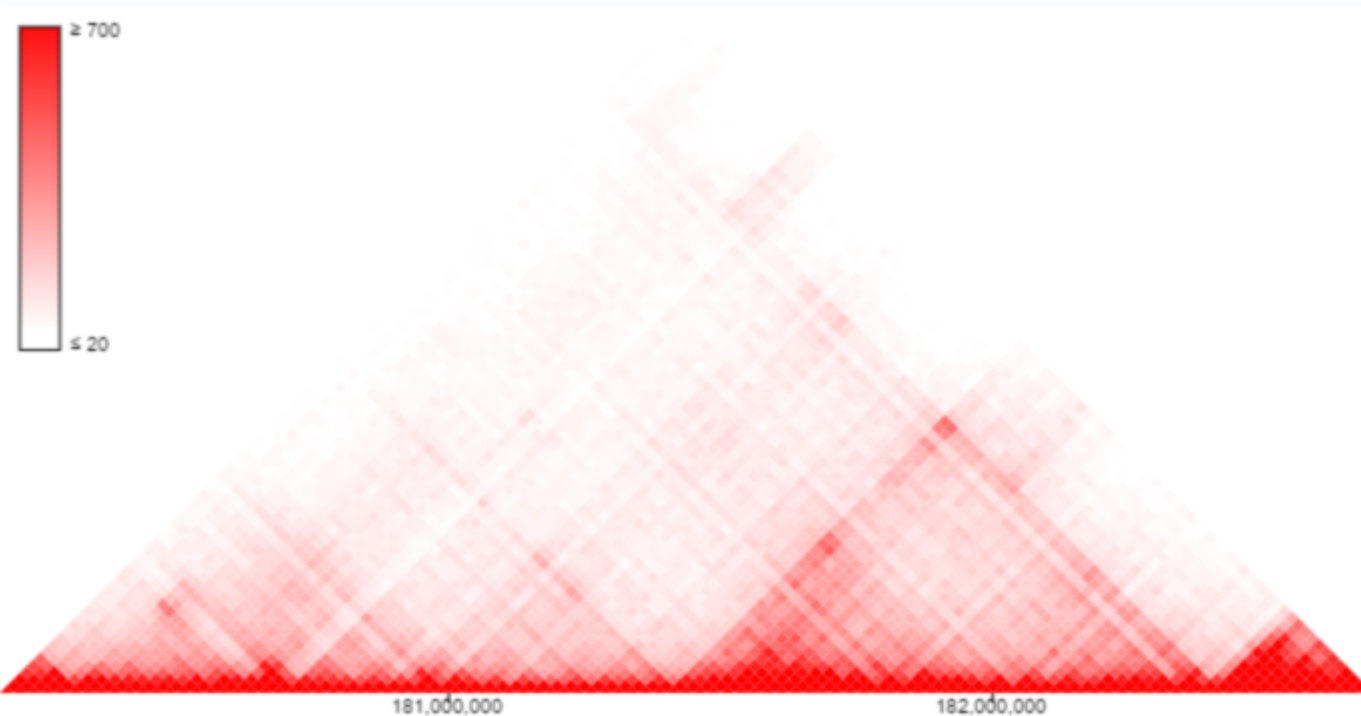
Hi-C data hg19 GM12878 at 25 kb is from "Rao SS, Huntley MH, Durand NC, Stamenova EK, Bochkov ID, Robinson JT, Sanborn AL, Machol I, Omer AD, Lander ES, Aiden EL. A 3D Map of the Human Genome at Kilobase Resolution Reveals Principles of Chromatin Looping. Cell. 2014;159:1665-1680."

