

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**).

The screenshot shows the HaploReg web interface. At the top, there is a header with the HaploReg logo and logos for the Broad Institute and MIT. Below the header is a brief description of the tool. The main interface has three tabs: "Build Query", "Set Options", and "Documentation". The "Build Query" tab is active. It contains instructions on how to enter a set of variants. There are three input methods: 1. A text input field for "Query (refSNP ID(s), comma-delimited):" with the value "rs3024505" entered. A red arrow labeled "1" points to this field. 2. A "Choose File" button for uploading a text file. 3. A dropdown menu for "or, select a GWAS:". Below these inputs is a "Submit" button, with a red arrow labeled "2" pointing to it. Below the "Submit" button, the query results are displayed: "Query SNP: rs3024505 and variants with $r^2 \geq 1$ ". A table of results is shown, with a red arrow labeled "3" pointing to the table. The table has columns for chromosome, position, LD, variant, reference/alternate alleles, population frequencies, conservation scores, histone marks, DNase, proteins, eQTL, motifs, Gencode genes, RefSeq genes, and dbSNP functional annotations.

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	1	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	1	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).

Enter dbSNP ID or 0-based coordinates: BED files, VCF files, GFF3 files (hg19).

rs3024505

Submit

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 search has evaluated 1 input line(s) and found 1 SNP(s).

Summary of SNP analysis

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

Showing 1 to 1 of 1 entries

Download BED GFF Full Output

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

Data supporting chr1:206939903 (rs3024505)

Score: 2b
Likely to affect binding

Human Feb 2009 (GRCh37/hg19) chr1:206,939,703-206,940,103 (401 bp)
100 bases hg19
206,939,750 206,939,800 206,939,850 206,939,900 206,939,950 206,940,000 206,940,050 206,940,100
RefSeq Genes
H3K27ac Mark (Often Found Near Active Regulatory Elements) on 7 cell lines from ENCODE
Detailed Chromatin States from ENCODE
Transcription Factor ChIP-seq from ENCODE
Phylo-P
Phylo-P: Phylogenetic Conservation by Phylo-P
Repeating Elements by RepeatMasker

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

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

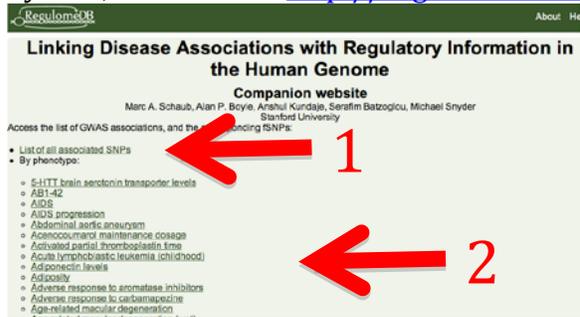
Histone modifications

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

Motifs

Method	Location	Motif	Cell Type	PWM	Reference
Footprinting	chr1:206939894..206939907	MZF1	Fibrop		2110904
Footprinting	chr1:206939894..206939907	MZF1	Gluc4		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.