Variant Annotation Using RegulomeDB and HaploReg

Jill E. Moore
Weng Lab
University of Massachusetts Medical School
June 29, 2015

Motivation

- The majority of variants report by GWAS are in noncoding regions of the genome

- The variant reported in the GWAS (lead/tagged variant) may not be causal but is in high linkage disequilibrium with the casual variant

- Using data from ENCODE, we can annotate noncoding regions of the genome and predict the function of disease associated noncoding variants

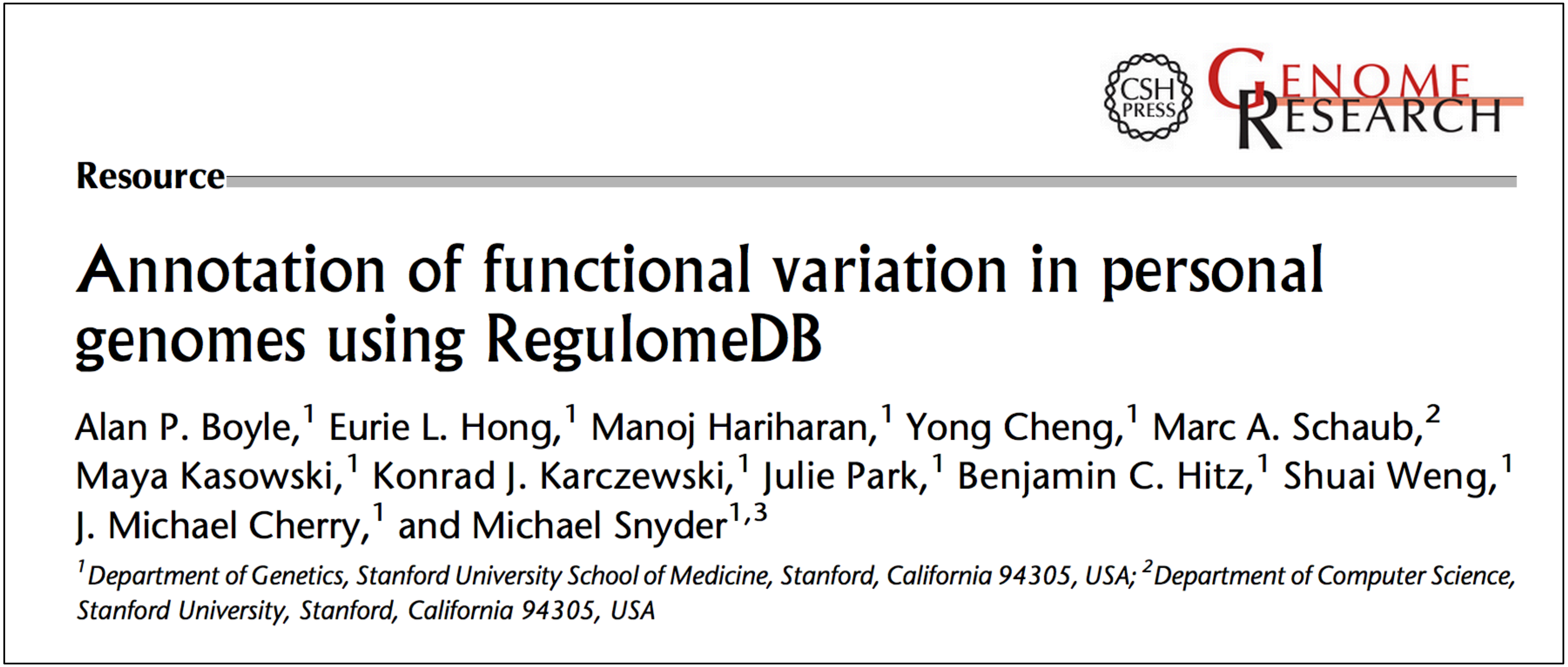
Variant Annotation Tools

RegulomeDB

http://www.regulomedb.org/

HaploReg

http://www.broadinstitute.org/mammals/haploreg/haploreg.php



CSH PRESS

Genome Research

Resource

Annotation of functional variation in personal genomes using RegulomeDB

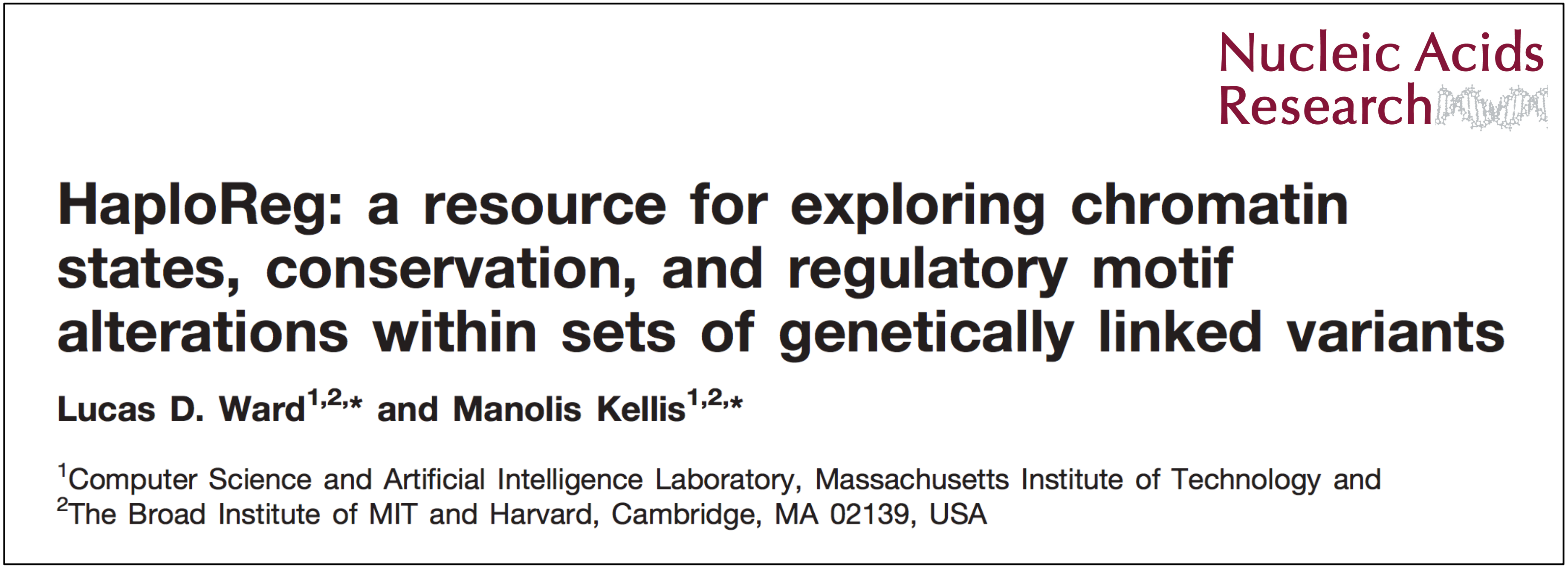
Alan P. Boyle,[1] Eurie L. Hong,[1] Manoj Hariharan,[1] Yong Cheng,[1] Marc A. Schaub,[2] Maya Kasowski,[1] Konrad J. Karczewski,[1] Julie Park,[1] Benjamin C. Hitz,[1] Shuai Weng,[1] J. Michael Cherry,[1] and Michael Snyder[1,3]

[1] Department of Genetics, Stanford University School of Medicine, Stanford, California 94305, USA; [2] Department of Computer Science, Stanford University, Stanford, California 94305, USA

Table 2. RegulomeDB variant classification scheme

Category scheme	
Category	Description
	Likely to affect binding and linked to expression of a gene target
1a	eQTL + TF binding + matched TF motif + matched DNase footprint + DNase peak
1b	eQTL + TF binding + any motif + DNase footprint + DNase peak
1c	eQTL + TF binding + matched TF motif + DNase peak
1d	eQTL + TF binding + any motif + DNase peak
1e	eQTL + TF binding + matched TF motif
1f	eQTL + TF binding/DNase peak
	Likely to affect binding
2a	TF binding + matched TF motif + matched DNase footprint + DNase peak
2b	TF binding + any motif + DNase footprint + DNase peak
2c	TF binding + matched TF motif + DNase peak
	Less likely to affect binding
3a	TF binding + any motif + DNase peak
3b	TF binding + matched TF motif
	Minimal binding evidence
4	TF binding + DNase peak
5	TF binding or DNase peak
6	Motif hit

Lower scores indicate increasing evidence for a variant to be located in a functional region. Category 1 variants have equivalents in other categories with the additional requirement of eQTL information.



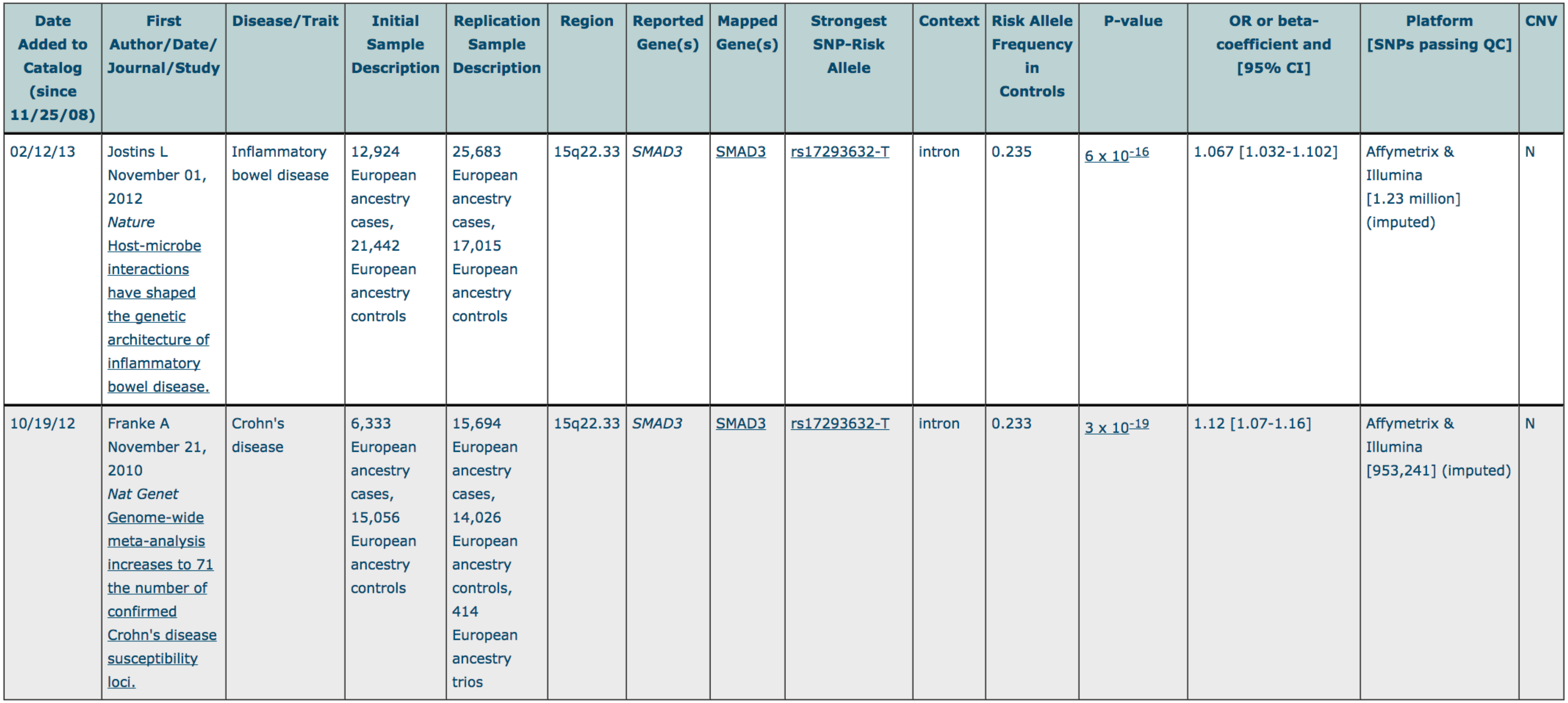
Nucleic Acids Research

HaploReg: a resource for exploring chromatin states, conservation, and regulatory motif alterations within sets of genetically linked variants

Lucas D. Ward[1,2,*] and Manolis Kellis[1,2,*]

[1]Computer Science and Artificial Intelligence Laboratory, Massachusetts Institute of Technology and
[2]The Broad Institute of MIT and Harvard, Cambridge, MA 02139, USA