Gene **SOX2** [NM_003106, ENSG00000181449, ENST00000325404]

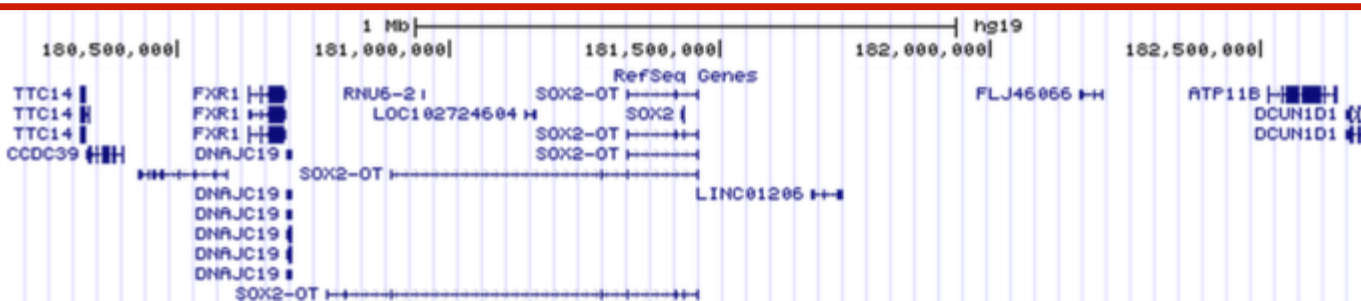
[Download Image](#)



[Intensity Adjustment Bar](#)



[Hi-C Heatmap](#)

