

Ensembl Regulation

The aim of Ensembl Regulation is to annotate the genome with functionally active regions. This is done using data from a variety of sources: at present **ENCODE** and **Roadmap Epigenomics**, with plans to extend this to other sources, such as **Blueprint**, in the future.

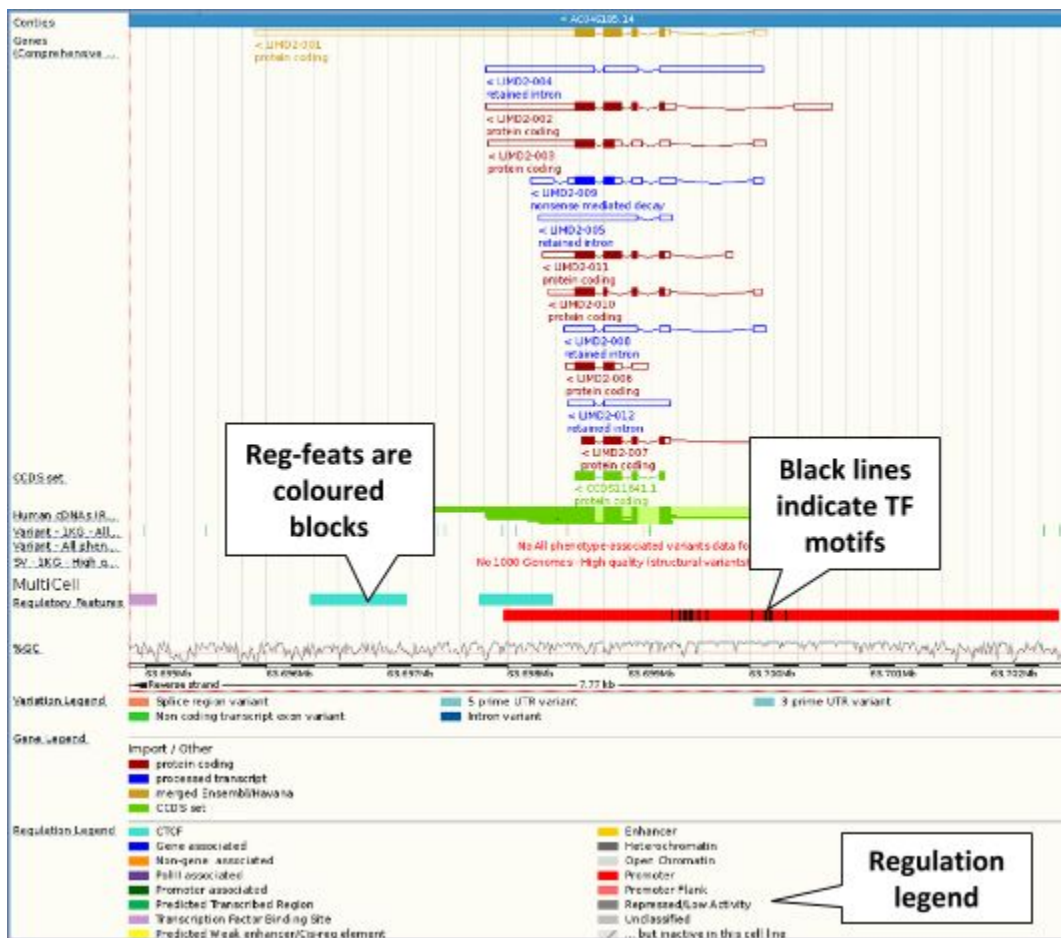
As well as presenting signal and peaks of the raw data from these sources, Ensembl integrate these to predict **Regulatory Features**, or Reg feats. Reg feats are things like promoters, enhancers, regions of open chromatin and CTCF binding sites. They have fixed boundaries and predicted activity across the cell types studied, based on the evidence from ENCODE, Roadmap Epigenomics *et al.*

For more information about Ensembl regulation, go to:
<http://www.ensembl.org/info/genome/funcgen/index.html>

Demo: Regulatory Features

We're going to have a look for regulatory features in the region of a gene and investigate their activity in different cell types. We'll start by searching for the gene *LIMD2* and jumping to the [location tab](#). Zoom out a little to see the gene plus some flanking regions.

The [MultiCell regulatory features](#) are shown by default.



In this region we can see a large red promoter, two turquoise CTCF binding sites and a lilac transcription factor binding site (don't worry if you have zoomed out further or not as far and can see more/less). Refer to the legend at the bottom to see what the colours mean.

You can also click on the regulatory features to learn more. Click on the red promoter to get a pop-up.

Regulatory Feature - MultiCell

Stable ID [ENSR00001537344](#)

Type Promoter

Core bp [17:63697964-63702563](#)

Attributes -

Motif Information

Name	PWM ID	Score
Egr1	MA0341.1	1
Egr1	MA0366.1	1
ETS1	MA0098.2	0.949
Egr1	MA0162.2	0.97
EBF1	MA0154.2	0.985
Egr1	MA0366.1	1
Egr1	MA0366.1	1
Egr1	MA0341.1	1
Egr1	MA0337.1	1
Egr1	MA0162.2	0.963
Egr1	MA0162.2	0.96
Egr1	MA0162.2	0.981
Egr1	MA0337.1	1
Egr1	MA0337.1	1
Egr1	MA0337.1	1

Stable ID of the feature

TF motifs within the feature

Click on the stable ID, [ENSR00001537344](#), to jump to the Regulation tab.