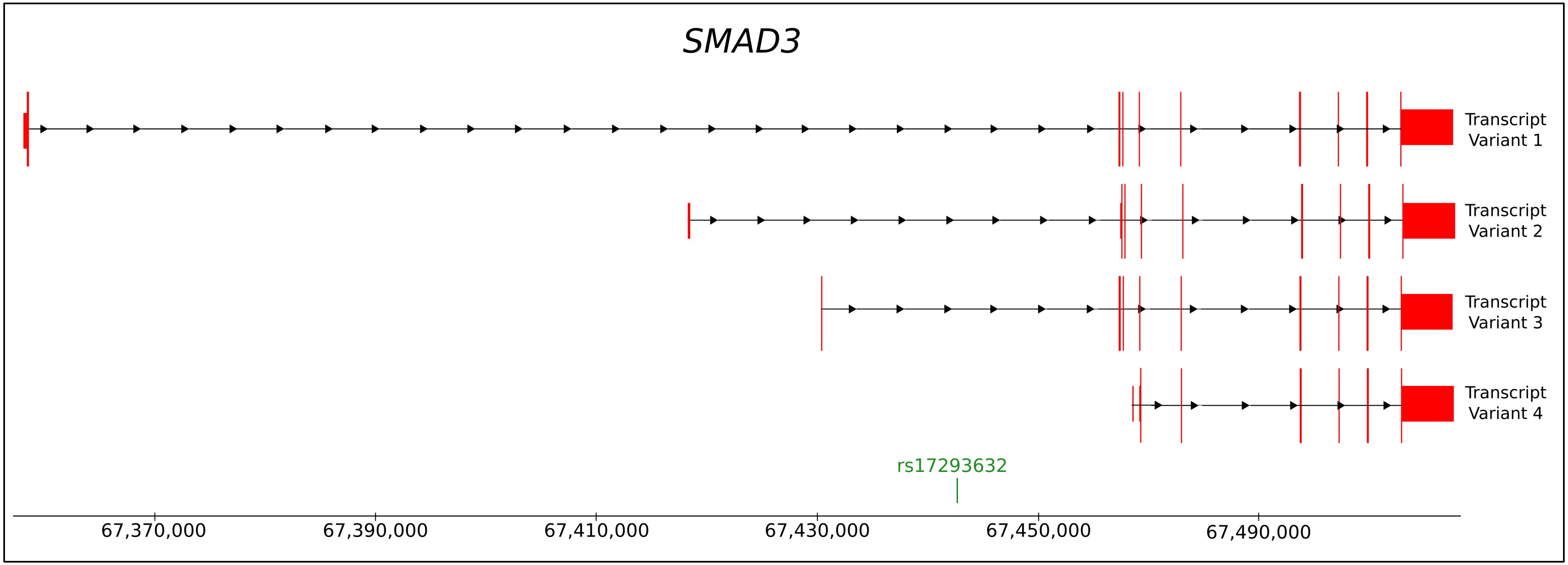
rs17293632 is Associated with IBD and Crohn's Disease



Date Added to Catalog (since 11/25/08)	First Author/Date/ Journal/Study	Disease/Trait	Initial Sample Description	Replication Sample Description	Region	Reported Gene(s)	Mapped Gene(s)	Strongest SNP-Risk Allele	Context	Risk Allele Frequency in Controls	P-value	OR or beta-coefficient and [95% CI]	Platform [SNPs passing QC]	CNV
02/12/13	Jostins L November 01, 2012 *Nature* Host-microbe interactions have shaped the genetic architecture of inflammatory bowel disease.	Inflammatory bowel disease	12,924 European ancestry cases, 21,442 European ancestry controls	25,683 European ancestry cases, 17,015 European ancestry controls	15q22.33	*SMAD3*	SMAD3	rs17293632-T	intron	0.235	6 x 10^{-16}	1.067 [1.032-1.102]	Affymetrix & Illumina [1.23 million] (imputed)	N
10/19/12	Franke A November 21, 2010 *Nat Genet* Genome-wide meta-analysis increases to 71 the number of confirmed Crohn's disease susceptibility loci.	Crohn's disease	6,333 European ancestry cases, 15,056 European ancestry controls	15,694 European ancestry cases, 14,026 European ancestry controls, 414 European ancestry trios	15q22.33	*SMAD3*	SMAD3	rs17293632-T	intron	0.233	3 x 10^{-19}	1.12 [1.07-1.16]	Affymetrix & Illumina [953,241] (imputed)	N

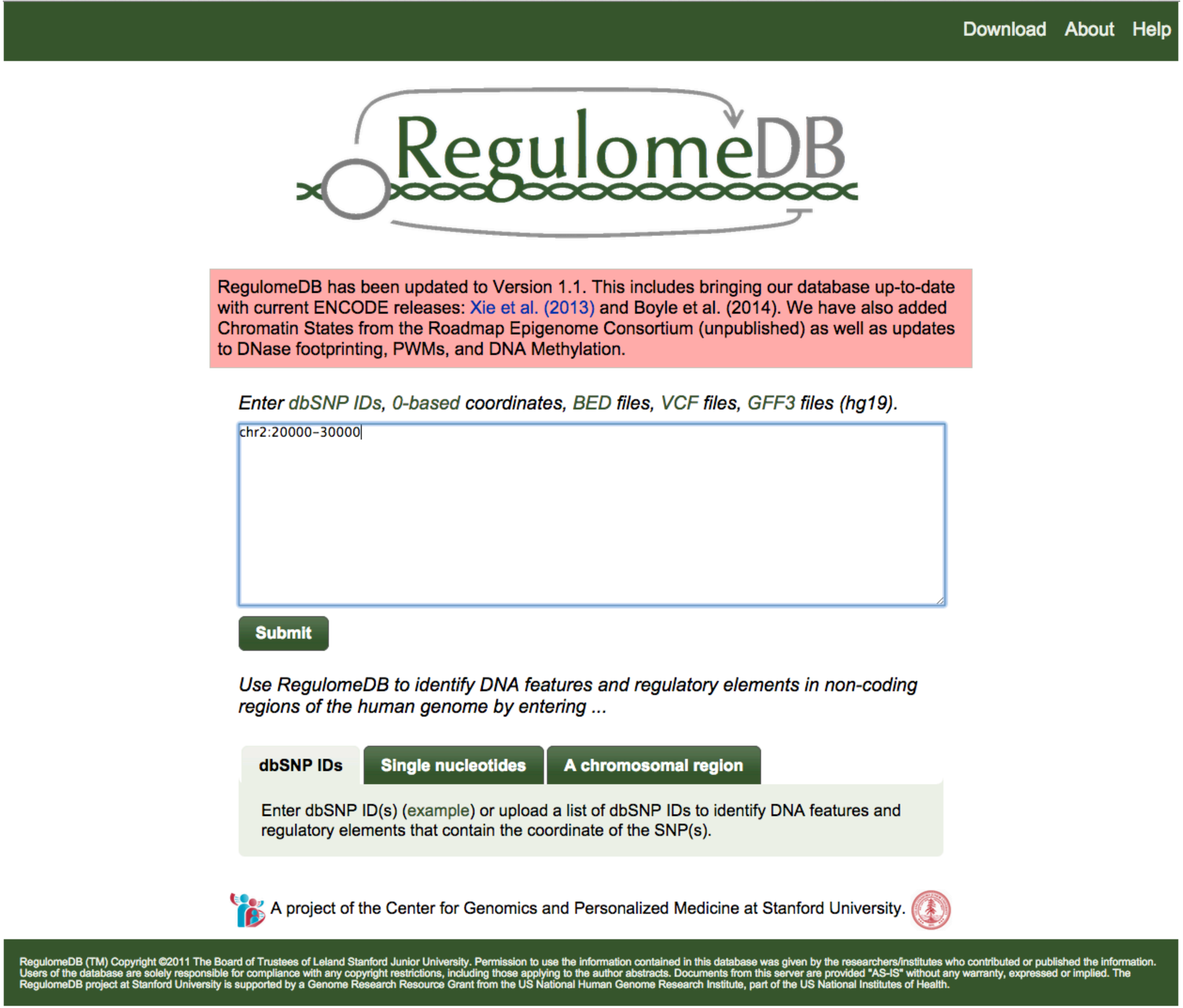
NHGRI GWAS Database https://www.genome.gov/26525384

rs17293632 is Associated with IBD and Crohn’s Disease



- rs17293632 is upstream of *SMAD3* transcript variant 4 and in the introns of transcript variants 1, 2 and 3.

Backup Slides



RegulomeDB has been updated to Version 1.1. This includes bringing our database up-to-date with current ENCODE releases: Xie et al. (2013) and Boyle et al. (2014). We have also added Chromatin States from the Roadmap Epigenome Consortium (unpublished) as well as updates to DNase footprinting, PWMs, and DNA Methylation.

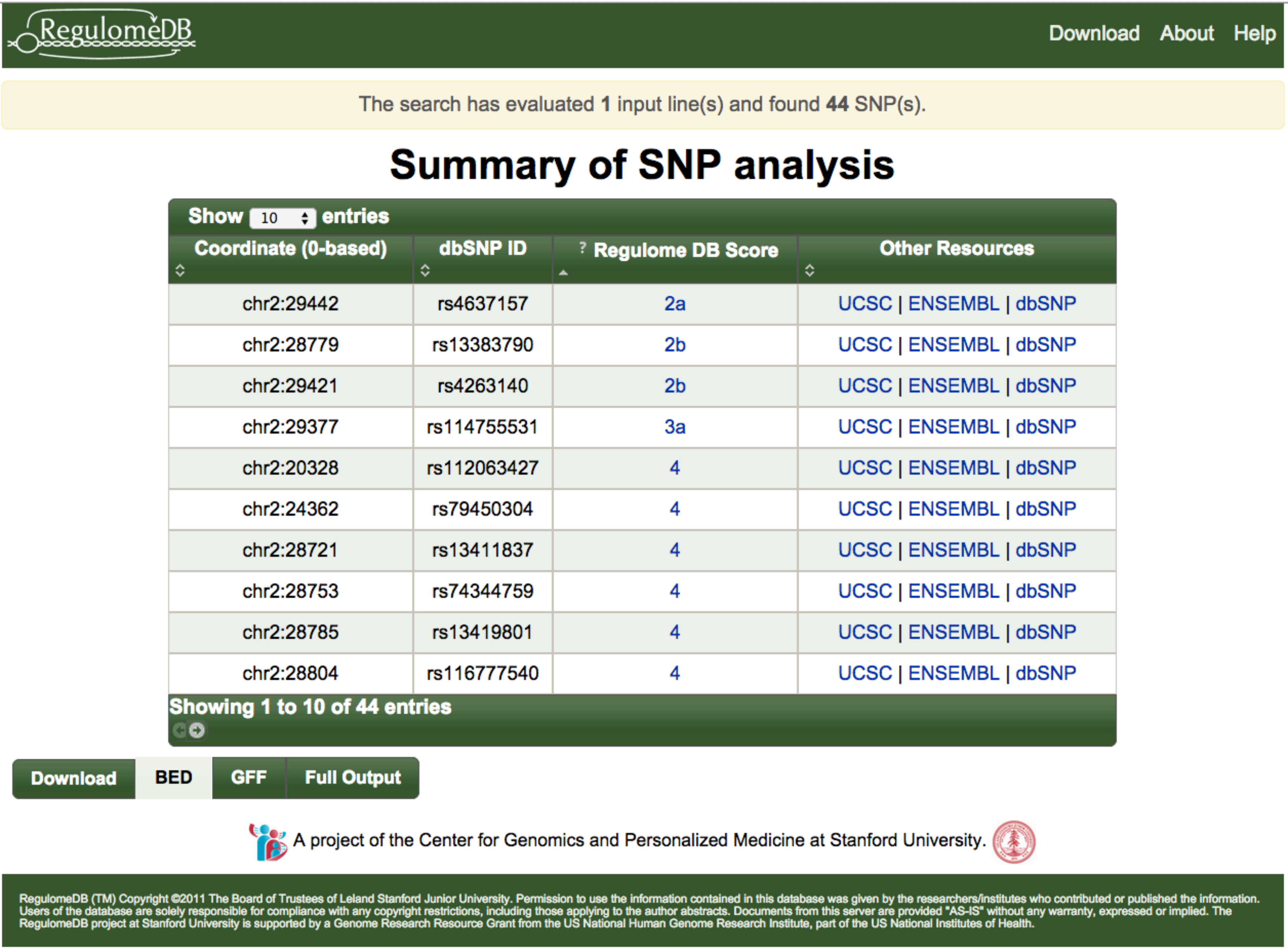
Enter dbSNP IDs, 0-based coordinates, BED files, VCF files, GFF3 files (hg19).

```
chr2:20000-30000
```

Use RegulomeDB to identify DNA features and regulatory elements in non-coding regions of the human genome by entering ...

Enter dbSNP ID(s) (example) or upload a list of dbSNP IDs to identify DNA features and regulatory elements that contain the coordinate of the SNP(s).

A project of the Center for Genomics and Personalized Medicine at Stanford University.



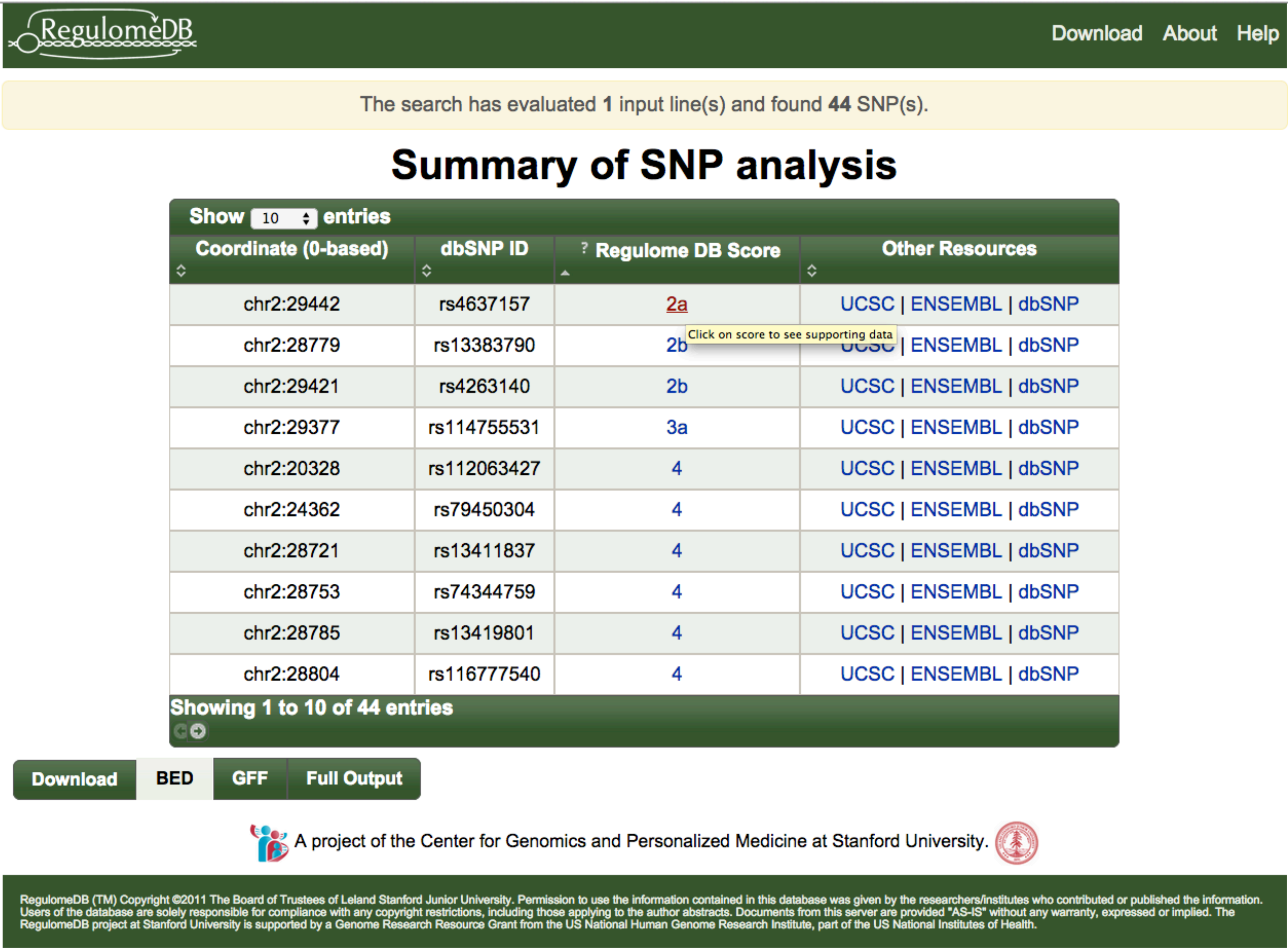
The search has evaluated **1** input line(s) and found **44** SNP(s).

