

ENCODE Element Browser

Goal: to navigate the candidate DNA elements predicted by the ENCODE consortium, including gene expression, DNase I hypersensitive sites, TF binding sites, and candidate enhancers/promoters.

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

Option 1: Search Gene expression by gene, across all cell types:

Click HUMAN or MOUSE in the navigation bar to choose the right species; type in the official gene name (for example, sox2) and click submit.

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

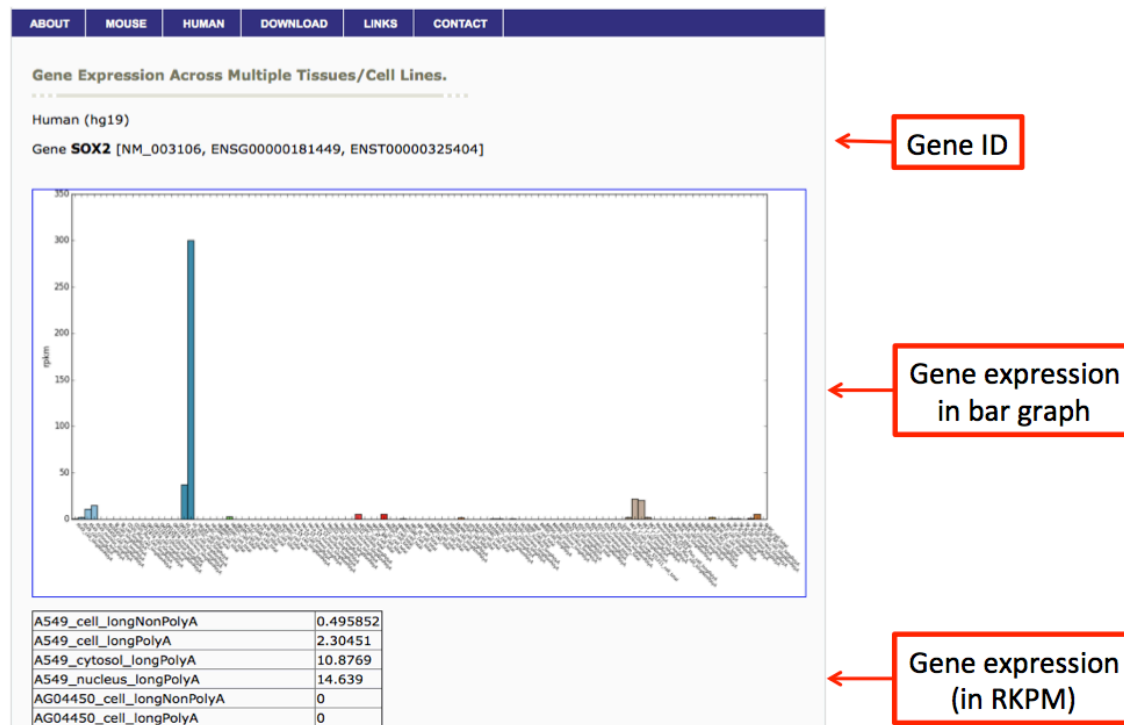
Human (hg19) Gene name(Sox2, Nanog ...) submit!

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!

You will be re-directed to the following result page.



In the upper panel of the result page, we plotted gene expression in a bar graph that you can directly use in your publication. In the table below the bar graph, we listed the RPKM values across all the tissue/cell types for the queried gene.

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

Under option 2, first choose chromosome from the dropdown menu and then type in the start and the end position (for example, chr3, 181000000, 182000000); click submit.