Regulatory Feature: ENSR00001537344

Summary | Details by Cell type | Feature Context | Source Data

Classification Promoter

Location Chromosome 17: 63,697,964-63,702,563

Active in 6/18 (DND-41, GM12878, HUVEC, HepG2, IMR90, Monocytes-CD14+)

Summary ⓘ

Select cells (showing 18/18)

Genes (Comprehensive...)

Contigs (Comprehensive...)

MultiCell

Regulatory Features

A549

Regulatory Features

DND-41

Regulatory Features

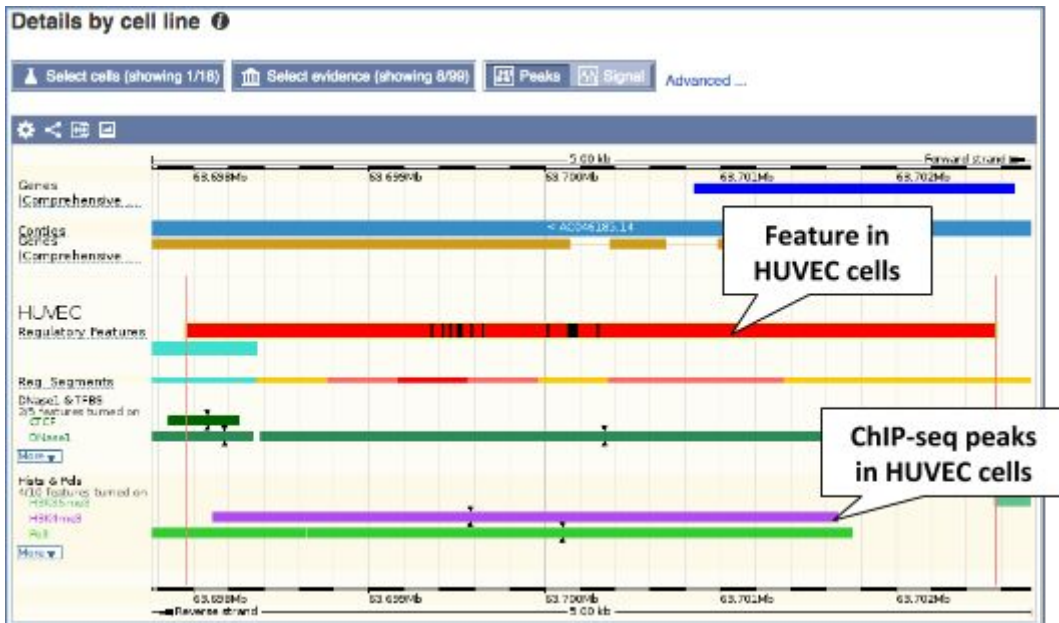
Different views available for the feature

Summary of cells the feature is active in

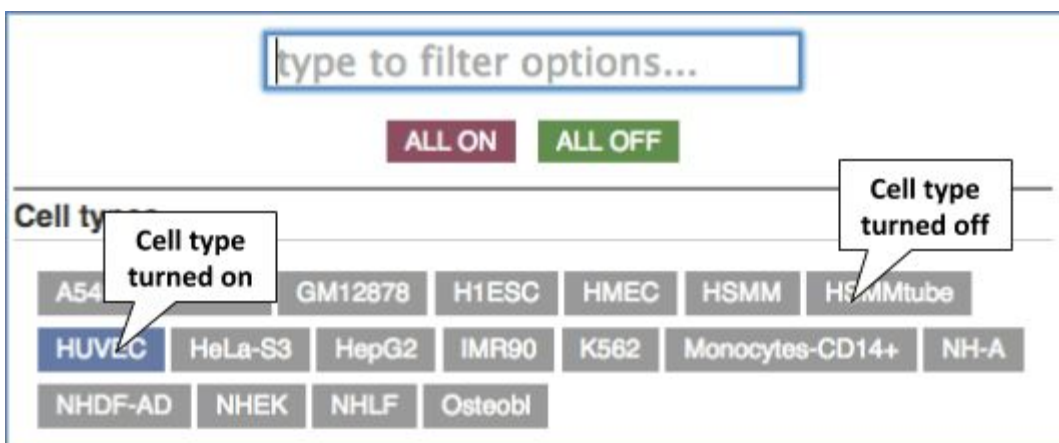
Inactive feature

Active feature

We can see that this promoter is active in six out of the 18 cell types currently in Ensembl. We can explore more detailed data in [Details by Cell type](#) – click on the button at the top.



At the moment, this page is only displaying data in HUVEC cells and only for a limited amount of evidence. Click on [Select cells](#) to add more.



We can add cells by clicking on them. If the cell type is turned on it's blue, if it's off it's grey. You can turn them on or off by clicking on them, or turn everything on or off using the buttons at the top.

Let's add a cell type where the promoter is inactive – [HeLa-S3](#).
Now close the menu.

We can change which evidence we can see, using the [Select evidence](#) button.



Select evidence (showing 0/99)

type to filter options...