Summary of SNP analysis

Coordinate (0-based)	dbSNP ID	? Regulome DB Score	Other Resources
chr2:29442	rs4637157	2a	UCSC | ENSEMBL | dbSNP
chr2:28779	rs13383790	2b	UCSC | ENSEMBL | dbSNP
chr2:29421	rs4263140	2b	UCSC | ENSEMBL | dbSNP
chr2:29377	rs114755531	3a	UCSC | ENSEMBL | dbSNP
chr2:20328	rs112063427	4	UCSC | ENSEMBL | dbSNP
chr2:24362	rs79450304	4	UCSC | ENSEMBL | dbSNP
chr2:28721	rs13411837	4	UCSC | ENSEMBL | dbSNP
chr2:28753	rs74344759	4	UCSC | ENSEMBL | dbSNP
chr2:28785	rs13419801	4	UCSC | ENSEMBL | dbSNP
chr2:28804	rs116777540	4	UCSC | ENSEMBL | dbSNP

Showing 1 to 10 of 44 entries

A project of the Center for Genomics and Personalized Medicine at Stanford University.



The search has evaluated **1** input line(s) and found **44** SNP(s).

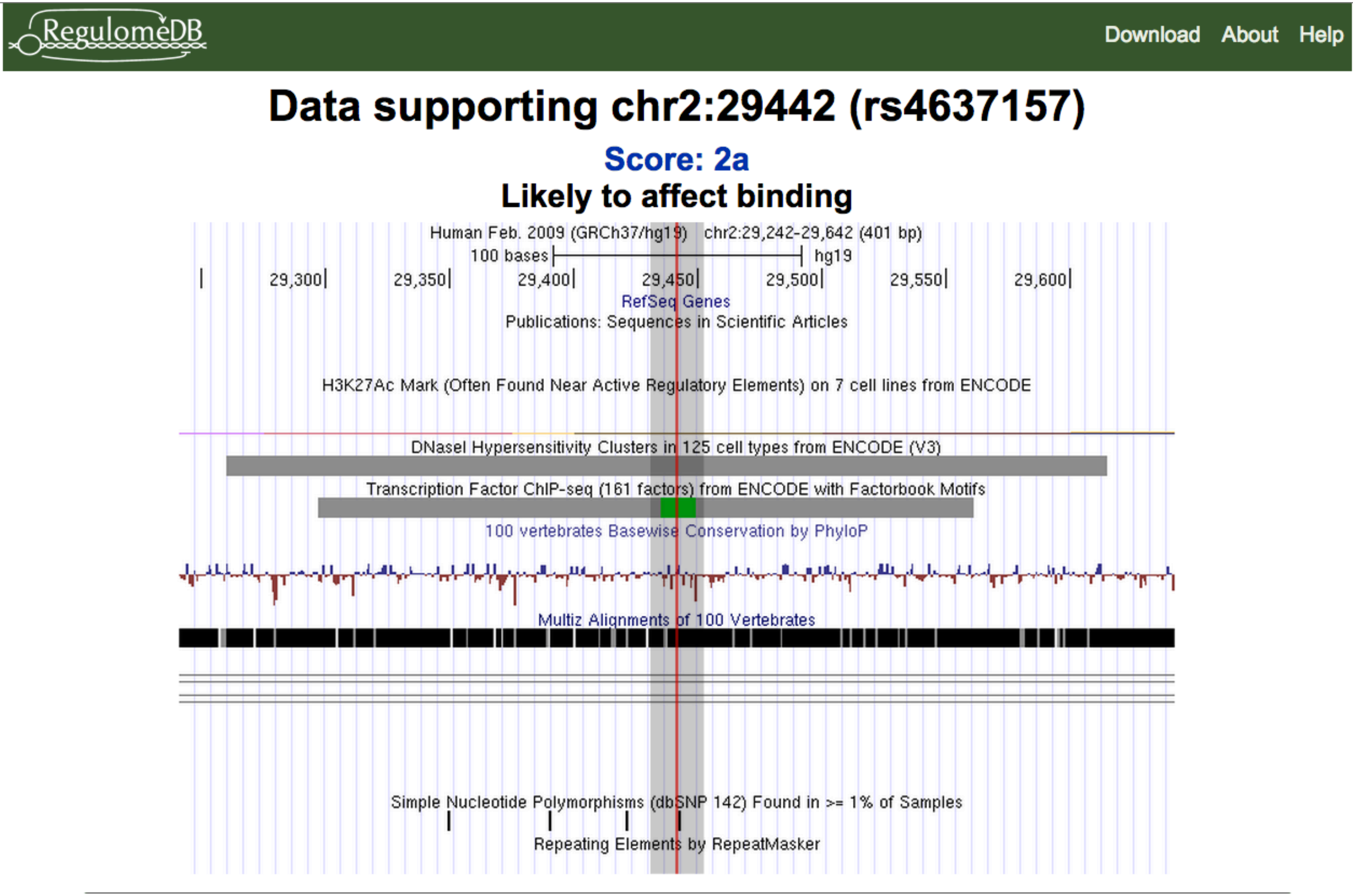
Summary of SNP analysis

Coordinate (0-based)	dbSNP ID	? Regulome DB Score	Other Resources
chr2:29442	rs4637157	2a	UCSC | ENSEMBL | dbSNP
chr2:28779	rs13383790	2b	UCSC | ENSEMBL | dbSNP
chr2:29421	rs4263140	2b	UCSC | ENSEMBL | dbSNP
chr2:29377	rs114755531	3a	UCSC | ENSEMBL | dbSNP
chr2:20328	rs112063427	4	UCSC | ENSEMBL | dbSNP
chr2:24362	rs79450304	4	UCSC | ENSEMBL | dbSNP
chr2:28721	rs13411837	4	UCSC | ENSEMBL | dbSNP
chr2:28753	rs74344759	4	UCSC | ENSEMBL | dbSNP
chr2:28785	rs13419801	4	UCSC | ENSEMBL | dbSNP
chr2:28804	rs116777540	4	UCSC | ENSEMBL | dbSNP

Click on score to see supporting data

Showing 1 to 10 of 44 entries

A project of the Center for Genomics and Personalized Medicine at Stanford University.

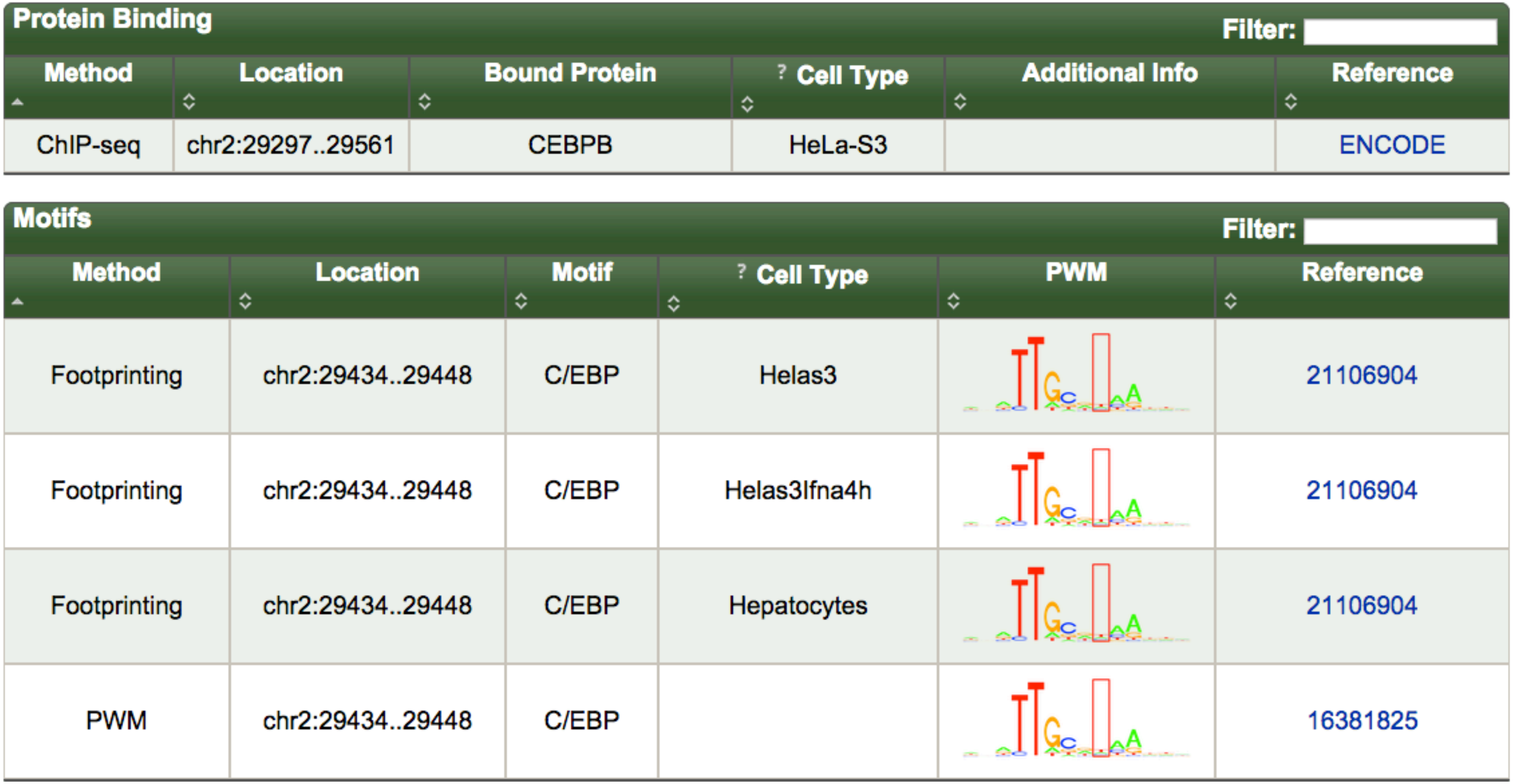

RegulomeDB
Download
About
Help

Data supporting chr2:29442 (rs4637157)

Score: 2a

Likely to affect binding

Human Feb. 2009 (GRCh37/hg19) chr2:29,242-29,642 (401 bp)
100 bases
hg19
29,300
29,350
29,400
29,450
29,500
29,550
29,600
RefSeq Genes
Publications: Sequences in Scientific Articles
H3K27Ac Mark (Often Found Near Active Regulatory Elements) on 7 cell lines from ENCODE
DNaseI Hypersensitivity Clusters in 125 cell types from ENCODE (V3)
Transcription Factor ChIP-seq (161 factors) from ENCODE with Factorbook Motifs
100 vertebrates Basewise Conservation by PhyloP
Multiz Alignments of 100 Vertebrates
Simple Nucleotide Polymorphisms (dbSNP 142) Found in >= 1% of Samples
Repeating Elements by RepeatMasker



