

Using HaploReg and RegulomeDB to mine ENCODE data:

(Updated 16 November 2012, Mike Pazin)

[HaploReg](#) and [RegulomeDB](#) are ENCODE-funded tools described in recent publications that retrieve ENCODE annotations at SNPs of interest. Use of HaploReg is briefly described on this page, and RegulomeDB is described on the following page.

HaploReg:

Go to the [HaploReg site](#), and enter the name of the SNP of interest (**Arrow 1**).

HaploReg BROAD INSTITUTE

HaploReg is a tool for exploring annotations of the noncoding genome at variants on haplotype blocks, such as candidate regulatory SNPs at disease-associated loci. Using LD information from the 1000 Genomes Project, linked SNPs and small indels can be visualized along with their predicted chromatin state in nine cell types, conservation across mammals, and their effect on regulatory motifs. HaploReg is designed for researchers developing mechanistic hypotheses of the impact of non-coding variants on clinical phenotypes and normal variation.

[Build Query](#) | [Set Options](#) | [Documentation](#)

Use one of the three methods below to enter a set of variants. If an r^2 threshold is specified (see the Set Options tab), results for each variant will be shown in a separate table along with other variants in LD. If r^2 is set to NA, only queried variants will be shown, together in one table.

Query (refSNP ID(s), comma-delimited): **Arrow 1**
or, upload a text file (one refSNP ID per line): no file selected
or, select a GWAS:

Arrow 2

Query SNP: **rs3024505** and variants with $r^2 \geq 1$ **Arrow 3**

chr	pos (hg19)	LD variant	Ref	Alt	ASN freq	CEU freq	YRI freq	GERP cons	SiPhy cons	Promoter histone marks	Enhancer histone marks	DNAse	Proteins bound	eQTL tissues	Motifs changed	GENCODE genes	RefSeq genes	dbSNP func annot
1	206939904	rs3024505	G	A	0.02	0.19	0.03			GM12878	K562, NHLF	74 cell types	22 bound proteins			1kb 3' of IL10	1kb 3' of IL10	
1	206942413	rs3024495	C	T	0	0.19	0.02				GM12878					IL10	IL10	intronic

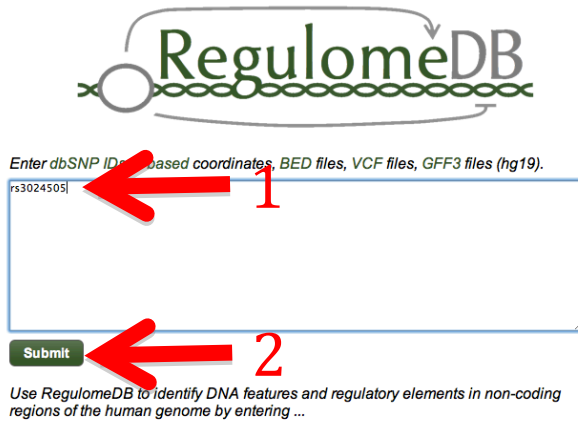
Click on the submit button (**Arrow 2**).

HaploReg retrieves the ENCODE annotation for the selected SNP, as well as other SNPs in LD (**arrow 3**).

Using the "Set Options" tab, the user can configure values such as the LD threshold and the population used from 1000 Genomes data used to calculate LD.

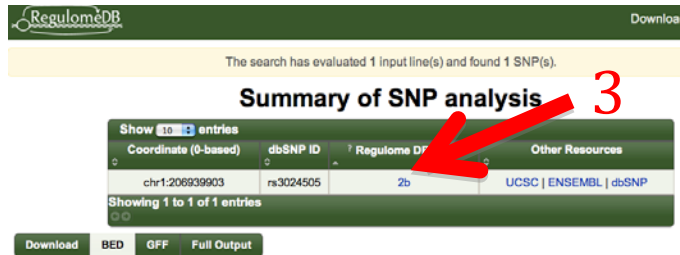
RegulomeDB:

Go to the RegulomeDB site and enter the name of the SNP of interest (**Arrow 1**).
Click on the submit button (**Arrow 2**).



The image shows the RegulomeDB search interface. At the top is the RegulomeDB logo. Below it is a search box with the text "Enter dbSNP ID or 0-based coordinates, BED files, VCF files, GFF3 files (hg19)". The search box contains the text "rs3024505" and is marked with a red arrow labeled "1". Below the search box is a "Submit" button, also marked with a red arrow labeled "2". Below the submit button is a small text box that says "Use RegulomeDB to identify DNA features and regulatory elements in non-coding regions of the human genome by entering ...".

RegulomeDB calculates a score for the regulatory potential of this region.