The screenshot shows a web interface for querying human ENCODE data. It features a navigation bar with links for HOME, MOUSE, HUMAN, DOWNLOAD, LINKS, and CONTACT. Below the navigation bar, there are four search options:

- Option 1: Search gene expression across ~ 60 human cell types (total 108 datasets)** - Includes a dropdown for 'Human (hg19)', a text input for 'Gene name (Sox2, Nanog ...)', and a 'submit!' button.
- Option 2: Search cis-elements in a given genomic region** - This option is highlighted with a red box. It includes a dropdown for 'human (hg19)', a dropdown for 'chr3', text inputs for 'start: 181000000' and 'end: 181200000', and a 'submit!' button.
- Option 3: search cis-elements surrounding a gene** - Includes a dropdown for 'Human (hg19)', a text input for 'Gene name (Sox2, Nanog ...)' with 'SOX2' entered, a text input for 'Extended region (default +/- 100kb)' with '20' entered, a 'kb' unit, and a 'submit!' button.
- Option 4: search cis-elements LINKED to a gene based on DNaseI HSS specificity** - Includes a dropdown for 'Human (hg19)', a text input for 'Gene name (Sox2, Nanog ...)', and a 'submit!' button.

First, we can see the experimentally determined DNase I Hypersensitive sites (DHS). The first column is the coordinates of DHS and the second column lists the tissue/cell type where the DHS is present. *Note:* at this time, the queried DHS sites only include 14 cell types where we have both DHS and H3K27ac ChIP-Seq data: *GM12878, H1-hESC, HeLa-S3, HepG2, HMEC, HSMM, HSMMtube, HUVEC, K562, NH-A, NHLF-Ad, NHEK, NHLF, Osteobl*. The list will be expanded soon to incorporate more cell types. For more information, please visit: <https://www.encodeproject.org/data/annotations/>

The screenshot shows the 'Candidate cis-elements in your queried region' section. It includes a navigation bar and a table of DNaseI Hypersensitive Sites. The table has two columns: 'Coordinate' and 'Tissue/cell type'. The species is identified as 'human_hg19' and the region as 'chr3:181000000-181200000'.

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

Scrolling down in the same page, we also listed the TF binding sites in this region. This table has three columns. The first column lists the coordinates of the TF binding sites. The second column list the transcription factor and the third column list in which tissue/cell types the TF binding sites are located.

TF binding Site:

Coordinate	TF	tissue
chr3:181010880-181011030	multiple	ATF2(GM12878), BATF(GM12878), EP300(GM12878), FOS(HUVEC), FOS(MCF10A-Er-Src), IRF4(GM12878), JUN(HUVEC), NFIC(GM12878), RUNX3(GM12878), STAT3(MCF10A-Er-Src)
chr3:181016320-181016470	CTCF, FOS, STAT3	CTCF(BE2_C), FOS(MCF10A-Er-Src), STAT3(MCF10A-Er-Src)
chr3:181034320-181034470	CEBPB, GATA3, REST, SP2, TCF7L2	CEBPB(HeLa-S3), GATA3(T-47D), REST(A549), REST(GM12878), REST(H1-hESC), REST(HeLa-S3), REST(HepG2), REST(K562), REST(PANC-1), REST(PFSK-1), REST(SK-N-SH), REST(U87), SP2(HepG2), TCF7L2(HCT-116)
chr3:181042360-181042510	CTCF, RAD21	CTCF(A549), CTCF(BE2_C), CTCF(Caco-2), CTCF(GM12892), CTCF(GM19240), CTCF(H1-hESC), CTCF(HEEpiC), CTCF(HMEC), CTCF(HepG2), CTCF(MCF-7), CTCF(NHEK), CTCF(RPTEC), CTCF(SAEC), CTCF(SK-N-SH_RA), CTCF(T-47D), CTCF(WERI-Rb-1), RAD21(H1-hESC), RAD21(HepG2), RAD21(SK-N-SH_RA)

Option 3: Search cis-elements surrounding a gene:

This option is similar to option 2, but it searches by gene names rather than a given genomic locus. Therefore, the users can conveniently type in the gene name of interest and specify the extended region around the gene TSS. The DHS sites and TF binding sites will be displayed in a similar pattern as in option 2. For example, please type in sox2 and 20 kb for the extended region.

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 ...) <input type="text"/> submit!					
Option 2: Search cis-elements in a given genomic region					
human (hg19) chr1 start: <input type="text"/> end: <input type="text"/> submit!					
Option 3: search cis-elements surrounding a gene					
Human (hg19) Gene name(Sox2, Nanog ...) <input type="text" value="sox2"/> ←					
Extended region (default +/- 100kb) <input type="text" value="20"/> ← kb submit!					
Option 4: search cis-elements LINKED to a gene based on DNaseI HSS specificity					
Human (hg19) Gene name(Sox2, Nanog ...) <input type="text"/> submit!					