ALL ON **ALL OFF**

Histone

H2A/B: H2AK5ac H2AZ H2BK5ac H2BK12ac H2BK15ac H2BK20ac H2BK120ac

H3:

H3K4ac H3K4me1 H3K4me2 H3K4me3 H3K9ac H3K9me1 H3K9me3 H3K14ac H3K18ac
H3K23ac H3K23me2 H3K27ac H3K27me3 H3K36me3 H3K56ac H3K79me1 H3K79me2

H4: H4K5ac H4K8ac H4K20me1 H4K91ac

Transcription Factor

ATF3 BAF155 BAF170 BATF BCL3 BCL11A BCLAF1 BHLHE40 Brg1 Cios Cjun Cmyc
CTCF CTCFL E2F8 EBF1 Egr1 ELF1 ETS1 FOSL1 FOSL2 FOXA1 FOXA2 Gabp Gata2
GTF2B HDAC2 HDAC8 HEY1 HNF4A HNF4G Ini1 IRF4 Junb Jund Max MEF2A MEF2C
Nanog Nfe2 NFKB NR4A1 Nrf1 Nrfp p300 Pax5 Pbx3 POU2F2 POU5F1 PU1 Rad21
RXRA Sin3Ak20 SIX5 SP1 SP2 Srf TAF1 TAF7 Tcf12 THAP1 Tr4 USF1 XRCC4 Yy1
ZBTB7A ZBTB33 ZEB1

Polymerase

PolII PolIII

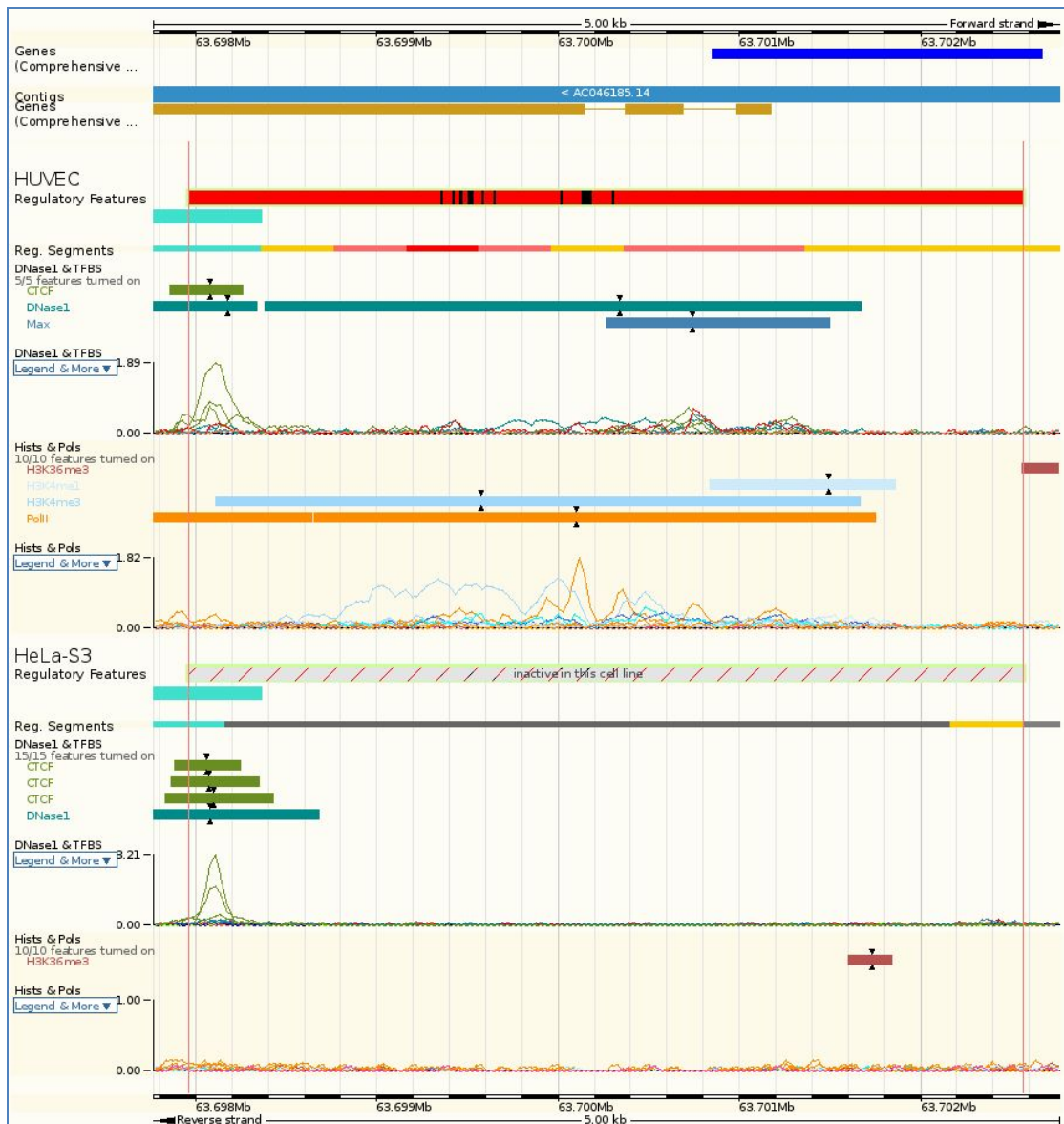
Open Chromatin

DNase1

Choose **ALL ON** to get all the possible evidence, then close the menu.

Lastly, we are currently only seeing the peaks read density. In order to see the signal too, select the [Signal](#) button.





Now we can see the active feature in HUVEC compared to the inactive feature in HeLa-S3. In HUVEC, we can see peaks of Max and PolII binding across the promoter, plus H3K4me3 and H3K4me1 modifications and DNase I sensitivity, whereas there is no such activity in HeLa-S3. In contrast, the CTCF binding site at the left is active, and shows CTCF binding and DNase I sensitivity in both cell types.

If you would like to see these data in table format, go to [Source data](#).

Source data ⓘ

Show **All** entries Show/hide columns Filter

Cell type	Evidence type	Feature name	Location
A549	DNase1 & TFBS	DNase1	19:7067084-7067515
A549	DNase1 & TFBS	DNase1	19:7069053-7070082
A549	Hists & Pols	H3K4me3	19:7068838-7069175
A549	Hists & Pols	H3K4me3	19:7069272-7070762
A549	Hists & Pols	H3K9ac	19:7069336-7070483
A549	Hists & Pols	H3K4me2	19:7069846-7070846

If you're interested in looking at regulatory features in detail across a region, you can do so in the [location tab](#). Click back using the tabs at the top.