Protein Binding

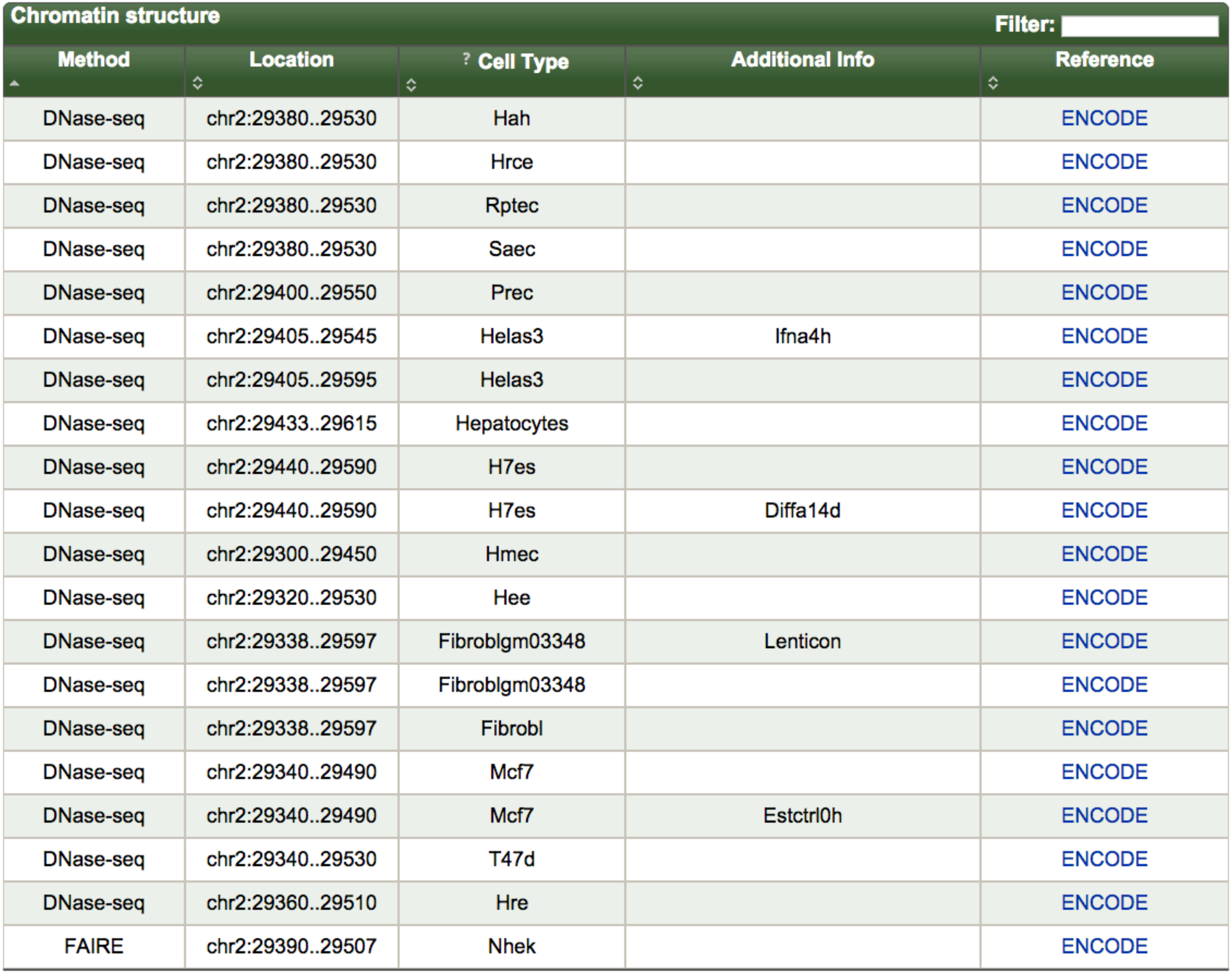
Filter:

Method	Location	Bound Protein	? Cell Type	Additional Info	Reference
ChIP-seq	chr2:29297..29561	CEBPB	HeLa-S3		ENCODE

Motifs

Filter:

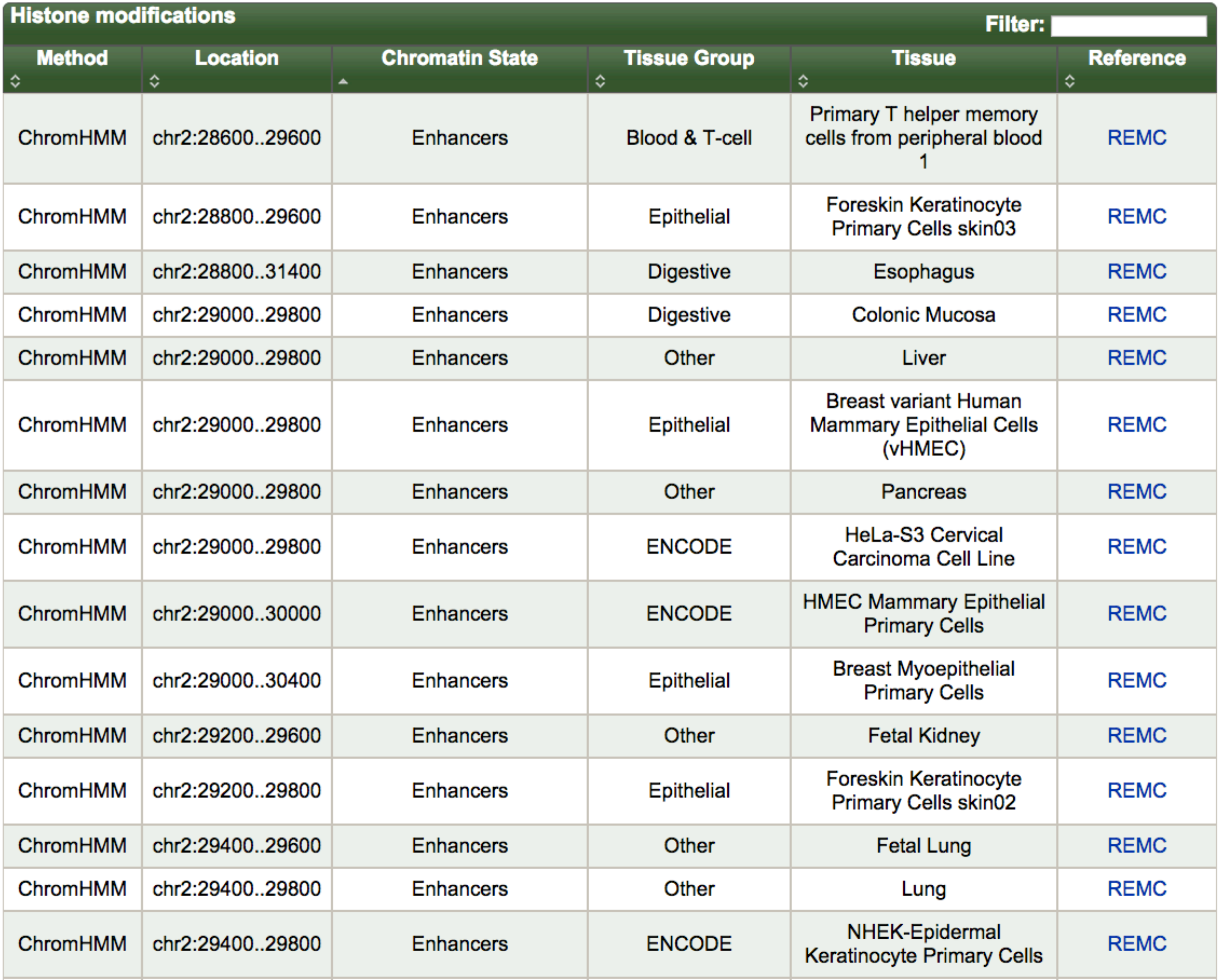
Method	Location	Motif	? Cell Type	PWM	Reference
Footprinting	chr2:29434..29448	C/EBP	Helas3		21106904
Footprinting	chr2:29434..29448	C/EBP	Helas3Ifna4h		21106904
Footprinting	chr2:29434..29448	C/EBP	Hepatocytes		21106904
PWM	chr2:29434..29448	C/EBP			16381825



Chromatin structure

Filter:

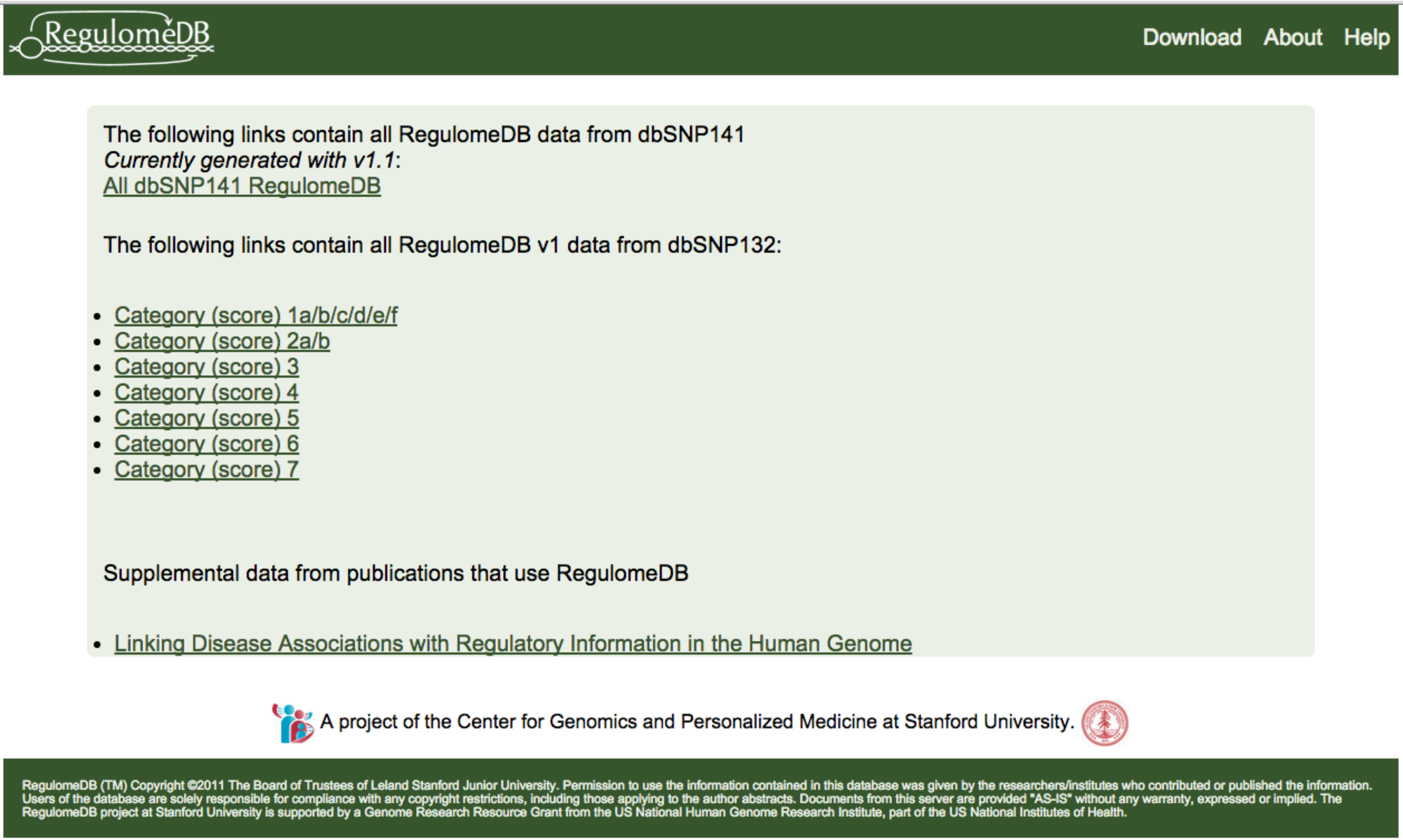
Method	Location	Cell Type	Additional Info	Reference
DNase-seq	chr2:29380..29530	Hah		ENCODE
DNase-seq	chr2:29380..29530	Hrce		ENCODE
DNase-seq	chr2:29380..29530	Rptec		ENCODE
DNase-seq	chr2:29380..29530	Saec		ENCODE
DNase-seq	chr2:29400..29550	Prec		ENCODE
DNase-seq	chr2:29405..29545	Helas3	Ifna4h	ENCODE
DNase-seq	chr2:29405..29595	Helas3		ENCODE
DNase-seq	chr2:29433..29615	Hepatocytes		ENCODE
DNase-seq	chr2:29440..29590	H7es		ENCODE
DNase-seq	chr2:29440..29590	H7es	Diffa14d	ENCODE
DNase-seq	chr2:29300..29450	Hmec		ENCODE
DNase-seq	chr2:29320..29530	Hee		ENCODE
DNase-seq	chr2:29338..29597	Fibroblgm03348	Lenticon	ENCODE
DNase-seq	chr2:29338..29597	Fibroblgm03348		ENCODE
DNase-seq	chr2:29338..29597	Fibrobl		ENCODE
DNase-seq	chr2:29340..29490	Mcf7		ENCODE
DNase-seq	chr2:29340..29490	Mcf7	Estctrl0h	ENCODE
DNase-seq	chr2:29340..29530	T47d		ENCODE
DNase-seq	chr2:29360..29510	Hre		ENCODE
FAIRE	chr2:29390..29507	Nhek		ENCODE



Histone modifications

Filter:

Method	Location	Chromatin State	Tissue Group	Tissue	Reference
ChromHMM	chr2:28600..29600	Enhancers	Blood & T-cell	Primary T helper memory cells from peripheral blood 1	REMC
ChromHMM	chr2:28800..29600	Enhancers	Epithelial	Foreskin Keratinocyte Primary Cells skin03	REMC
ChromHMM	chr2:28800..31400	Enhancers	Digestive	Esophagus	REMC
ChromHMM	chr2:29000..29800	Enhancers	Digestive	Colonic Mucosa	REMC
ChromHMM	chr2:29000..29800	Enhancers	Other	Liver	REMC
ChromHMM	chr2:29000..29800	Enhancers	Epithelial	Breast variant Human Mammary Epithelial Cells (vHMEC)	REMC
ChromHMM	chr2:29000..29800	Enhancers	Other	Pancreas	REMC
ChromHMM	chr2:29000..29800	Enhancers	ENCODE	HeLa-S3 Cervical Carcinoma Cell Line	REMC
ChromHMM	chr2:29000..30000	Enhancers	ENCODE	HMEC Mammary Epithelial Primary Cells	REMC
ChromHMM	chr2:29000..30400	Enhancers	Epithelial	Breast Myoepithelial Primary Cells	REMC
ChromHMM	chr2:29200..29600	Enhancers	Other	Fetal Kidney	REMC
ChromHMM	chr2:29200..29800	Enhancers	Epithelial	Foreskin Keratinocyte Primary Cells skin02	REMC
ChromHMM	chr2:29400..29600	Enhancers	Other	Fetal Lung	REMC
ChromHMM	chr2:29400..29800	Enhancers	Other	Lung	REMC
ChromHMM	chr2:29400..29800	Enhancers	ENCODE	NHEK-Epidermal Keratinocyte Primary Cells	REMC