Option 4: Search cis-elements linked to a gene based on DHS:

This option will help users predict the linkage between distal regulatory elements with their target genes. The prediction is based the tissue specificity of DHS across 129 cell types. Users can type in the gene name under option 4 and then click submit (sox2). Below is the result:

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

Columns 1-3 are the loci for the proximal DHS near the queried gene. Column 4 is the gene name. Columns 5-7 are the distal DHS that are associated with the queried gene. The last column is the Pearson correlation between the distal and proximal DHS (for more details, please refer to Thurman, et al, Science 2012)

Data: All the human data used in this query website is from:

<https://www.encodeproject.org/data/annotations/>. All the mouse data can be downloaded from: <http://mouseencode.org>.

3D genome browser

Goal: To develop user-friendly browser to visualize Hi-C and Chia-PET data.

Main features:

1. Easy navigation functions such as zooming in/out, moving to left/right.
2. The ability to incorporate customized UCSC genome browser session.
3. Gene centric view: allow users to predict linkage between enhancers and their target genes, based on integrated views of Hi-C/ChIA-PET data and the linkage based on correlation of distal and proximal DHS sites.
4. Allow query human trait and disease-associate SNPs and find regions interact with them through chromatin looping.
5. Use your own Hi-C data.

Link: <http://3dgenome.org>

Generate Hi-C Interaction heatmap

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 <input type="text" value="human"/> Assembly <input type="text" value="hg19"/>						
Step 2. Choose the source of the data:						
<input checked="" type="radio"/> Browse Available Hi-C Data						
Tissue <input type="text" value="GM12878"/> Resolution <input type="text" value="25kb"/>						
OR						
<input type="radio"/> Use Your Own Data						
Data URL <input type="text"/> Resolution (in kb) <input type="text"/>						
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"/>				Chromosome <input type="text" value="chr1"/>		
<input type="text"/>				Start <input type="text"/>		
<input type="text"/>				End <input type="text"/>		
<input type="button" value="Show Interaction"/>				<input type="button" value="Show Interaction"/>		

1. Click "HI-C INTERACTIONS" tab;
2. Using dropdown menu to choose species and assembly.
3. Choose the source of Hi-C data:

Option 1. Use dropdown menu to choose the published Hi-C data;

Option 2. Specify the location of your own Hi-C data. Note: it needs to be converted to a binary file using our provided script.

4. Choose genomic region to show Hi-C interaction

Option 1: Search by Gene Name(RefSeq ID or Uniprot ID). By default, this regions will be ± 50 bins, with each bin size = matrix resolution;

Option 2: Search by genomic locus.

Example: 1) Choose human, hg19; 2) choose “Browse available Hi-C data”, GM12878, 25kb 3) type sox2 in the under Option 1 and click “show Interaction”. Below is the Hi-C interaction map around sox2 region.