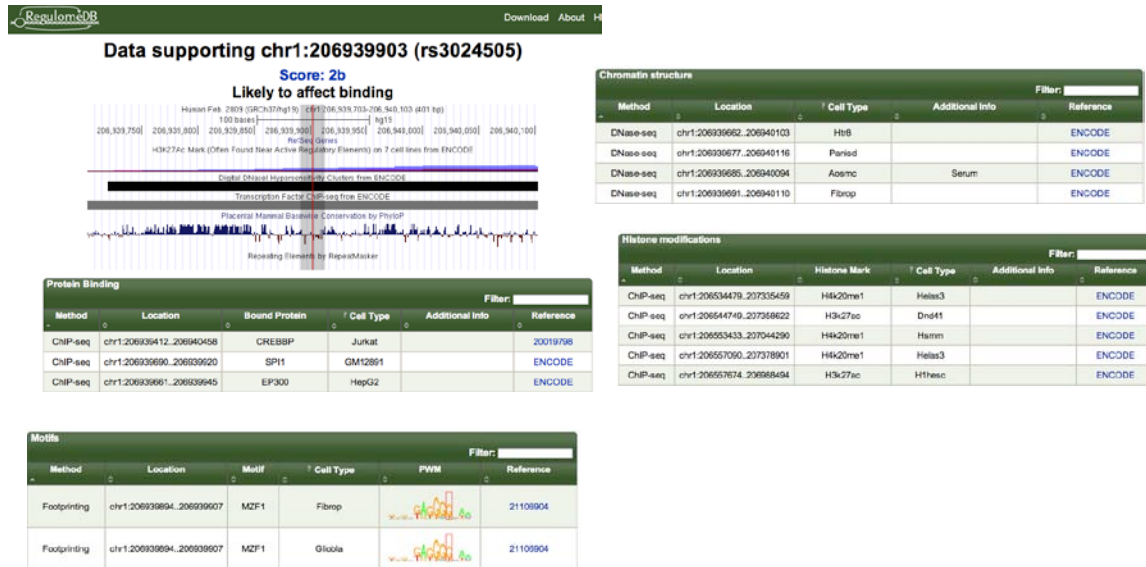


The image shows the RegulomeDB search results page. At the top is the RegulomeDB logo and a "Download" button. Below that is a message: "The search has evaluated 1 input line(s) and found 1 SNP(s)". The main heading is "Summary of SNP analysis" with a red arrow labeled "3" pointing to the "Regulome DB" column. Below the heading is a table with the following data:

Coordinate (0-based)	dbSNP ID	Regulome DB	Other Resources
chr1:206939903	rs3024505	2b	UCSC ENSEMBL dbSNP

Below the table is a "Showing 1 to 1 of 1 entries" message and a "Download" button. Below the download button are three buttons: "BED", "GFF", and "Full Output".

Clicking on the score retrieves the ENCODE annotation for the region (**arrow 3**).



The image shows the detailed results page for the SNP rs3024505. The main heading is "Data supporting chr1:206939903 (rs3024505)". Below the heading is a "Score: 2b" and "Likely to affect binding". The main content is a genomic track showing various annotations: Human Fek 2809 (GRCh37/hg19), hg19, RefSeq Genes, H3K27ac Mark (Often Found Near Active Regulatory Elements) on 7 cell lines from ENCODE, DNase-seq, Transcription Factor ChIP-seq from ENCODE, Phylo-P, and Repeating Elements by RepeatMasker. To the right of the track are two tables: "Chromatin structure" and "Histone modifications".

Chromatin structure

Method	Location	Cell Type	Additional Info	Reference
DNase-seq	chr1:206939662..206940103	H1h9		ENCODE
DNase-seq	chr1:206939677..206940116	Parise1		ENCODE
DNase-seq	chr1:206939685..206940094	Aoemc	Serum	ENCODE
DNase-seq	chr1:206939691..206940110	Fibrop		ENCODE



Histone modifications

Method	Location	Histone Mark	Cell Type	Additional Info	Reference
ChIP-seq	chr1:206534479..207335459	H4k20me1	HeLa3		ENCODE
ChIP-seq	chr1:206644749..207358622	H3k27ac	Dnd41		ENCODE
ChIP-seq	chr1:206553433..207044290	H4k20me1	Hamm		ENCODE
ChIP-seq	chr1:206557050..207378901	H4k20me1	HeLa3		ENCODE
ChIP-seq	chr1:206557674..206868494	H3k27ac	Hthesc		ENCODE

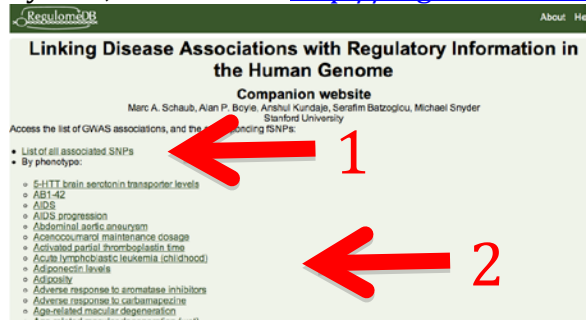
Protein Binding

Method	Location	Bound Protein	Cell Type	Additional Info	Reference
ChIP-seq	chr1:206939412..206940458	CREBBP	Jurkat		20071978
ChIP-seq	chr1:206939690..206939920	SP11	GM12891		ENCODE
ChIP-seq	chr1:206939661..206939945	EP300	HepG2		ENCODE

Motifs

Method	Location	Motif	Cell Type	PWM	Reference
Footprinting	chr1:206939894..206939907	MZF1	Fibrop		2110904
Footprinting	chr1:206939894..206939907	MZF1	Gluc04		2110904

RegulomeDB also has a database of predicted functional SNPs, by disease/trait and by SNP, available at: <http://regulome.stanford.edu/GWAS>



There is a list of over 4700 SNPs associated with human traits and disease (arrow 1), as well as a list of over 470 human traits and diseases (arrow 2).

Clicking on a trait/disease returns a list of SNPs that have been associated with that trait or disease:



Clicking on a SNP (red arrow) returns the evidence for the association:



As well as the annotation for the lead SNP, and other SNPs in LD that, based on functional annotation, are candidates for the functional variant:



One can follow the links to view the genomic annotation of these SNPs in the genome browser.