The following links contain all RegulomeDB data from dbSNP141
Currently generated with v1.1:
All dbSNP141 RegulomeDB

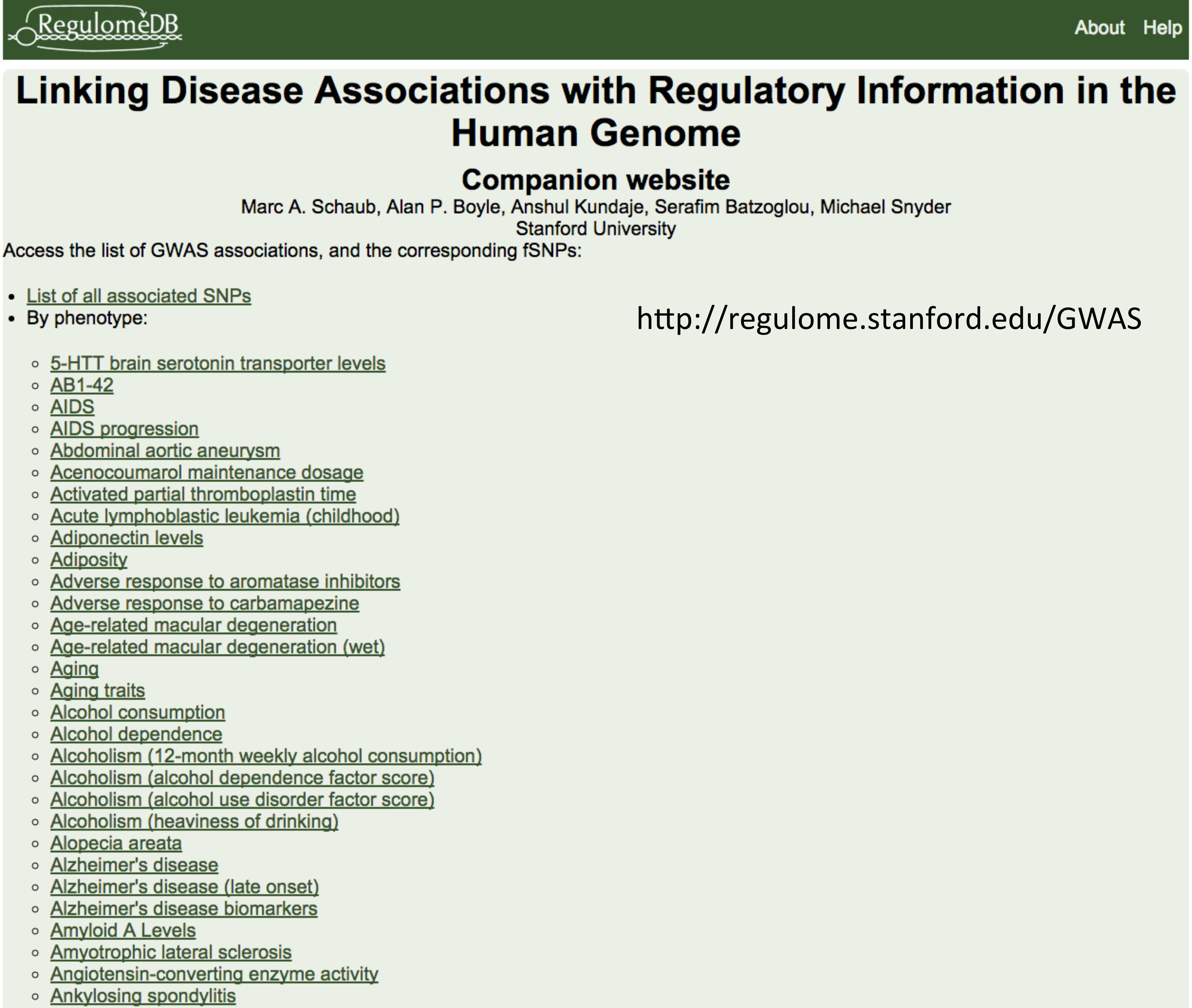
The following links contain all RegulomeDB v1 data from dbSNP132:

- Category (score) 1a/b/c/d/e/f
- Category (score) 2a/b
- Category (score) 3
- Category (score) 4
- Category (score) 5
- Category (score) 6
- Category (score) 7

Supplemental data from publications that use RegulomeDB

- Linking Disease Associations with Regulatory Information in the Human Genome

A project of the Center for Genomics and Personalized Medicine at Stanford University.



Linking Disease Associations with Regulatory Information in the Human Genome

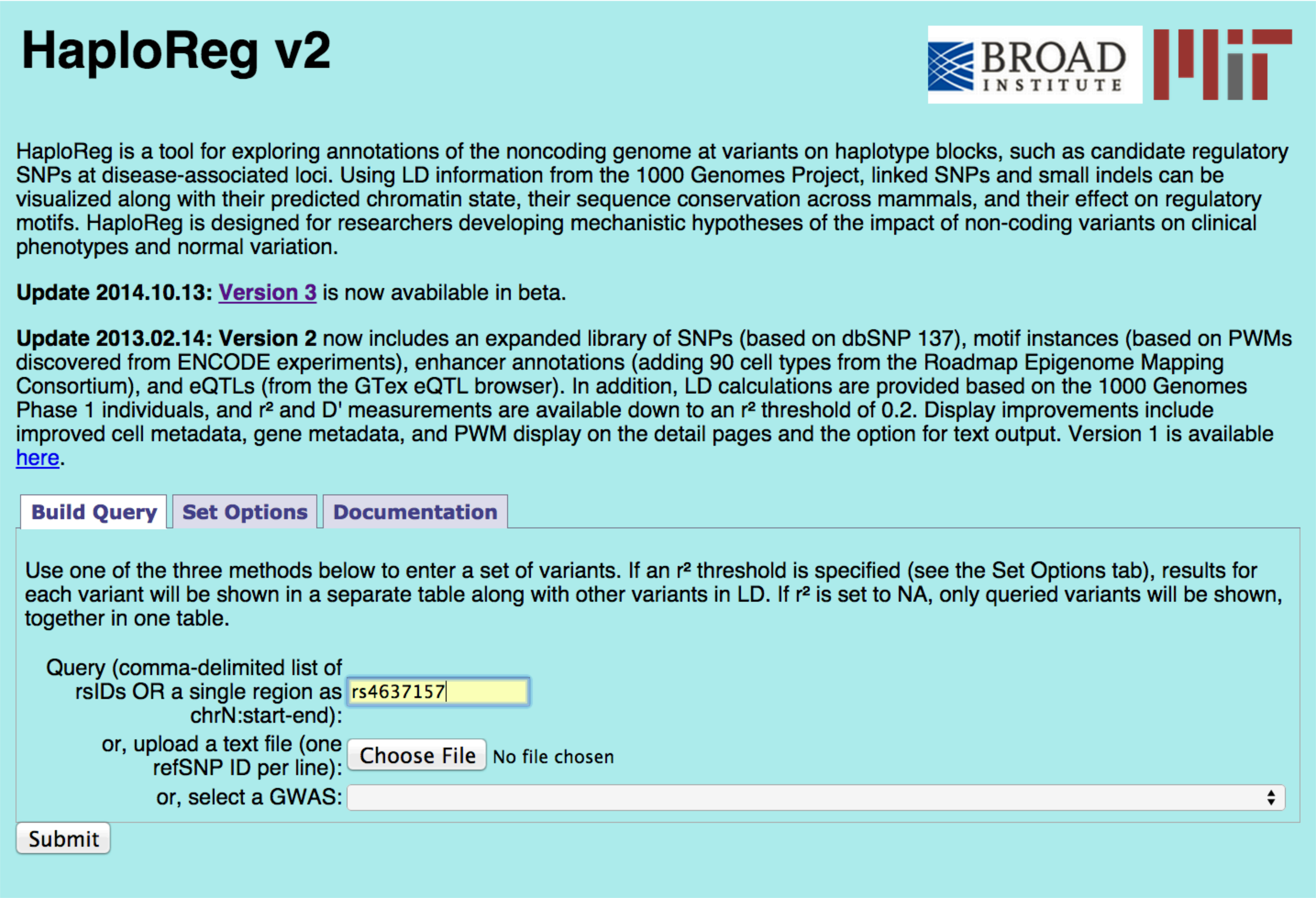
Companion website

Marc A. Schaub, Alan P. Boyle, Anshul Kundaje, Serafim Batzoglou, Michael Snyder
Stanford University

Access the list of GWAS associations, and the corresponding fSNPs:

- List of all associated SNPs
- By phenotype:
  - 5-HTT brain serotonin transporter levels
  - AB1-42
  - AIDS
  - AIDS progression
  - Abdominal aortic aneurysm
  - Acenocoumarol maintenance dosage
  - Activated partial thromboplastin time
  - Acute lymphoblastic leukemia (childhood)
  - Adiponectin levels
  - Adiposity
  - Adverse response to aromatase inhibitors
  - Adverse response to carbamapezine
  - Age-related macular degeneration
  - Age-related macular degeneration (wet)
  - Aging
  - Aging traits
  - Alcohol consumption
  - Alcohol dependence
  - Alcoholism (12-month weekly alcohol consumption)
  - Alcoholism (alcohol dependence factor score)
  - Alcoholism (alcohol use disorder factor score)
  - Alcoholism (heaviness of drinking)
  - Alopecia areata
  - Alzheimer's disease
  - Alzheimer's disease (late onset)
  - Alzheimer's disease biomarkers
  - Amyloid A Levels
  - Amyotrophic lateral sclerosis
  - Angiotensin-converting enzyme activity
  - Ankylosing spondylitis

http://regulome.stanford.edu/GWAS



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 2014.10.13: Version 3 is now avabilable in beta.

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 r² and D' measurements are available down to an r² 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.

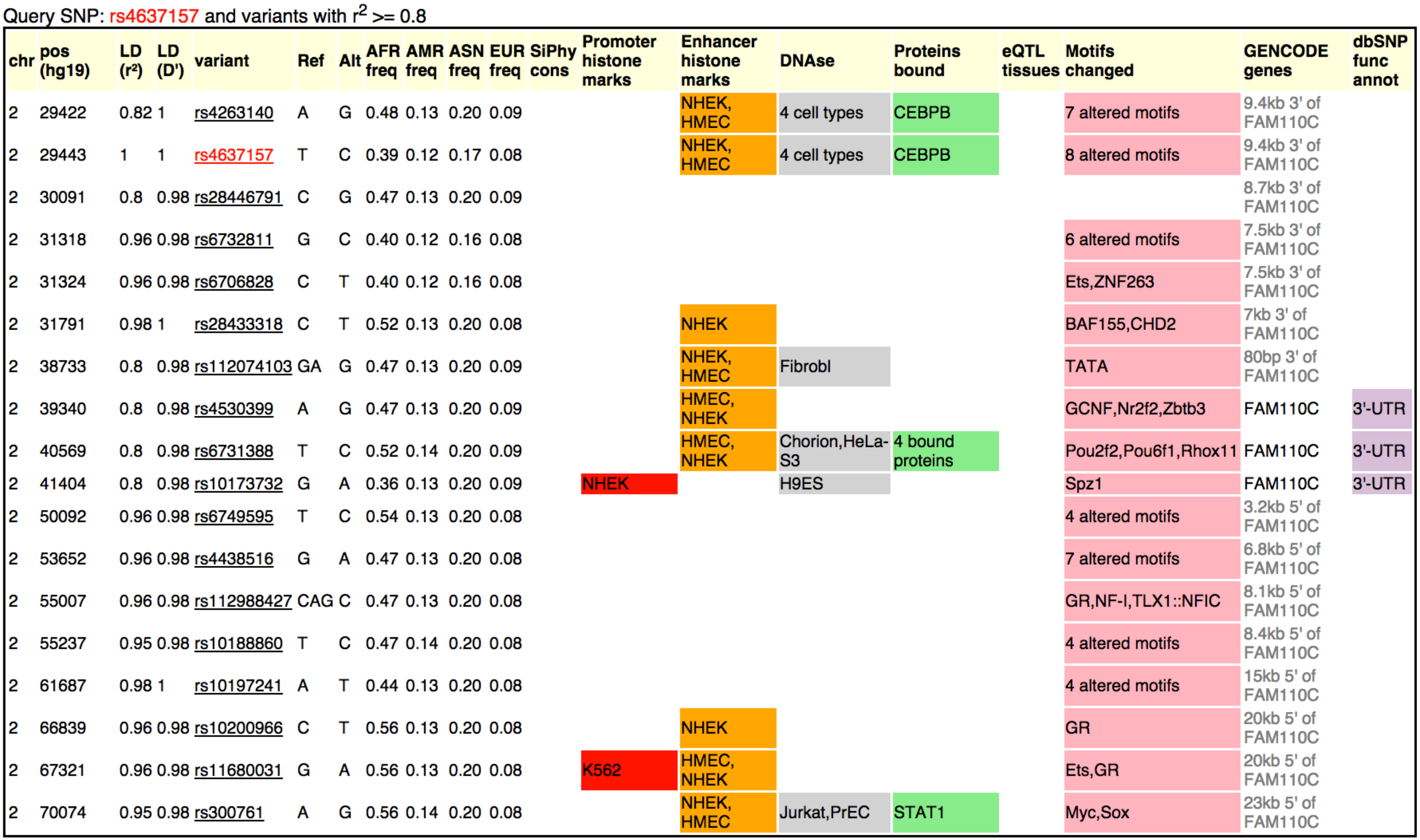
Use one of the three methods below to enter a set of variants. If an r² threshold is specified (see the Set Options tab), results for each variant will be shown in a separate table along with other variants in LD. If r² is set to NA, only queried variants will be shown, together in one table.