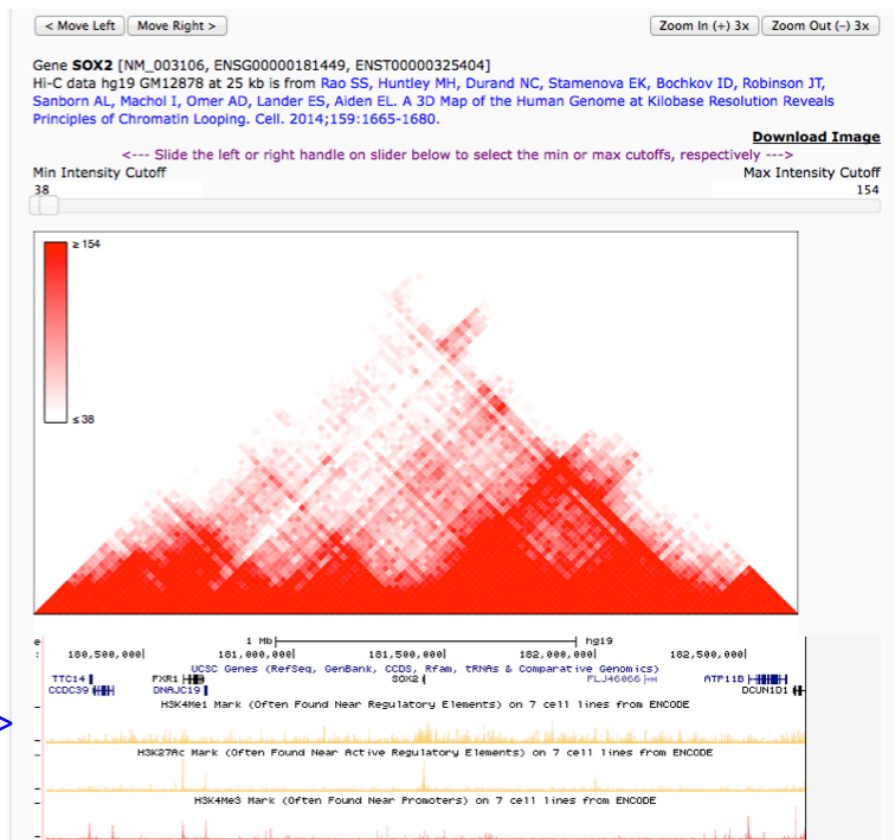
Navigation: ->

Data source: ->

Dynamic
color bar: ->

Hi-C Heatmap ->

Imbedded UCSC
Genome browser->



Navigate Hi-C data:

1. The "Move Left" and "Move Right" buttons allow users traversing the current genomic regions in the upstream or downstream directions, respectively.

2. The "Zoom In" and "Zoom Out" buttons allow users zoom in or zoom out from the current genomic region by three times. In order to not overload the server, we only support a maximum region of 1500 bins currently.

3. Adjusting color scale: The minimum and maximum intensity cutoff values are important in determining the colors of Hi-C heatmaps. As the values inside the bins of the contact matrix reflect the number of interactions, any values below the minimum cutoff value would be shown as white and any values above the maximum cutoff value would be shown as red.

Download Image: To save the image, right click this link and select "Save Link As" (this might be different depending on the web browser). In the resulting dialog box, enter an appropriate name and click "Save" to save the heatmap as a jpeg.

Use your own UCSC genome browser session

Users can locate your UCSC genome browser session and copy and paste the session ID in the 3D genome browser, so that they can overlay the Hi-C data with other genomics/epigenomics data. The UCSC genome browser session ID is the string after "hgsid=".

genome.ucsc.edu/cgi-bin/hgTracks?db=mm9&position=chr2%3A4756914-4883237&hgsid=424958157_7YvpPIULEHA4KnAWcZQAzf4JM7n7

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

User your own Hi-C data

A useful feature for our 3D genome browser is that you can visualize your own Hi-C data. This feature is important to work with unpublished data, as only the users can know and specify the location of their Hi-C data. However, users need to convert the standard Hi-C interaction matrix to binary format, using a script that we provide. Next, a user can put the binary Hi-C data in an internet accessible place and copy the link in our browser.

Potential pitfall: Please make sure the genome assembly and bin size is set correctly.

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:

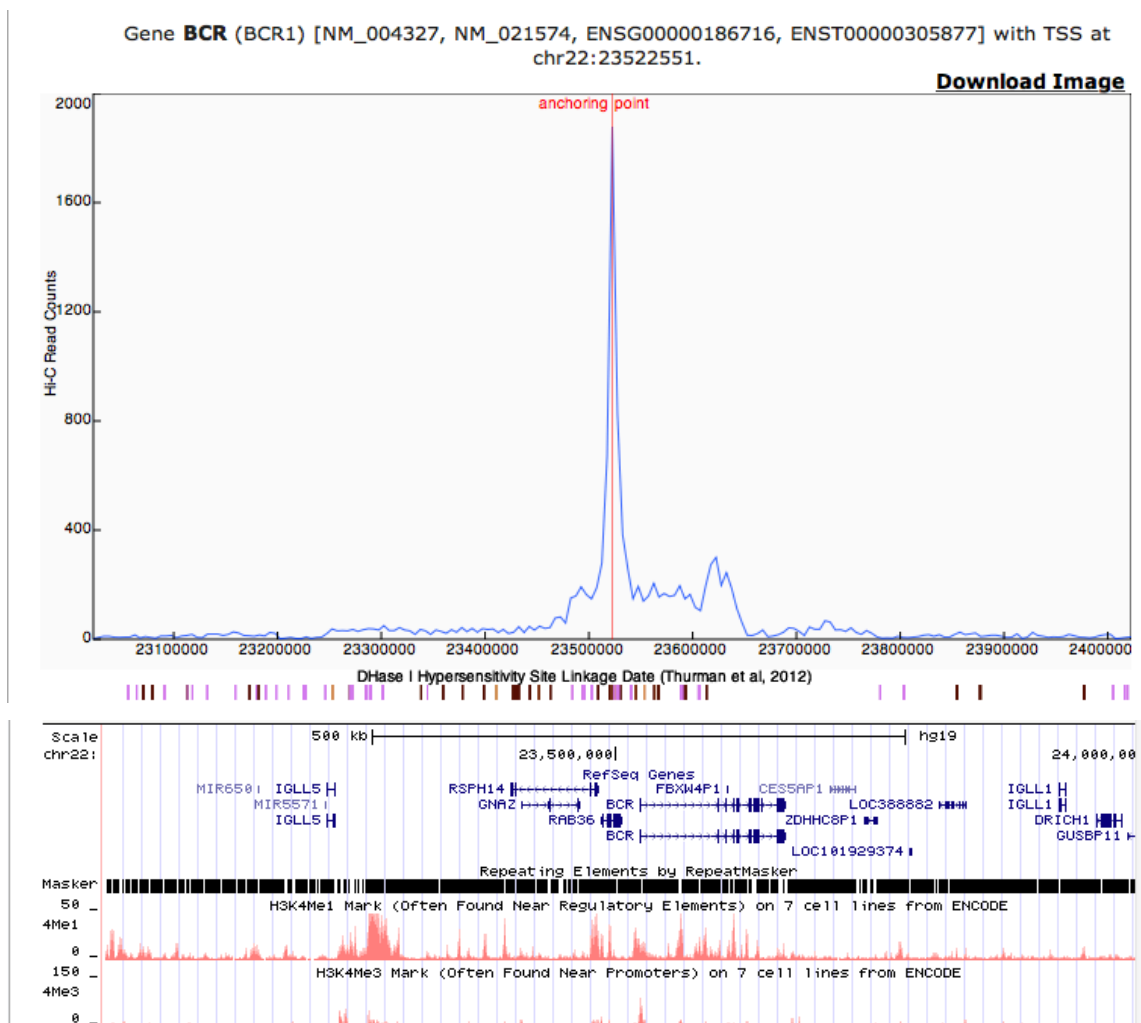
Browse Available Hi-C Data
Tissue Resolution

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.

Virtual 4C



Rather than visualizing **all** the interactions in a given region, some users are more interested to see the chromatin interactions with one anchor point, for example, a gene

promoter region. This information might be useful to predict the enhancers that regulate their genes of interests. To address this request, we can extract a row of data from a Hi-C interaction matrix, centered at the queried gene, and plot it in the follow figure. This figure is termed “virtual 4C”, as the way the data is presented is the same as 4C (Circularized Chromosome Conformation Capture), but we use Hi-C instead.

Example: In the above example, click virtual 4C, choose human, hg19 and type in BRC1. The vertical red line is the promoter of BRC1. Be default, we display +-500kb around the anchor point. The value in y-axis is the interaction frequencies measured by Hi-C. For this feature, you can either use gene name or a SNP ID.

Immediately below the virtual 4C track, the color bars indicate the distal DHS sites that are correlated to the proximal DHS near the queried gene.

On the lower panel of the result page, we show the histone modification in the same cell type, which can be used to predict candidate enhancers.

Therefore, by visualizing the virtual 4C, linkage by DHS and Epigenomics data generated by ENCODE, the query provide a powerful tool to predict the linkage between target genes and their distal regulatory elements.