Now click on [Configure this page](#). Go to [Regulatory features](#) in the left hand menu.

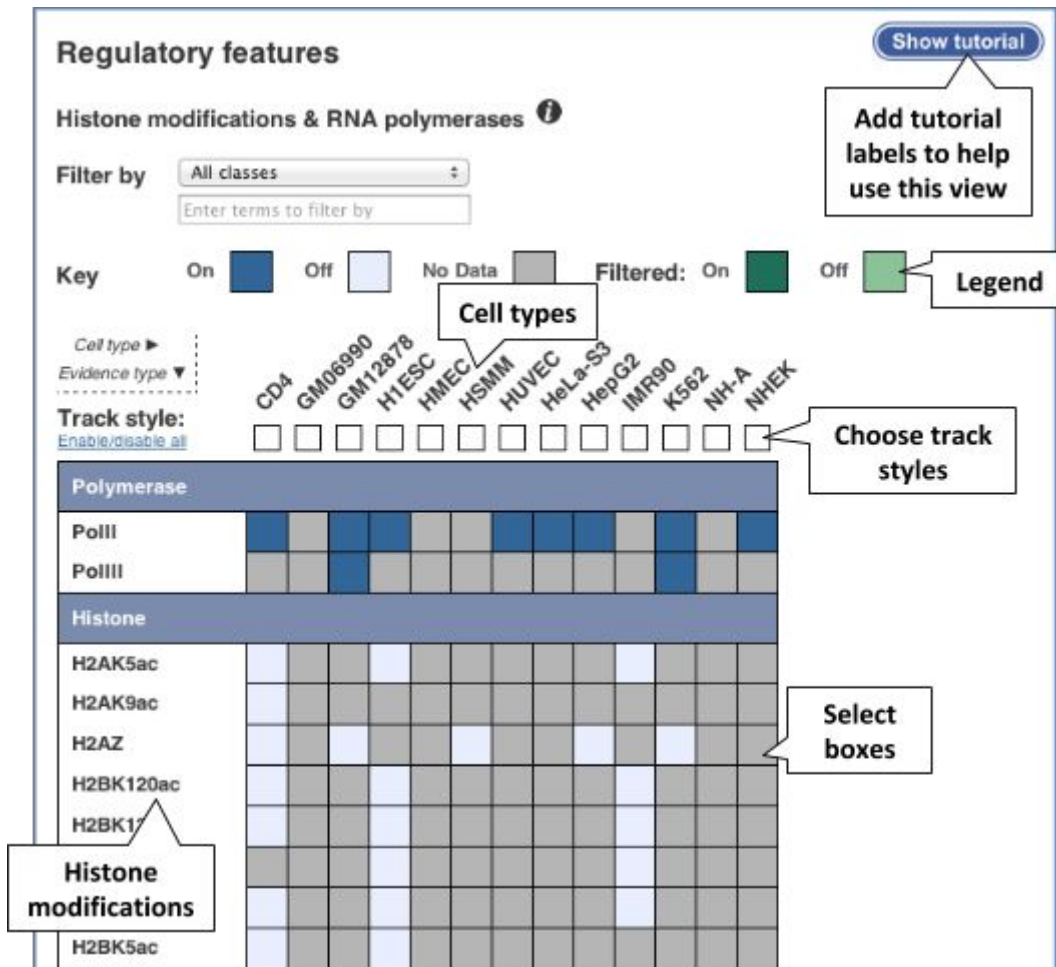
Regulation

Enable/disable all Regulatory features

- Reg. Feats ★ ⓘ
- Reg. Feats: A549 ★ ⓘ
- Reg. Feats: DND-41 ★ ⓘ
- Reg. Feats: GM12878 ★ ⓘ
- Reg. Feats: H1ESC ★ ⓘ
- Reg. Feats: HMEC ★ ⓘ
- Reg. Feats: HSMM ★ ⓘ
- Reg. Feats: HSMMtube ★ ⓘ
- Reg. Feats: HUVEC ★ ⓘ
- Reg. Feats: HeLa-S3 ★ ⓘ

The MultiCell [Reg. Feats](#) are already on. Turn on the tracks for the [Reg. Feats: HUVEC](#) and [Reg. Feats: HeLa-S3](#).

We can also turn on the evidence tracks. There are two menus for this: [Open chromatin & TFBSs](#) and [Histones & Polymerases](#). Open the menu for [Histones & Polymerases](#).



You can turn on a single track by clicking on the box in the matrix. Note that certain tracks are selected for all cell lines by default (PolII, PolIII, H3K27me3, H3K36me3, H3K4me3, H3K9me3). These will appear in the Region in detail view only if you specify a track style for the cell lines.

Turn on all the tracks for [HeLa-S3](#) and [HUVEC](#). Hover over the cell line name then select All.

