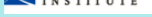


(Updated 17 May 2013, Mike Pazin)

HaploReg v2:

HaploReg v2



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, their sequence 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.

Update 2013.02.14: Version 2 now includes an expanded library of SNPs (based on dbSNP 137), motif instances (based on PWMs discovered from ENCODE experiments), enhancer annotations (adding 90 cell types from the Roadmap Epigenome Mapping Consortium), and eQTLs (from the GTex eQTL browser). In addition, LD calculations are provided based on the 1000 Genomes Phase 1 individuals, and r^2 and D' measurements are available down to an r^2 threshold of 0.2. Display improvements include improved cell metadata, gene metadata, and PWM display on the detail pages and the option for text output. Version 1 is available [here](#).

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 (comma-delimited list of rsIDs OR a single region as chrN:start-end):

or, upload a text file (one refSNP ID per line): Choose File no file selected

or, select a GWAS:

Submit

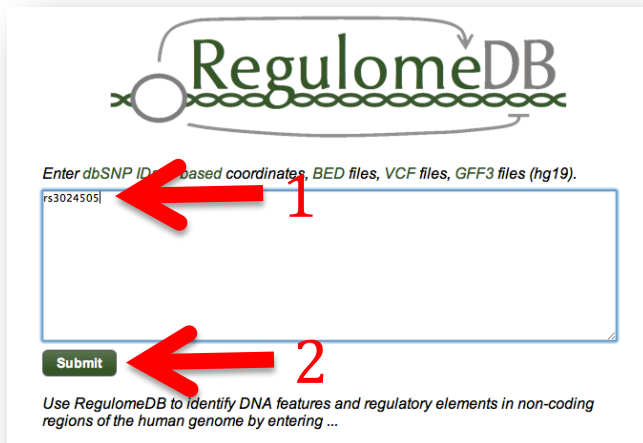
selected SNP, as well as other SNPs in LD (arrow 3).



Query SNP: rs4810485 and variants with $r^2 \geq 0.8$																			
chr	pos (hg19)	LD (r ²)	LD (D')	variant	Ref	Alt	AFR freq	AMR freq	ASN freq	EUR freq	SiPhy cons	Promoter histone marks	Enhancer histone marks	DNAse	Proteins bound	eQTL tissues	Motifs changed	GENCODE genes	dbSNP func annot
20	44730245	0.98	0.99	rs6032660	G	A	0.98	0.73	0.59	0.75							Mtf1, Zfx	12kb 5' of NCOA5	
20	44732089	0.97	0.99	rs2024568	T	C	0.97	0.73	0.58	0.75							BDP1, GCNF, NR2f2	13kb 5' of NCOA5	
20	44734310	0.98	0.99	rs6032662	C	T	0.98	0.73	0.59	0.75							Zfp410	13kb 5' of CD40	
20	44735263	0.95	0.99	rs6032663	T	G	0.98	0.72	0.58	0.74							RFK5	12kb 5' of CD40	
20	44735854	0.97	0.99	rs6065826	A	G	0.99	0.76	1.00	0.75			GM12878	HMVEC-Lly			HMG-I, Y, PU.1	11kb 5' of CD40	
20	44739419	0.98	0.99	rs6032664	A	T	0.98	0.73	0.59	0.75			GM12878				Spdef	7.5kb 5' of CD40	
20	44740196	0.95	0.99	rs6074022	C	T	0.97	0.73	0.58	0.74		HSMM	GM12878	7 cell types	6 bound proteins		CHD2, Nr1-2	6.7kb 5' of CD40	
20	44742064	0.98	0.99	rs1569723	C	A	0.98	0.73	0.59	0.75			HMEC	ProgBio			Irf	4.8kb 5' of CD40	
20	44746982	1	1	rs1883832	T	C	0.98	0.73	0.59	0.75		8 cell types	NHLF	LNCAp, Chorion, GM19239	13 bound proteins		4 altered motifs		5'-UTR
20	44747947	1	1	rs4810485	T	G	0.94	0.73	0.59	0.75		4 cell types	NHEK, H1	10 cell types	4 bound proteins		STAT	CD40	intronic
20	44749251	0.88	1	rs4239702	T	C	0.85	0.70	0.60	0.72		GM12878	Huvec	6 cell types			Myf, Sox, Zfp105	CD40	intronic

RegulomeDB:

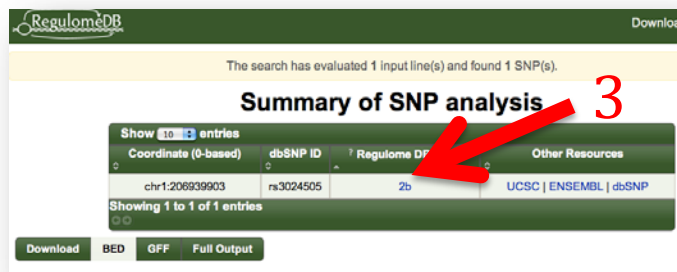
Go to the [RegulomeDB](http://regulomedb.org) site and enter the name of the SNP of interest (**Arrow 1**).