Query (comma-delimited list of rsIDs OR a single region as chrN:start-end): rs4637157

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

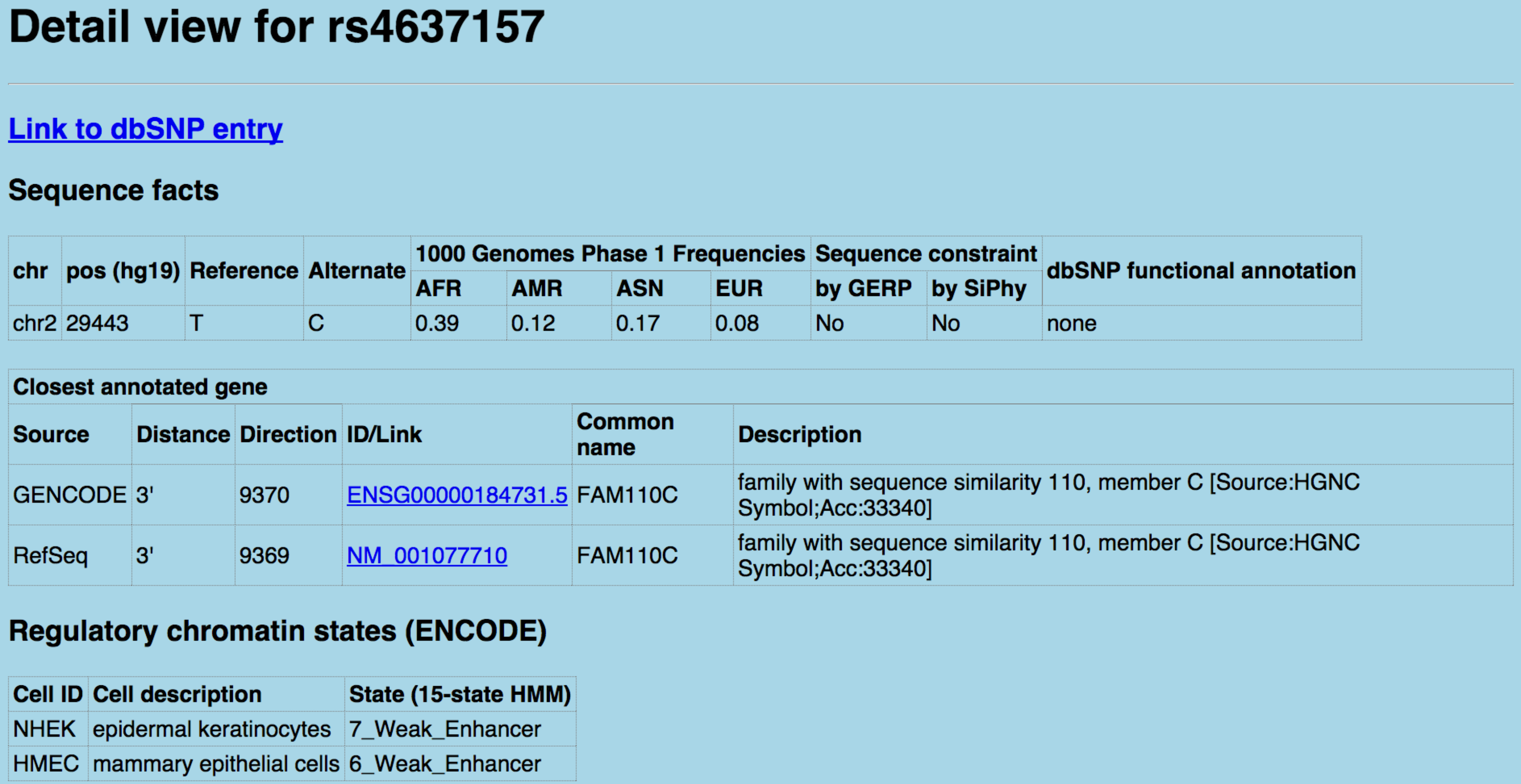
or, select a GWAS:

Submit



Query SNP: rs4637157 and variants with r^2 >= 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
2	29422	0.82	1	rs4263140	A	G	0.48	0.13	0.20	0.09			NHEK, HMEC	4 cell types	CEBPB		7 altered motifs	9.4kb 3' of FAM110C	
2	29443	1	1	rs4637157	T	C	0.39	0.12	0.17	0.08			NHEK, HMEC	4 cell types	CEBPB		8 altered motifs	9.4kb 3' of FAM110C	
2	30091	0.8	0.98	rs28446791	C	G	0.47	0.13	0.20	0.09								8.7kb 3' of FAM110C	
2	31318	0.96	0.98	rs6732811	G	C	0.40	0.12	0.16	0.08							6 altered motifs	7.5kb 3' of FAM110C	
2	31324	0.96	0.98	rs6706828	C	T	0.40	0.12	0.16	0.08							Ets,ZNF263	7.5kb 3' of FAM110C	
2	31791	0.98	1	rs28433318	C	T	0.52	0.13	0.20	0.08			NHEK				BAF155,CHD2	7kb 3' of FAM110C	
2	38733	0.8	0.98	rs112074103	GA	G	0.47	0.13	0.20	0.09			NHEK, HMEC	Fibrobl			TATA	80bp 3' of FAM110C	
2	39340	0.8	0.98	rs4530399	A	G	0.47	0.13	0.20	0.09			HMEC, NHEK				GCNF,Nr2f2,Zbtb3	FAM110C	3'-UTR
2	40569	0.8	0.98	rs6731388	T	C	0.52	0.14	0.20	0.09			HMEC, NHEK	Chorion,HeLa-S3	4 bound proteins		Pou2f2,Pou6f1,Rhox11	FAM110C	3'-UTR
2	41404	0.8	0.98	rs10173732	G	A	0.36	0.13	0.20	0.09		NHEK		H9ES			Spz1	FAM110C	3'-UTR
2	50092	0.96	0.98	rs6749595	T	C	0.54	0.13	0.20	0.08							4 altered motifs	3.2kb 5' of FAM110C	
2	53652	0.96	0.98	rs4438516	G	A	0.47	0.13	0.20	0.08							7 altered motifs	6.8kb 5' of FAM110C	
2	55007	0.96	0.98	rs112988427	CAG	C	0.47	0.13	0.20	0.08							GR,NF-I,TLX1::NFIC	8.1kb 5' of FAM110C	
2	55237	0.95	0.98	rs10188860	T	C	0.47	0.14	0.20	0.08							4 altered motifs	8.4kb 5' of FAM110C	
2	61687	0.98	1	rs10197241	A	T	0.44	0.13	0.20	0.08							4 altered motifs	15kb 5' of FAM110C	
2	66839	0.96	0.98	rs10200966	C	T	0.56	0.13	0.20	0.08			NHEK				GR	20kb 5' of FAM110C	
2	67321	0.96	0.98	rs11680031	G	A	0.56	0.13	0.20	0.08		K562	HMEC, NHEK				Ets,GR	20kb 5' of FAM110C	
2	70074	0.95	0.98	rs300761	A	G	0.56	0.14	0.20	0.08			NHEK, HMEC	Jurkat,PrEC	STAT1		Myc,Sox	23kb 5' of FAM110C	



Detail view for rs4637157

Link to dbSNP entry

Sequence facts

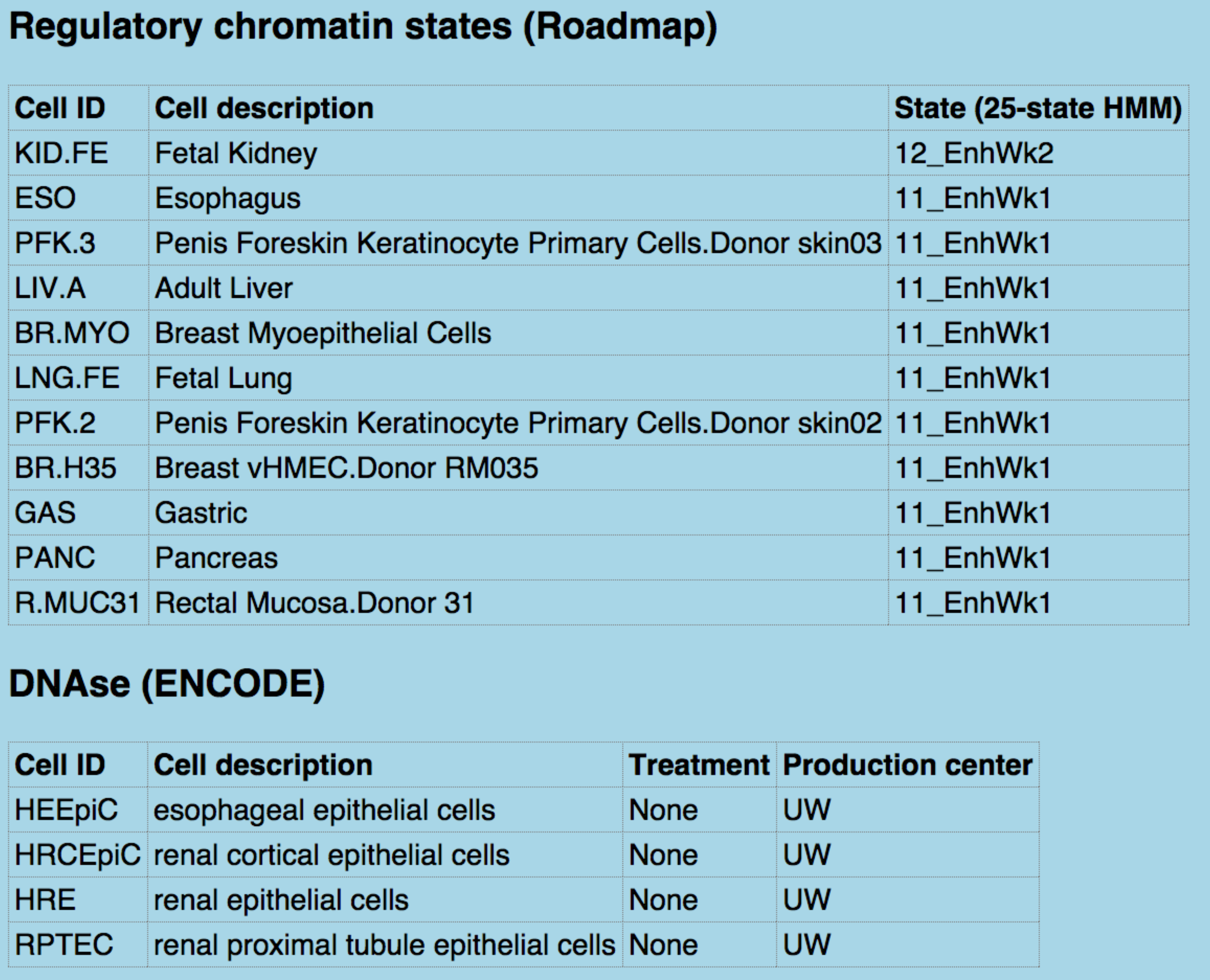
chr	pos (hg19)	Reference	Alternate	1000 Genomes Phase 1 Frequencies				Sequence constraint		dbSNP functional annotation
				AFR	AMR	ASN	EUR	by GERP	by SiPhy	
chr2	29443	T	C	0.39	0.12	0.17	0.08	No	No	none

Closest annotated gene

Source	Distance	Direction	ID/Link	Common name	Description
GENCODE	3'	9370	ENSG00000184731.5	FAM110C	family with sequence similarity 110, member C [Source:HGNC Symbol;Acc:33340]
RefSeq	3'	9369	NM_001077710	FAM110C	family with sequence similarity 110, member C [Source:HGNC Symbol;Acc:33340]

Regulatory chromatin states (ENCODE)

Cell ID	Cell description	State (15-state HMM)
NHEK	epidermal keratinocytes	7_Weak_Enhancer
HMEC	mammary epithelial cells	6_Weak_Enhancer