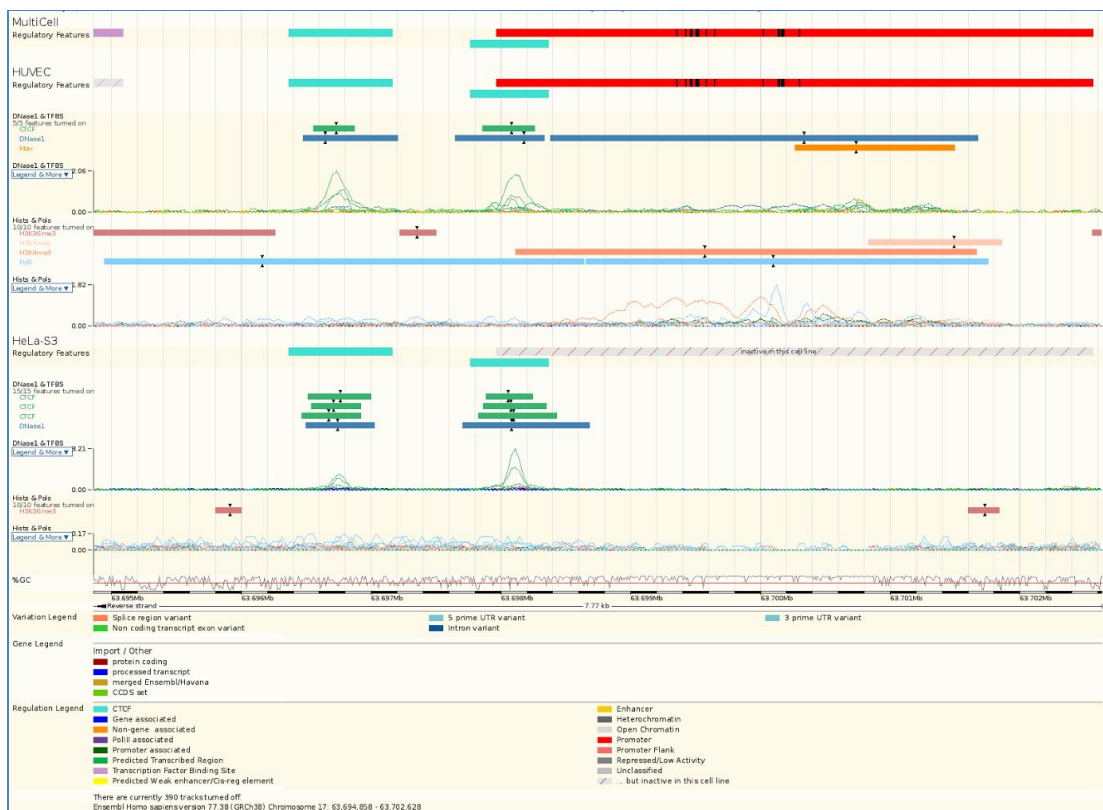


Now choose the track style for the tracks you've switched on. Click on the track style box for [HeLa-S3](#) and [HUVEC](#) and select [Both](#).



There is a similar matrix for [Open chromatin & TFBS](#). Use this to turn on all tracks for [HeLa-S3](#) and [HUVEC](#) in [Both](#). Now close the menu.

We can now see regulatory activity across the region in both cell types.



You can also get regulation data in the gene tab, by clicking on [Regulation](#) in the left-hand menu.











Demo: Regulation track hubs

Our regulatory data incorporates data from sources such as ENCODE, Blueprint and Roadmap Epigenomics. To see the full data directly from these sources, you can add track hubs.

From ensembl.org, click on [Trackhubs](#).



This page lists various track hubs that can be added to Ensembl.

Datahub name	Description	Species and assembly
Blueprint Hub	Blueprint Epigenomics Data Hub	 Human (GRCh37)
ENCODE Analysis Hub	ENCODE Integrative Analysis Data Hub	 Human (GRCh37)
Broad Improved Canine Annotation v1	Broad Institute CanFam3 Improved Annotation Data v1	 Dog (CanFam3)
Cancer genome polyA site & usage	An in-depth map of polyadenylation sites in cancer (matched-pair tissues and cell lines)	 Human (GRCh37)
CEMT (CEEHRC)	Epigenomic Data tracks from BCGSC, Vancouver	 Human (GRCh37)
DNA Methylation	DNA Methylation Hundreds of analyzed methylomes from bisulfite sequencing data	 Human (GRCh37)
		 Human (NCBI36)
		 Mouse (GRCm38)
		 Mouse (NCBI37)
		 Chimpanzee (CHIMP2.1.4)

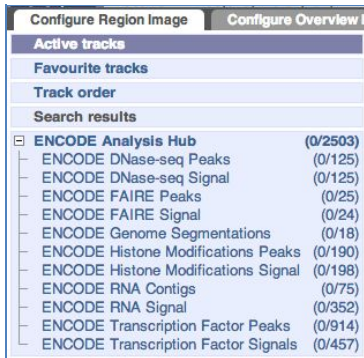
The table contains a brief description of the hub, plus the assembly that the hub is based on, as a link. Click on the link to turn on the hub. If the hub is based on a genome assembly which is not the current assembly in Ensembl, the link will also jump you to an archive with the previous assembly.

These often contain vast amounts of data, which can slow Ensembl down, so only add them if you need them, and trash them when you are finished with them.