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

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

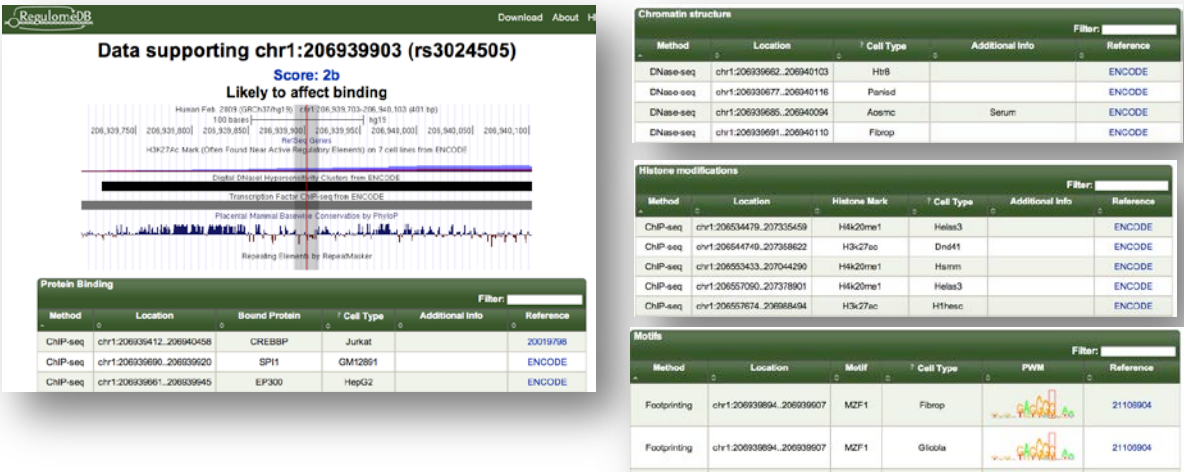
RegulomeDB calculates a score for the regulatory potential of this region.



The image shows the RegulomeDB search results page. At the top is the RegulomeDB logo and a "Download" link. Below it is a yellow banner with the text "The search has evaluated 1 input line(s) and found 1 SNP(s)". Below the banner is a section titled "Summary of SNP analysis". A red arrow labeled "3" points to the "Regulome DB" column of the table. The table has four columns: "Coordinate (0-based)", "dbSNP ID", "Regulome DB", and "Other Resources". The table contains one row with the following data: "chr1:206939903", "rs3024505", "2b", and "UCSC | ENSEMBL | dbSNP". Below the table is a section titled "Showing 1 to 1 of 1 entries". At the bottom are three buttons: "Download", "BED", and "GFF".

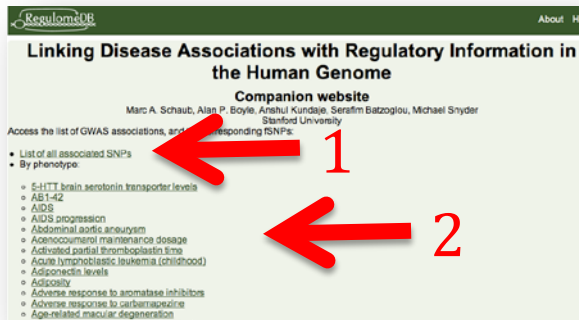
Coordinate (0-based)	dbSNP ID	Regulome DB	Other Resources
chr1:206939903	rs3024505	2b	UCSC ENSEMBL dbSNP

Clicking on the score (arrow 3) retrieves the ENCODE (and other) annotation for the region, including transcription factor binding, chromatin structure (DNase, FAIRE, and histone modifications), transcription factor motifs and eQTL.



RegulomeDB Disease Association Database, a database of predicted functional SNPs, organized by disease/trait and by SNP, is 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:

Lead SNP
rs3024505
Position: chr1:206,939,904 ([Open in UCSC Genome Browser](#))
Distance to nearest TSS: 18,466 bp
GENCODE v7 location: Intergenic region
RegulomeDB Score: 2b - ChIP-seq peak + any motif + matched DNase Footprint + DNase-seq peak ([Open in RegulomeDB](#))

Linkage disequilibrium region
Linkage disequilibrium threshold:
- In all HapMap 2 populations: $r^2 \geq 0.8$ $r^2 \geq 0.9$ $r^2 = 1.0$
- In the HapMap 2 CEU population $r^2 \geq 0.8$ $r^2 \geq 0.9$ $r^2 = 1.0$
SNPs in the linkage disequilibrium region sorted by decreasing amount of evidence supporting a functional role for the SNP:

rs3024493
Position: chr1:206,943,968 ([Open in UCSC Genome Browser](#))
Distance to lead SNP: 4,064 bp
Distance to nearest TSS: 22,530 bp
GENCODE v7 location: Intron
RegulomeDB Score: 2b - ChIP-seq peak + any motif + matched DNase Footprint + DNase-seq peak ([Open in RegulomeDB](#))
Linkage disequilibrium with Lead SNP (HapMap 2): CEU: $D'=1.0$, $r^2=1.0$ / CHB: $D'=1.0$, $r^2=1.0$ / JPT: $D'=1.0$, $r^2=1.0$ / YRI: $D'=1.0$, $r^2=1.0$

rs3024495
Position: chr1:206,942,413 ([Open in UCSC Genome Browser](#))
Distance to lead SNP: 2,509 bp
Distance to nearest TSS: 20,975 bp
GENCODE v7 location: Intron
RegulomeDB Score: 5a - ChIP-seq peak ([Open in RegulomeDB](#))
Linkage disequilibrium with Lead SNP (HapMap 2): CEU: $D'=1.0$, $r^2=1.0$ / CHB: $D'=1.0$, $r^2=1.0$ / JPT: $D'=1.0$, $r^2=1.0$ / YRI: $D'=1.0$, $r^2=1.0$

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