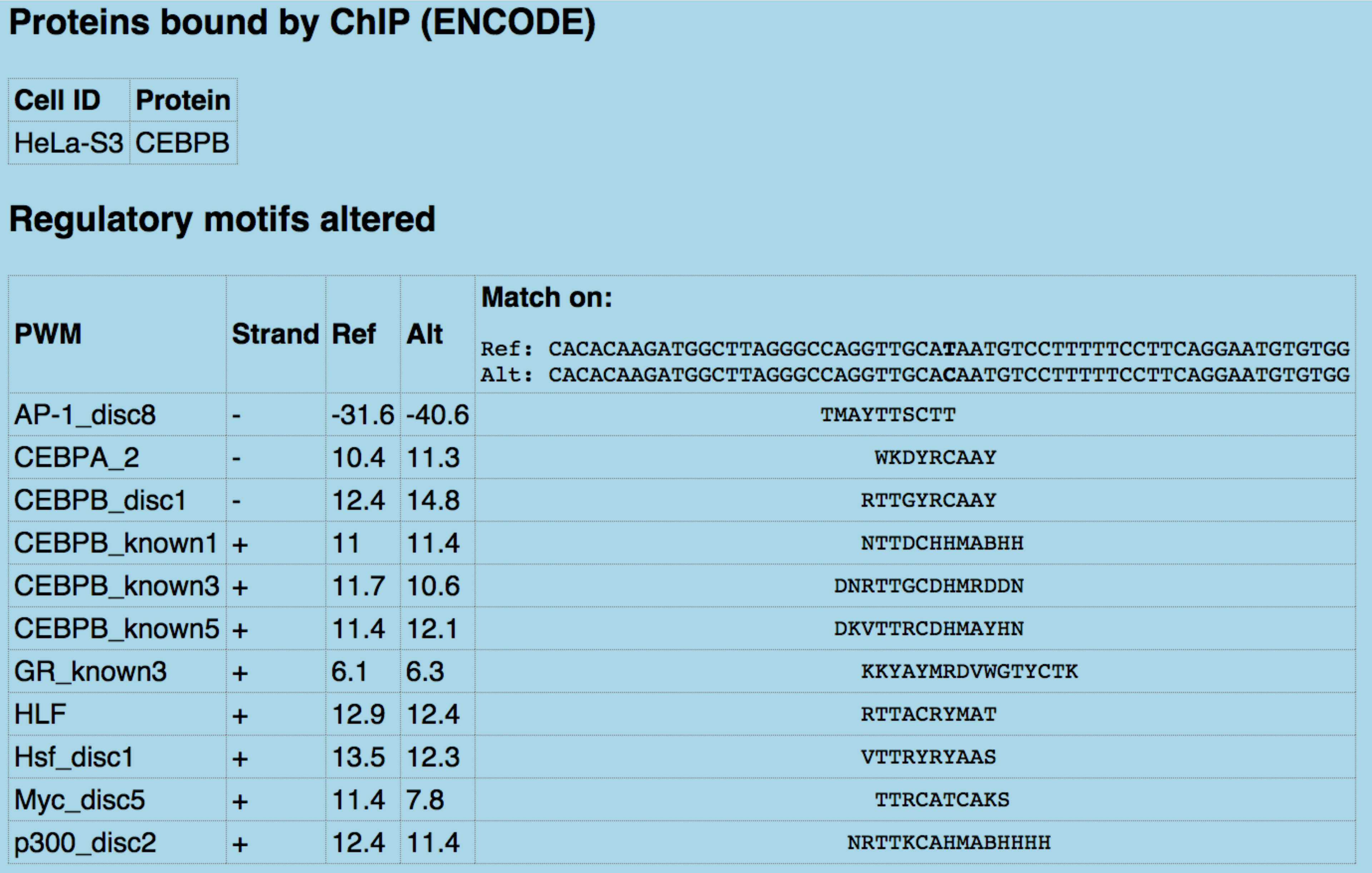


Regulatory chromatin states (Roadmap)

Cell ID	Cell description	State (25-state HMM)
KID.FE	Fetal Kidney	12_EnhWk2
ESO	Esophagus	11_EnhWk1
PFK.3	Penis Foreskin Keratinocyte Primary Cells.Donor skin03	11_EnhWk1
LIV.A	Adult Liver	11_EnhWk1
BR.MYO	Breast Myoepithelial Cells	11_EnhWk1
LNG.FE	Fetal Lung	11_EnhWk1
PFK.2	Penis Foreskin Keratinocyte Primary Cells.Donor skin02	11_EnhWk1
BR.H35	Breast vHMEC.Donor RM035	11_EnhWk1
GAS	Gastric	11_EnhWk1
PANC	Pancreas	11_EnhWk1
R.MUC31	Rectal Mucosa.Donor 31	11_EnhWk1

DNAse (ENCODE)

Cell ID	Cell description	Treatment	Production center
HEEpiC	esophageal epithelial cells	None	UW
HRCEpiC	renal cortical epithelial cells	None	UW
HRE	renal epithelial cells	None	UW
RPTEC	renal proximal tubule epithelial cells	None	UW

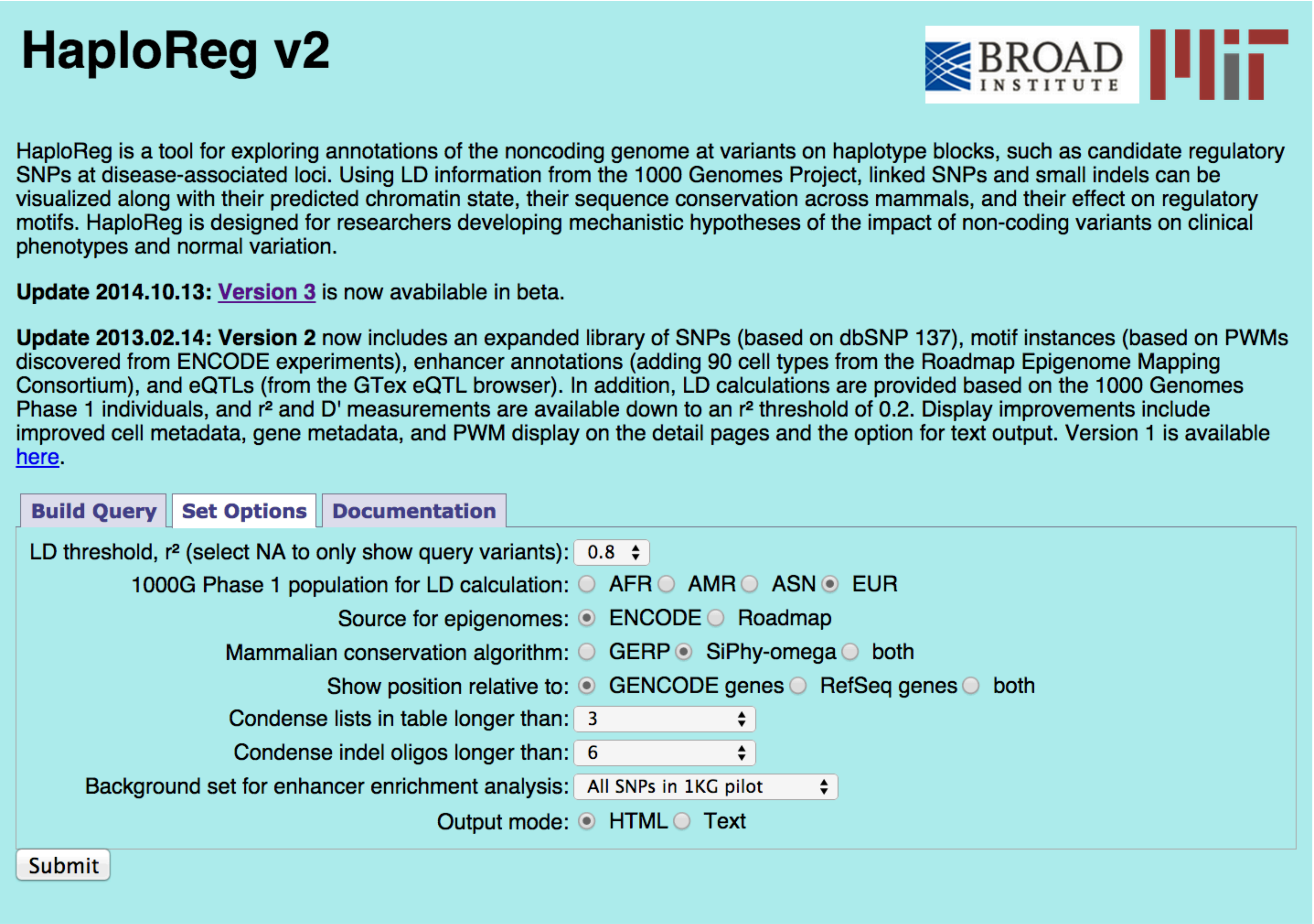


Proteins bound by ChIP (ENCODE)

Cell ID	Protein
HeLa-S3	CEBPB

Regulatory motifs altered

PWM	Strand	Ref	Alt	Match on: Ref: CACACAAGATGGCTTAGGGCCAGGTTGCA**T**AATGTCCTTTTTCCTTCAGGAATGTGTGG Alt: CACACAAGATGGCTTAGGGCCAGGTTGCA**C**AATGTCCTTTTTCCTTCAGGAATGTGTGG
AP-1_disc8	-	-31.6	-40.6	TMAYTTSCTT
CEBPA_2	-	10.4	11.3	WKDYRCAAY
CEBPB_disc1	-	12.4	14.8	RTTGYRCAAY
CEBPB_known1	+	11	11.4	NTTDCHHMABHH
CEBPB_known3	+	11.7	10.6	DNRTTGCDHMRDDN
CEBPB_known5	+	11.4	12.1	DKVTTRCDHMAYHN
GR_known3	+	6.1	6.3	KKYAYMRDVWGTYCTK
HLF	+	12.9	12.4	RTTACRYMAT
Hsf_disc1	+	13.5	12.3	VTTRYRYAAS
Myc_disc5	+	11.4	7.8	TTRCATCAKS
p300_disc2	+	12.4	11.4	NRTTKCAHMABHHHH



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 2014.10.13: Version 3 is now avabilable in beta.

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 r² and D' measurements are available down to an r² 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.

LD threshold, r² (select NA to only show query variants): 0.8

1000G Phase 1 population for LD calculation: ○ AFR ○ AMR ○ ASN ● EUR

Source for epigenomes: ● ENCODE ○ Roadmap

Mammalian conservation algorithm: ○ GERP ● SiPhy-omega ○ both

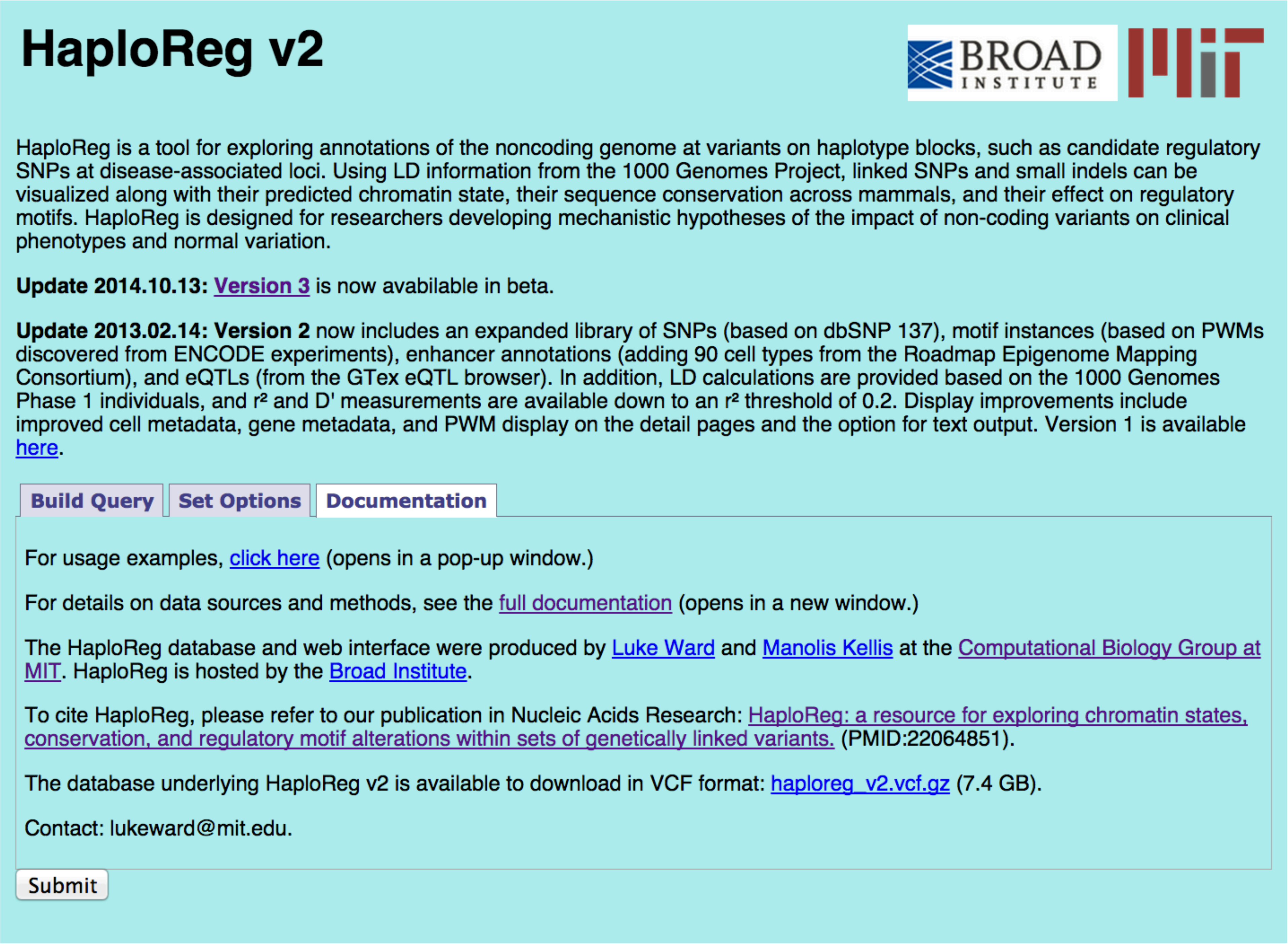
Show position relative to: ● GENCODE genes ○ RefSeq genes ○ both

Condense lists in table longer than: 3

Condense indel oligos longer than: 6

Background set for enhancer enrichment analysis: All SNPs in 1KG pilot

Output mode: ● HTML ○ Text



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 2014.10.13: Version 3 is now avabilable in beta.

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.

For usage examples, click here (opens in a pop-up window.)