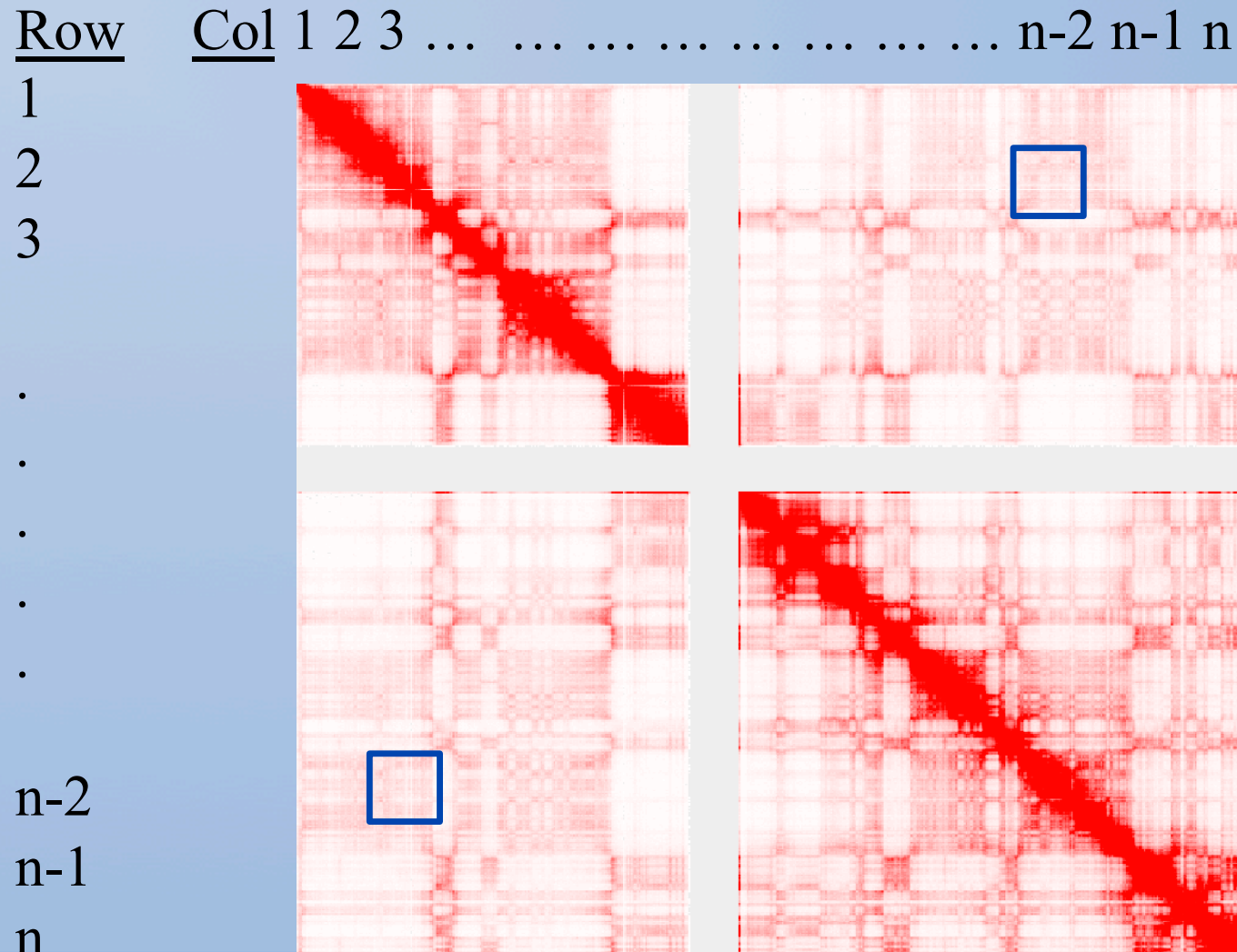


[UCSC Genome Browser](#)

