



Predicting Distal Regulatory Regions For A Gene:

(Updated 24 October 2013, Mike Pazin)

The [Regulatory Elements Database](#) is an ENCODE-funded tool described in a recent [publication](#) that can be used to make predictions about the linkage between regulatory regions and genes, based on the statistical association of DNase I Hypersensitive Sites (DHS) and gene expression across more than 100 samples consisting of over 70 diverse cell types.

From the [Regulatory Elements Database](#) site, click on the “By GENE” link (Arrow 1).

Regulatory Elements Database

This database provides a user interface to the results of the analysis presented in Sheffield et al. (2013), signal in 112 human samples. Questions? contact Nathan Sheffield. See also: Supplemental Files

There are 5 ways to explore the data:

- By CELLTYPE - Select cell-types to include or exclude.
- By CLUSTER - View promoter, CpG-island, and conserved element overlap for all clusters, and select:
 - Muscle-specific cluster: cluster 1520
 - Prostate and hepatocyte cluster: cluster 910
 - Prostate-only cluster: cluster 2483
 - Hematopoietic cluster: cluster 25
 - Pluripotency cluster: cluster 104
- By GENE - Search by gene of interest ← 1
 - MyoG
 - RBFOX1
 - LIN28A
 - HBG1
- By COORDINATE - Give chr, start, stop to find all regulatory elements in a region.
 - IRF2 regulator: chr4: 185240845-185240995
 - MyoD1 regulator: chr11: 17828545-17828695
 - Blood regulator: chr3: 128166420-128166570
- By FACTOR - find a specific TF of interest
 - CTCF
 - Myf

In the new screen, enter a gene name (such as IL10) in the box, and click the “Submit” button.

Gene: IL10
chr1: 206940946-206945839

Expression Profile
(log scale)

Graphical Connections

Distribution of linked clusters:
You may find it interesting if many of the connected DHS sites belong to the same cluster.

| Cluster Profile | Cluster ID | Number of DHS sites in this cluster |
|-----------------|------------|-------------------------------------|
| | 626 | 4 |
| | 675 | 4 |

RESOURCES

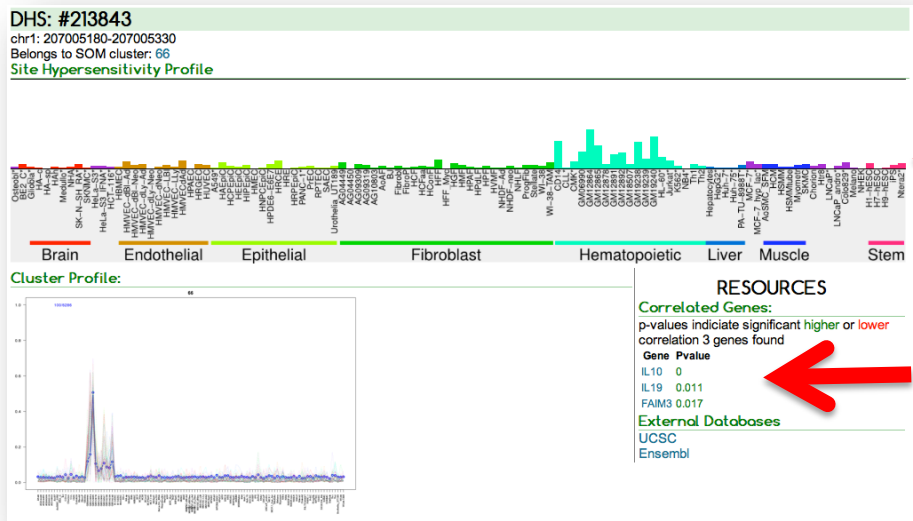
BED file downloads
Connects to 43 DHS sites
Download top 10 DHS sites. Download

Top 10 Correlated DHS Sites
p-values indicate significant higher or lower correlation

| Location | Cluster P |
|-------------------------------|-----------|
| chr1:206900205-206900355 509 | 0.000 |
| chr1:207005180-207005330 509 | 0.000 |
| chr1:206946500-206946550 15 | 0.000 |
| chr1:206947600-206947750 1368 | 0.000 |
| chr1:206853100-206853250 175 | 0.003 |
| chr1:206970325-206970475 875 | 0.008 |
| chr1:206903800-206903750 2055 | 0.009 |
| chr1:206946000-206946150 875 | 0.011 |
| chr1:206912940-206913090 2219 | 0.011 |
| chr1:206970165-206970315 2055 | 0.012 |
| chr1:206967920-206968070 2168 | 0.012 |

View top 100 connections. View

A list of the 10 DHS that are best correlated with the gene of interest is displayed at the right (near **Arrow 2**), and the expression profile of the gene across cell types is also shown (**Arrow 3**). For each DHS, the display indicates the genomic coordinate, p value (green for positive correlation with expression, red for negative correlation with expression), and the cluster number (hyperlink to DHS with a similar signal profile across cell types). The list can be downloaded as a BED file (which can be opened in excel); the complete list can also be displayed. To display details of any particular DHS, click on the location, which is also a hyperlink (eg., **Arrow 2**).



The new display indicates the genes that are predicted to be regulated by this DHS at the right (**arrow 4**), and the DHS signal profile across individual cell types (**arrow 5**).