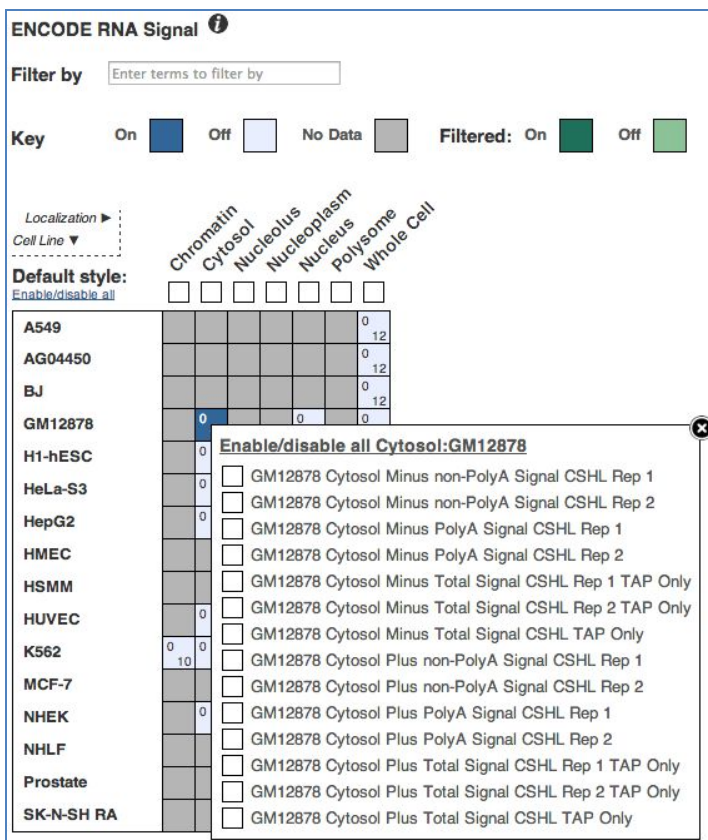
Click on the link [Human \(GRCh37\)](#) for the [ENCODE Analysis Hub](#).

This will take you directly to the [Personal data](#) menu in the [Region in detail](#) view. Because this is a GRCh37 hub, this has taken you to our dedicated [GRCh37 site, grch37.ensembl.org](http://grch37.ensembl.org).

Go to [Configure Region Image](#) to see that a new category has been added to your menu.



Open these menus to find the ENCODE matrices, which work in the same way as the Open chromatin & TFBS and Histones & polymerases matrices, except that some have multiple options (indicated by numbers within the boxes). If you click on these boxes, you can choose which of these options to add.



Demo: BioMart

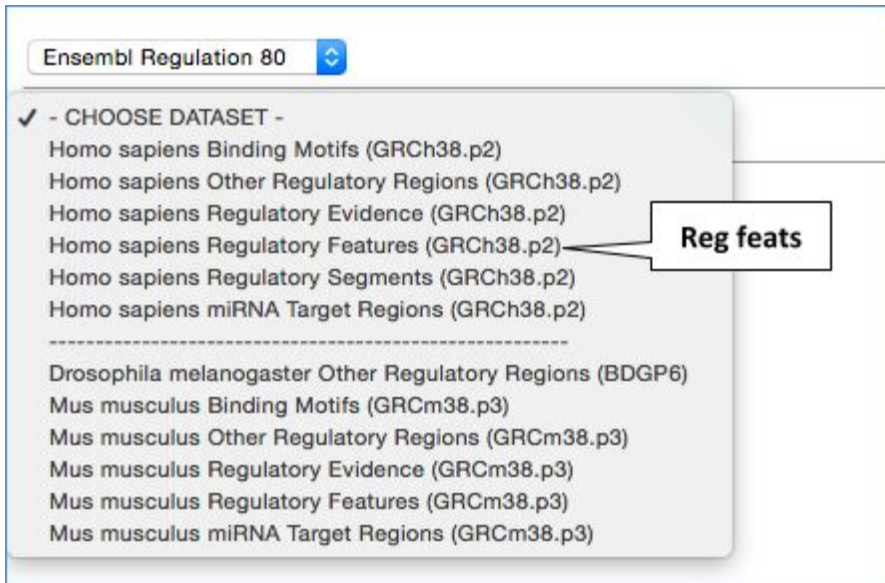
Regulatory features and evidence are also available via Ensembl BioMart. We'll do a query where we filter by a list of regulatory feature IDs to determine what kind of features they are and where they are in the genome. Here is the list of IDs:

ENSR00001601181	ENSR00001567543
ENSR00001601182	ENSR00000556855
ENSR00001601183	ENSR00000556857
ENSR00001601184	ENSR00000556858
ENSR00001601185	ENSR00000556859
ENSR00001567544	ENSR00000556863
ENSR00000556865	ENSR00000556867
ENSR00001567547	

Start at ensembl.org and click on [BioMart](#) in the top blue bar.

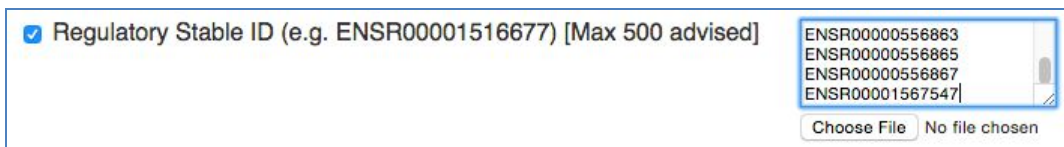


Choose [Ensembl Regulation 80](#) as the database. This gives you an option to choose a further database. Since we are working with a list of regulatory features, choose the [Homo sapiens Regulatory Features \(GRCh38.p2\)](#) database.



This will make the [Filters](#) and [Attributes](#) options appear in the left-hand column. You can do filters and attributes in any order, but we'll start by clicking on [Filters](#).

Scroll down to find [Regulatory Stable ID](#), then paste in the list of IDs.



That's all our filtering. Now go to [Attributes](#) on the left-hand column.

[Chromosome Name](#), [Start \(bp\)](#), [End \(bp\)](#) and [Feature Type](#) are already selected by default. Also select [Regulatory Stable ID](#) to get back our original input.

REGULATORY FEATURES

- Regulatory Stable ID
- Bound Start (bp)
- Bound End (bp)
- Chromosome Name
- Start (bp)
- End (bp)
- Feature Type

Now click on [Results](#) at the top.