inspired
together

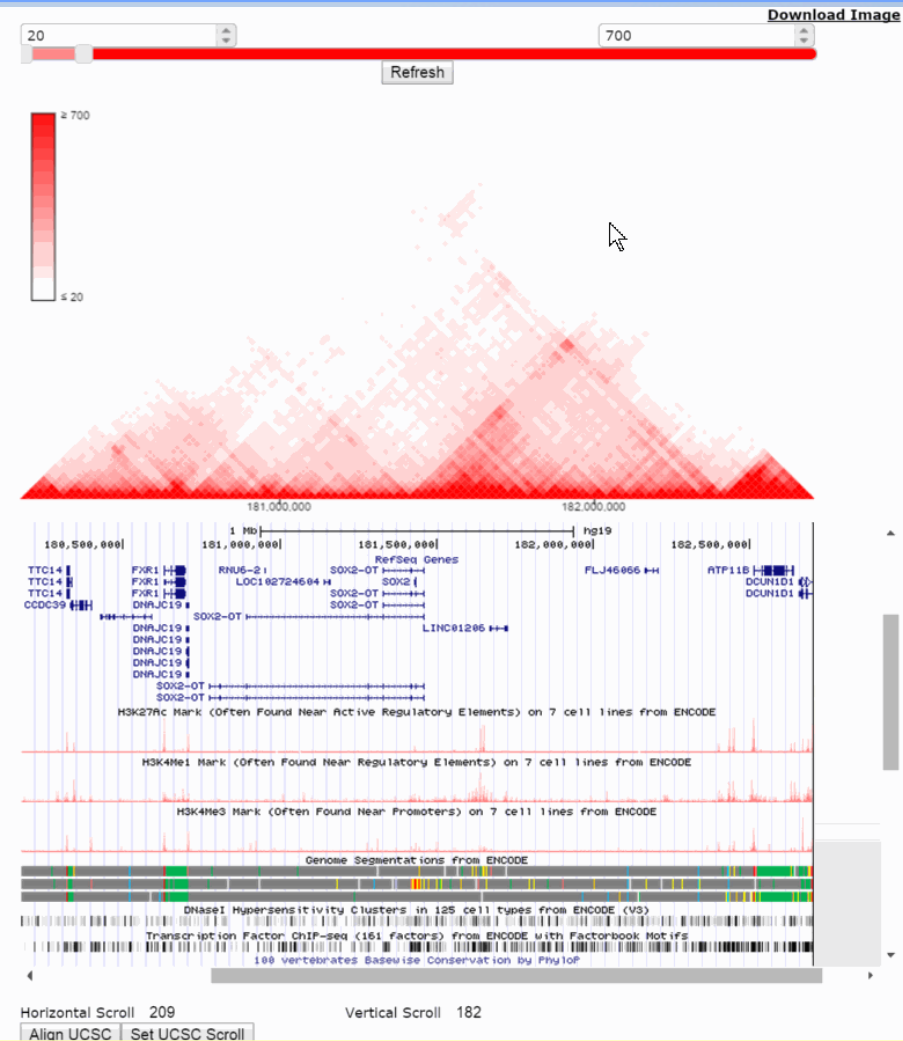
College of Medicine

The Contact Matrix yielded by Hi-C Data

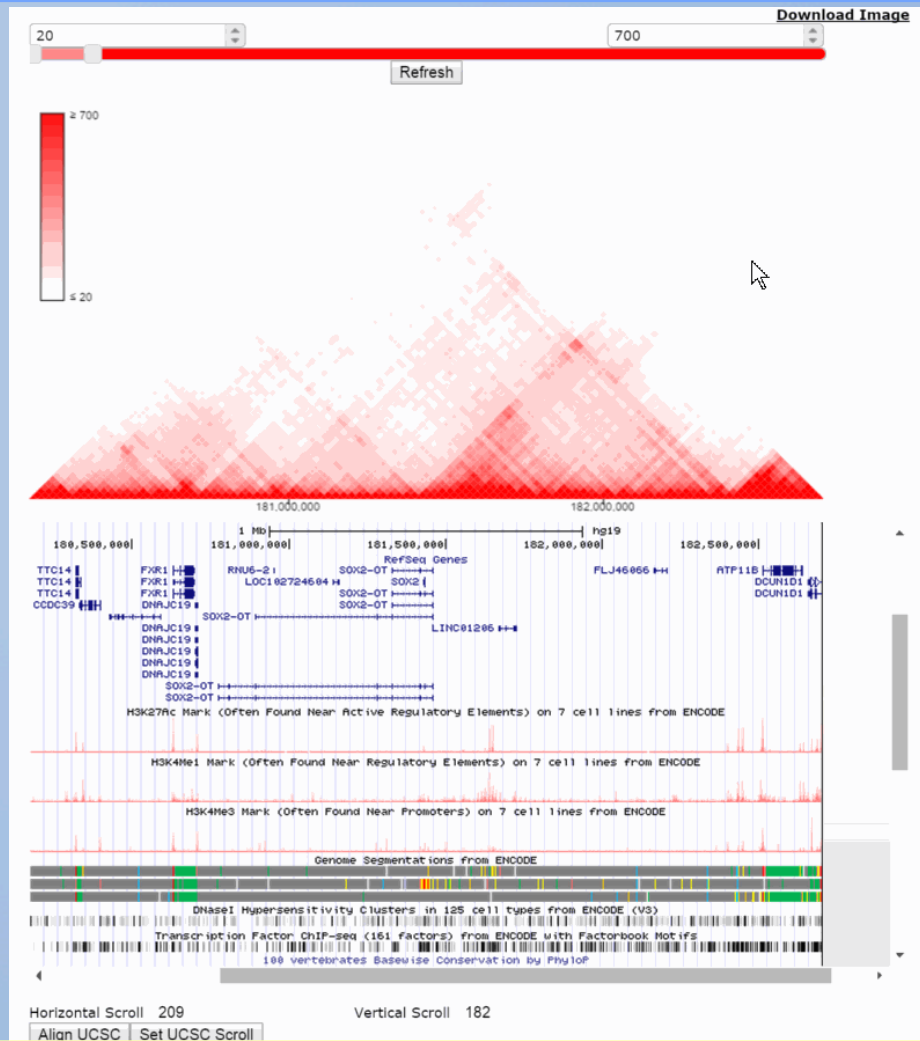


ired
together

Interpreting the Data