BioMart is showing us multiple lines per feature. This is because BioMart gives us a new line if there could possibly be new data in the table. In this case, it's giving us a new line for each cell type, as this is data we could have selected. Choose [Unique results only](#) to only give one line per feature.

Export all results to Unique results only

Email notification to

View rows as Unique results only

Chromosome Name	Start (bp)	End (bp)	Feature Type	Regulatory Stable ID
11	560972	561971	Promoter	ENSR00000556855
11	560972	561971	Promoter	ENSR00000556855
11	560972	561971	Promoter	ENSR00000556855
11	560972	561971	Promoter	ENSR00000556855
11	560972	561971	Promoter	ENSR00000556855
11	560972	561971	Promoter	ENSR00000556855
11	560972	561971	Promoter	ENSR00000556855
11	560972	561971	Promoter	ENSR00000556855
11	560972	561971	Promoter	ENSR00000556855
11	560972	561971	Promoter	ENSR00000556855

Download the results

Choose to only see one line per feature

You can also download the results in various formats.

Browser Exercises

Gene regulation: Human *STX7*

(a) Find the Location tab ([Region in detail](#) page) for the *STX7* gene. Are there any predicted enhancers in this gene region? If so, where in the gene do they appear?

(b) Click [Configure this page](#) and on the [Regulatory features](#) menu in the left hand side. Turn on [Regulatory features](#) for [HUVEC](#), [HeLa-S3](#), and [HepG2](#) cell types. Are the predicted enhancers active in any of these cell lines?

(c) Use [Configure this page](#) to add supporting data indicating open chromatin for HeLa-S3 cells. Are there sites enriched for marks of open chromatin (DNase1) in HeLa cells at the 5' end of *STX7*?

(d) [Configure this page](#) once again to add histone modification supporting data for the same cell type as above (ie HeLa-S3). Which ones are present at the 5' end of *STX7*?

(e) Is there any data to support methylated CpG sites in this region (5' end) of *STX7* in Jurkat cells?

Regulatory features in human

(a) Go to the Location tab ([Region in detail](#) page) for human *APOE* and zoom out a little to see the flanking region. Is there a regulatory feature annotated at the 5' end of the gene? What kind of feature is it and what is its stable ID? Does it contain any transcription factor binding motifs?

(b) In which cell types is this feature active?

(c) Can you observe the relevant transcription factor binding to the transcription factor binding motifs you identified in part (a) in any cell types? What other transcription factors are also found at this location in this/these cell type(s)?

Using the ENCODE track hub

(a) If you have not already done so, add the ENCODE track-hub to view in GRCh37 and go to the configuration menu for the hub.

How many RNA Signal tracks are available for the nucleus of HeLa-S3 cells? What is the difference between the tracks?

(b) Turn on both repeats of the total signal on both strands. Go to the locus of the *SMC3* gene. Can you see any correlation between the signal and gene models? You may find it easier to move the tracks up and down (click and drag the left-hand bar) to view side-by-side.

(c) What does this suggest about the transcriptional activity of *SMC3* in HeLa-S3 cells? Is there a promoter at the 5' end of *SMC3* that is predicted to be active in HeLa-S3 cells?

BioMart Exercises

Regulatory features by region

An ~250kb deletion on chromosome 18 (2257243-2521438) has been identified in a patient with skeletal defects. This region is non-genic and you suspect it may contain an enhancer

(a) Export a list of predicted enhancers within this locus.

(b) You are particularly interested in enhancers that are active in HSMM (Human Skeletal Muscle Myoblasts) cells. Which of the enhancers in this region have evidence of activity in these cells?

Regulatory evidence

Find all peaks of H3K9ac in NH-A cells on chromosome band 7p21.1. (Note that data peaks, such as histone modifications, as classified as Regulatory evidence in BioMart.)

Variants in a regulatory feature

Export a list of variants found in the following promoters:
ENSR00000507583, ENSR00001505213, ENSR00000507555,
ENSR00000507578, ENSR00000507715

(Note that you will need to use Variation Mart for this query.)

Ensembl Exercise Answers

Browser

Gene regulation: Human *STX7*

(a) Search for **human *STX7*** from the home page. Click on **Location** in the search results.

Click on the **Reg. Feats** track name to jump to an article explaining the underlying data. Click and drag the **Reg. Feats** track next to the **Genes (Merged Ensembl/Havana)** track to better compare where the Regulatory features are in the gene.

Regulatory features from the Ensembl 'regulatory build' are based on indicators of open chromatin such as CTCF binding sites, DNase I hypersensitive sites, and Transcription Factor binding sites. The Regulatory features are turned on by default in the Region in detail view.

See the legend below the Region in detail view to find the predicted enhancers are coloured in yellow. There are five enhancers predicted in the region of *STX7*, two near the 3' end, two near the middle and one near the 5' end.

(b) Two enhancers appear in the HUVEC cell type only (out of the three cells chosen).

(c) **Configure this page** and click on **Open chromatin & TFBS**. Turn on both peaks and signal for **DNase 1** in **HeLa-S3** cells (the boxes in this configure this page window will turn blue. For more information on how to select and view the supporting data, click on **Show tutorial** in the pop up window). Close the menu.

There's a DNase 1 hypersensitive site in the 5' exon of *STX7*. Click on the coloured block to find out that the DNase1 enriched sites in HeLa-S3 cells come from the ENCODE project.

(d) **Configure this page** and click on **Histones & polymerases**. Change the **Filter by** menu from **All classes** to **Histone**. Select the all the histone modifications available for **HeLa cells** (some of them might be on by default). Save and close the menu.