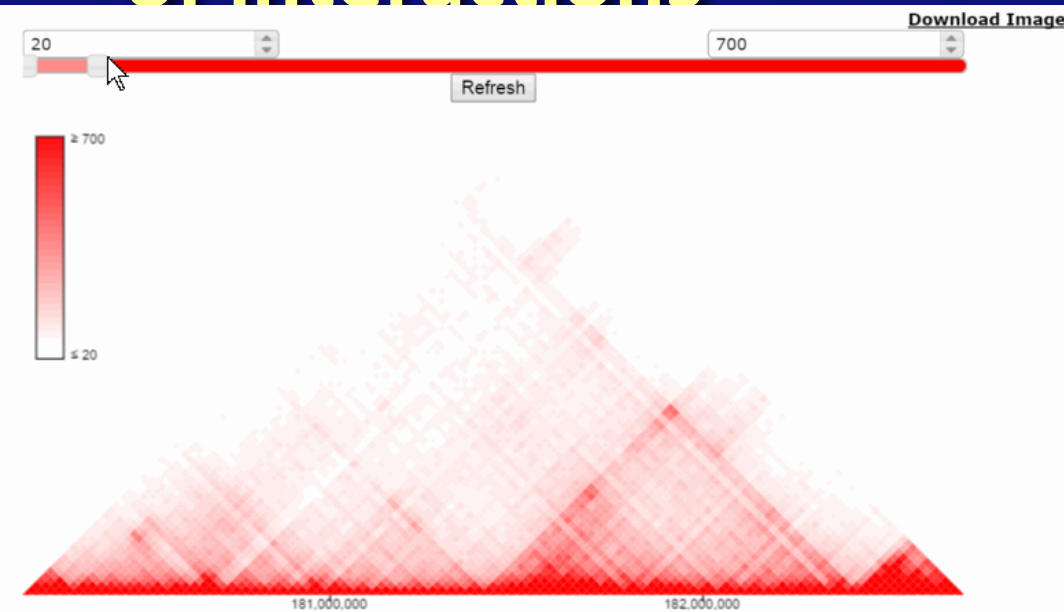


Single click on heatmap region to obtain the genomic regions



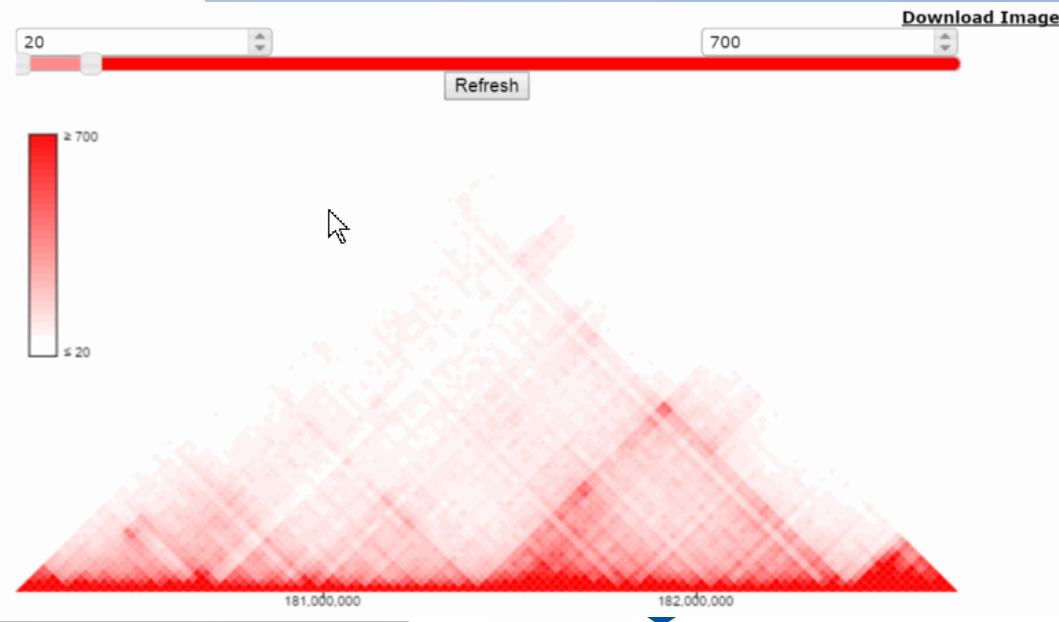
Double click on a region of interest to zoom in (Don't forget the navigation bar)

Adjusting intensity to highlight the scope of interactions

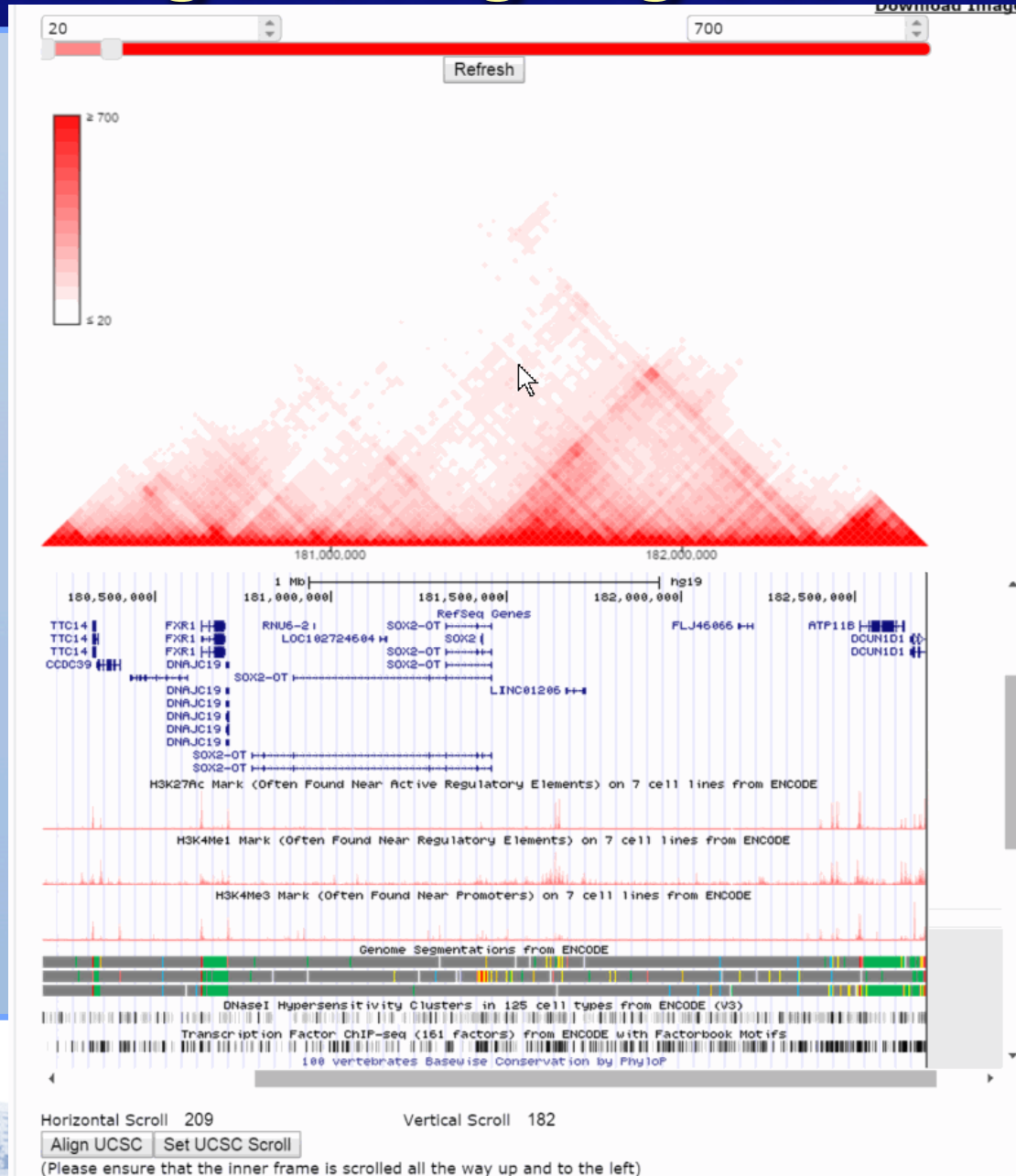


Adjust the minimum and maximum intensity cutoffs with the slider.