H3K4me3, H3K9ac, H3K27ac H3K36me3 and HsK4me2 sites have been found in the 5' region of *STX7* in HeLa-S3 cells.

(e) Click on configure this page and choose the [DNA Methylation](#) menu. Turn on the track [Jurkat RRBS ENCODE](#). Save and close the menu.

Some CpG sites at the 5' end of *STX7* are not highly methylated (note the yellow bars) whilst others are (blue bars). Yellow, green, and blue bars represent unmethylated, intermediately methylated, and methylated regions, respectively. For more information on human DNA methylation tracks, see:

www.ensembl.org/info/docs/funcgen/index.html

Regulatory features in human

(a) Search for **human gene** *APOE* from the home page. Click on [Location](#) in the search results. Hold down shift and drag out a box in the middle display to zoom out. The gene is positive stranded so look for features at the left hand side. Click on the features to get their IDs.

There is a pink promoter flank at the 5' end of *APOE*. Click on it to get a pop-up with its ID: ENSR00000347288. It contains two black lines, which are transcription factor binding motifs. The pop-up reveals that they are binding sites for SP1.

(b) Click on the stable ID [ENSR00000347288](#). Click on the ID to go to the regulation tab.

ENSR00001636517 is active in 6/18 cell types studied:
A549, H1ESC, HMEC, HepG2, IMR90 and K562

(c) Click on [Details by cell type](#), then open the [Select cells](#) menu. Choose **ALL ON** to select all cell types, then close the menu. Open the [Select evidence](#) menu and choose **SP1** only, then close.

SP1 binding is only observed in H1ESC cells, at both of the binding motifs identified.

Open [Select evidence](#) again and choose all the transcription factors, then close. You may find it easier to see if you also go into

Select cells and turn the cell types [ALL OFF](#), then turn on [H1ESC](#) only.

TAF1 binds covering both SP1 binding sites. TAF7 and USF1 bind only to the left SP1 binding site.

Using the ENCODE track hub

(a) Go to grch37.ensembl.org and click on [Trackhubs](#) at the bottom left. The ENCODE track hub is near the top of the list. Add it by clicking on [Human \(GRCh37\)](#) in the right-hand column.

You will reach a page entitled [Your data](#), which lists the hub. Click on [Configure hub](#) to get to the configuration menu.

Click on [ENCODE RNA Signal](#) to open up the matrix of RNA Signal tracks. Find the intersection of [HeLa-S3](#) (Y-axis) and [Nucleus](#) (X-axis). Click on the box to see the track names.

There are 12 tracks (number in the bottom right of the box). There are PolyA, non-PolyA and total tracks, from the positive and negative strand, two repeats of each from Cold Spring Harbor.

(b) Click on the track names to turn them on, then save and close the menu. You will be in the Region in detail view.

Type [SMC3](#) into the [Gene](#) box, then select the gene when the name pops up.

The positive strand RNA Signals mostly line up with the exons of *SMC3*.

(c) These data suggest that *SMC3* is expressed in HeLa-S3 cells.

Click on [Configure this page](#) and scroll down to open the [Regulatory features](#) menu. Turn on the [Reg. Feats: HeLa-S3](#) track and close the menu.

There is a predicted promoter at the 5' end of *SMC3* and it is active (red) in HeLa-S3 cells.

BioMart

Regulatory features by region

(a) Start at ensembl.org/biomart/martview.
Choose the [ENSEMBL Regulation 80](#) database.
Choose the [Homo sapiens Regulatory Features \(GRCh38.p2\)](#) dataset.

Click on [Filters](#) in the left panel.
Select [Chromosome – 18](#).
Select [Base Pair](#) and input the coordinates.
Select [Feature Type – Enhancer](#).

Click on [Attributes](#) in the left panel.
Select [Regulatory stable ID](#) along with the default attributes.

Click the [Results](#) button on the toolbar.
Select [View All rows as HTML](#) or export all results to a file.

(b) Click on [Filters](#) again.
Select [Cell Type – HSMM](#).

Click on [Attributes](#).
Select [Cell Type](#) and [Has evidence](#).

Click the [Results](#) button.
Select [View All rows as HTML](#) or export all results to a file.
ENSR00001642219 and ENSR00001512731 are active (has evidence:true) in HSMM cells.

Regulatory evidence

Click [New](#).
Choose the [ENSEMBL Regulation 80](#) database.
Choose the [Homo sapiens Regulatory Features \(GRCh38.p2\)](#) dataset.

Click on [Filters](#) in the left panel.
Select [Chromosome – 7](#).
Select [Band](#) and choose [p21.1](#) as both [Start](#) and [End](#).

Select [Feature Type](#) – H3K9ac.
Select [Cell Type](#) – NH-A.

Click the [Results button](#) on the toolbar.
Select [View All rows as HTML](#) or export all results to a file.

Variants in a regulatory feature

Click [New](#).
Choose the [ENSEMBL Variation 80](#) database.
Choose the [Homo sapiens Short Variation \(SNPs and indels\) \(GRCh38\)](#) dataset.

Click on [Filters](#) in the left panel.
Expand the [REGULATORY REGION ASSOCIATED INFORMATION FILTERS](#) section by clicking on the + box.
Select [Filter by Regulatory Stable ID\(s\)](#) and enter the list of IDs.

Click on [Attributes](#) in the left panel.
Select the [Variation](#) attributes page.
Expand the [REGULATORY REGION ASSOCIATED INFORMATION FILTERS](#) by clicking on the + box.
Select [Regulatory Feature Stable ID](#).

Click the [Results button](#) on the toolbar.
Select [View All rows as HTML](#) or export all results to a file.