Change the intensity cutoffs with textbox arrows or directly enter the values, then clicking *refresh*



Customizing and Aligning UCSC Browser



inspired
together

Use Your Own Hi-C Data

Step 2. Choose the source of the data:

☐ Browse Available Hi-C Data

Tissue Resolution

OR

☒ Use Your Own Data

Data URL

Please specify the URL of the data (e.g. <http://promoter.bx.psu.edu/hi-c/example/GM12878.25kb.butlr>). Note that the data must be in BUTLR (**B**inary **U**pper **T**riang**L**ar Mat**R**ix) format. Ensure that the server supports HTTP and that the permissions are set appropriately.

- Similar to UCSC Genome Browser's visualization of bigWig/bigBed on remote servers, the 3D Genome Browser could visualize Hi-C matrices as **BUTLR** (**B**inary **U**pper **T**riang**L**ar Mat**R**ix) files.
- The **BUTLR** file format decreases disk memory usage and allows random access.

inspired
together

The Virtual 4C and ChIA-PET (*BCR1*)

[HOME](#) [HI-C INTERACTIONS](#) [VIRTUAL 4C](#) [DOWNLOAD](#) [TUTORIAL](#) [CITATION](#) [CONTACT](#)

Virtual 4C Created from Hi-C Data with DHS-Linkage and ChIA-PET

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

OR

☐ Use Your Own Data

Data URL

Step 3. Enter Gene Symbol or SNP rsid:

Gene Name/SNP rsid Extended region in kb (default $\pm$ 500kb)

ChIA-PET Data Selection

☒ RAD21 (Synder)

Optional: User-defined UCSC Browser Session

The browser is developed with Google Chrome 42*. While we test our browser on a variety of web browsers and operating systems, we recommend our users to use Chrome for the most optimal experience.

Gene Symbol, Refseq,
Uniprot
SNP rsid

pired
together

HERSHEY
College of Medicine

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, Alden EL. A 3D Map of the Human Genome at Kilobase Resolution Reveals Principles of Chromatin Looping. Cell. 2014;159:1665-1680.

Region or TSS @ chr22:23522551

Region or TSS @ chr22:23558935

Choice of Region

Gene **BCR** (BCR1) [NM_004327, NM_021574, ENSG00000186716, ENST00000305877] with TSS at chr22:23522551.[Download Image](#)

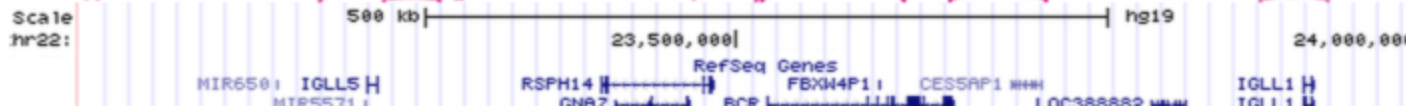
Virtual 4C Plot

[DHS-Linkage](#)

together

[ChIA-PET](#)

JE HERSHEY

[UCSC Genome Browser](#)

Thank you!

- Yue Laboratory
- Ross Hardison
- John Stamatoyannopoulos
- Zhiping Weng & ENCODE Encyclopedia Group
- ENCODE DCC
- ENCODE Outreach Group
- Funding: PhRMA foundation, Penn STATE CTSI, 5U01CA176063-02 (UCI)
- Feedbacks Welcome!



**inspired
together**

THE HUCK INSTITUTES
OF THE LIFE SCIENCES

PENN STATE HERSEY
 College of Medicine