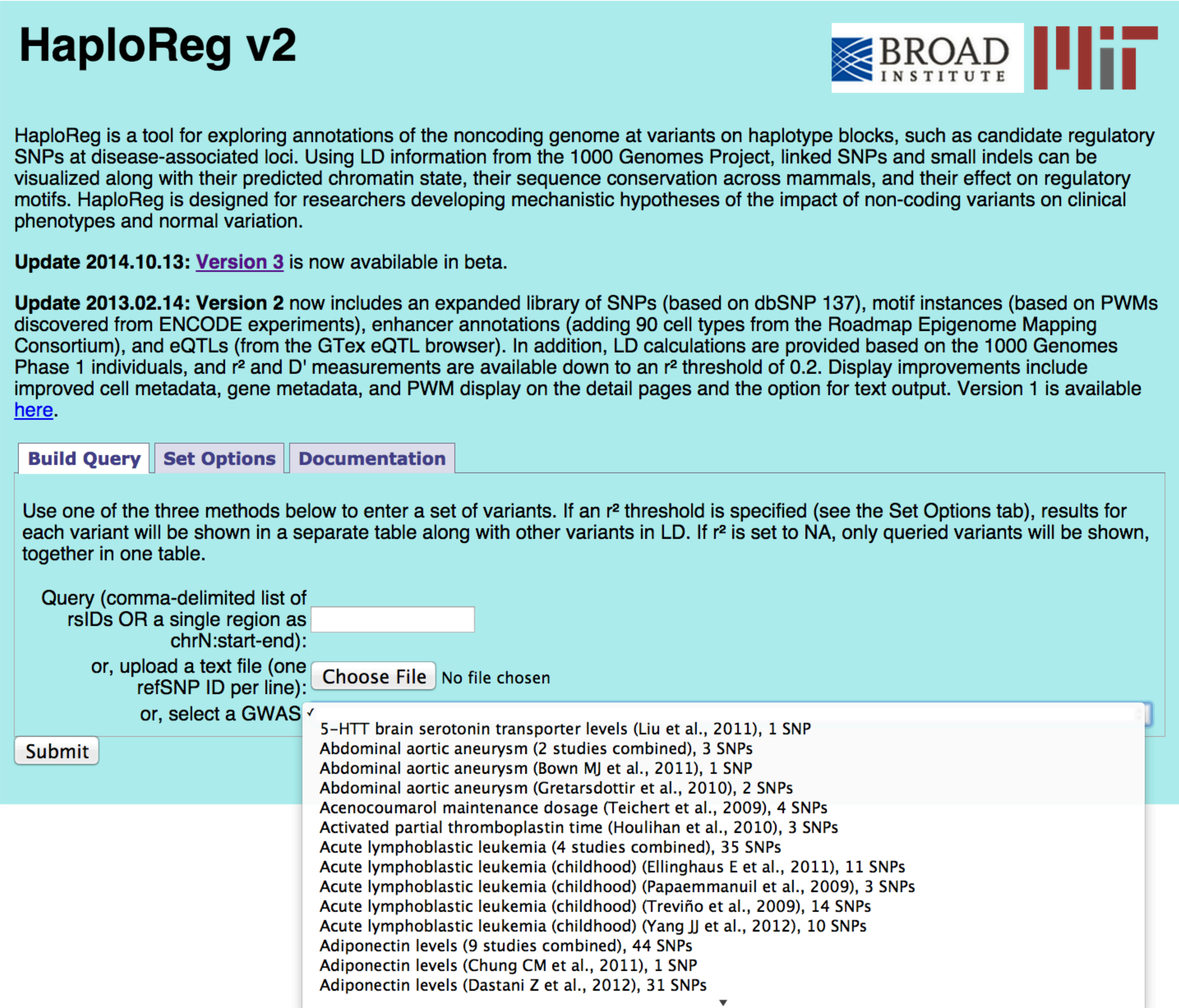
For details on data sources and methods, see the full documentation (opens in a new window.)

The HaploReg database and web interface were produced by Luke Ward and Manolis Kellis at the Computational Biology Group at MIT. HaploReg is hosted by the Broad Institute.

To cite HaploReg, please refer to our publication in Nucleic Acids Research: HaploReg: a resource for exploring chromatin states, conservation, and regulatory motif alterations within sets of genetically linked variants. (PMID:22064851).

The database underlying HaploReg v2 is available to download in VCF format: haploreg_v2.vcf.gz (7.4 GB).

Contact: lukeward@mit.edu.



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 2014.10.13: Version 3 is now avabilable in beta.

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 r² and D' measurements are available down to an r² 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.

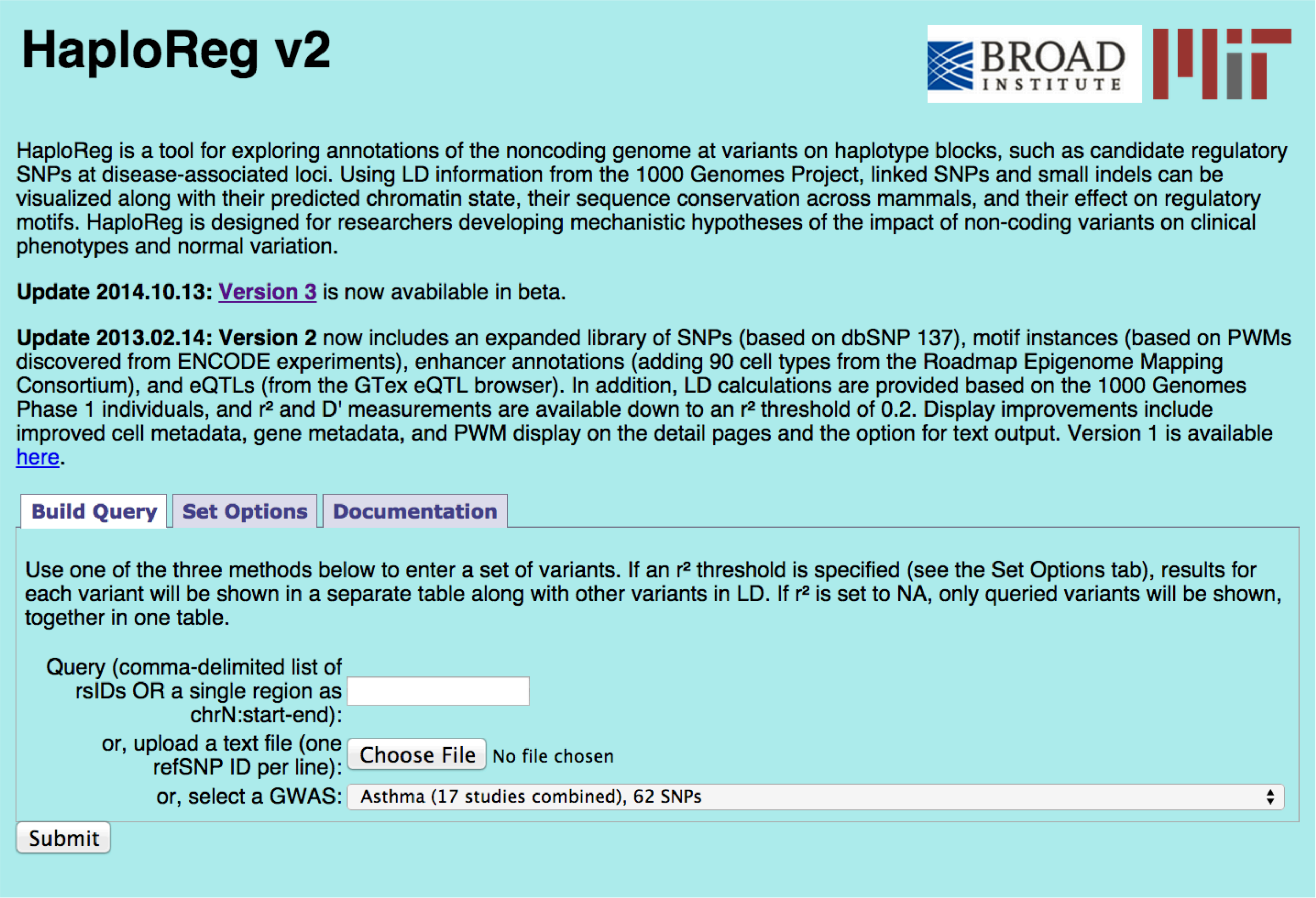
Use one of the three methods below to enter a set of variants. If an r² threshold is specified (see the Set Options tab), results for each variant will be shown in a separate table along with other variants in LD. If r² is set to NA, only queried variants will be shown, together in one table.

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 chosen

or, select a GWAS ✓

5-HTT brain serotonin transporter levels (Liu et al., 2011), 1 SNP
Abdominal aortic aneurysm (2 studies combined), 3 SNPs
Abdominal aortic aneurysm (Bown MJ et al., 2011), 1 SNP
Abdominal aortic aneurysm (Gretarsdottir et al., 2010), 2 SNPs
Acenocoumarol maintenance dosage (Teichert et al., 2009), 4 SNPs
Activated partial thromboplastin time (Houlihan et al., 2010), 3 SNPs
Acute lymphoblastic leukemia (4 studies combined), 35 SNPs
Acute lymphoblastic leukemia (childhood) (Ellinghaus E et al., 2011), 11 SNPs
Acute lymphoblastic leukemia (childhood) (Papaemmanuil et al., 2009), 3 SNPs
Acute lymphoblastic leukemia (childhood) (Treviño et al., 2009), 14 SNPs
Acute lymphoblastic leukemia (childhood) (Yang JJ et al., 2012), 10 SNPs
Adiponectin levels (9 studies combined), 44 SNPs
Adiponectin levels (Chung CM et al., 2011), 1 SNP
Adiponectin levels (Dastani Z et al., 2012), 31 SNPs



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 2014.10.13: Version 3 is now avabilable in beta.

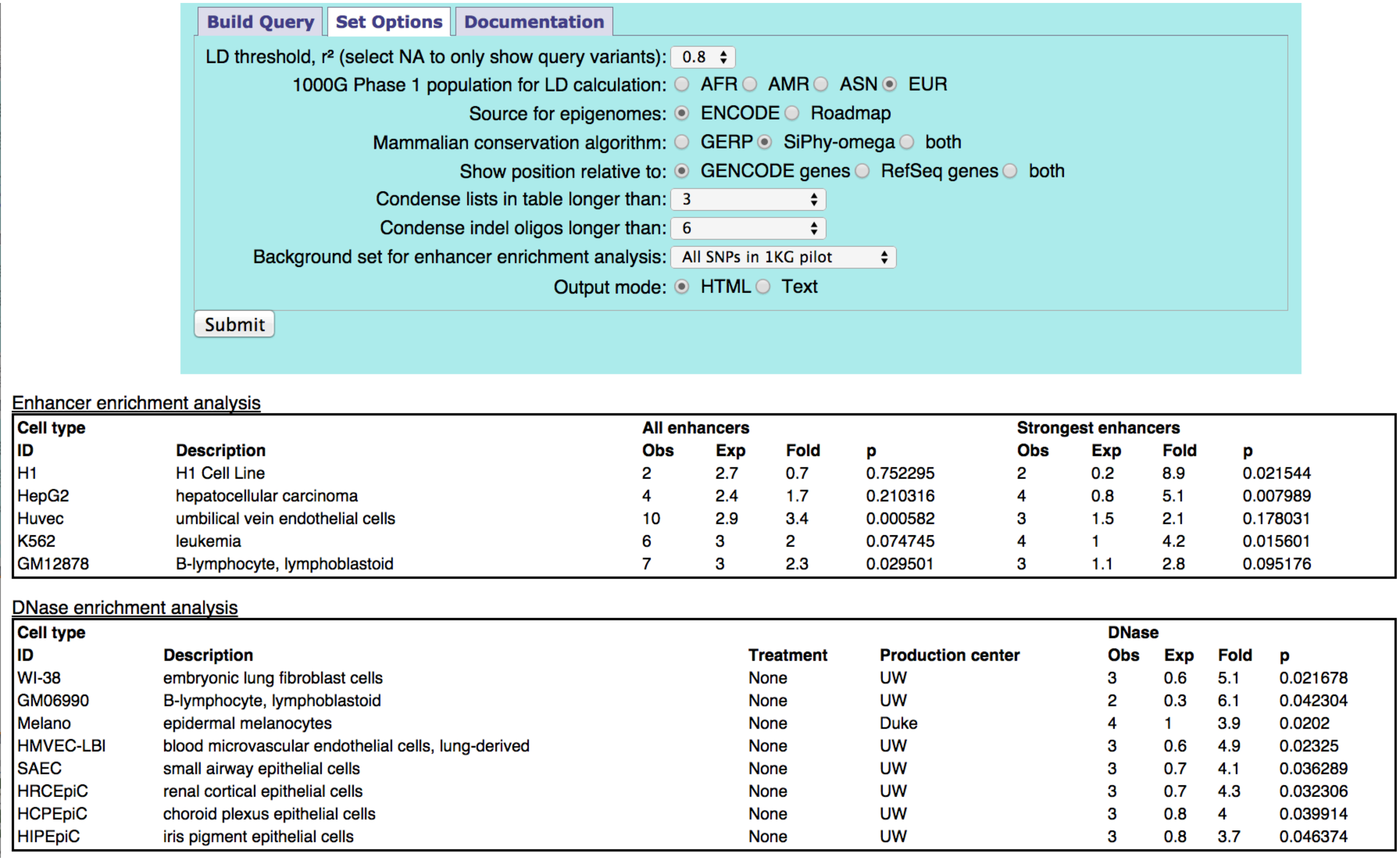
Update 2013.02.14: Version 2 now includes an expanded library of SNPs (based on dbSNP 137), motif instances (based on PWMs discovered from ENCODE experiments), enhancer annotations (adding 90 cell types from the Roadmap Epigenome Mapping Consortium), and eQTLs (from the GTex eQTL browser). In addition, LD calculations are provided based on the 1000 Genomes Phase 1 individuals, and r² and D' measurements are available down to an r² 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.

Use one of the three methods below to enter a set of variants. If an r² 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 r² 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): No file chosen

or, select a GWAS: Asthma (17 studies combined), 62 SNPs



LD threshold, r² (select NA to only show query variants): 0.8

1000G Phase 1 population for LD calculation: ○ AFR ○ AMR ○ ASN ● EUR

Source for epigenomes: ● ENCODE ○ Roadmap

Mammalian conservation algorithm: ○ GERP ● SiPhy-omega ○ both

Show position relative to: ● GENCODE genes ○ RefSeq genes ○ both

Condense lists in table longer than: 3

Condense indel oligos longer than: 6

Background set for enhancer enrichment analysis: All SNPs in 1KG pilot

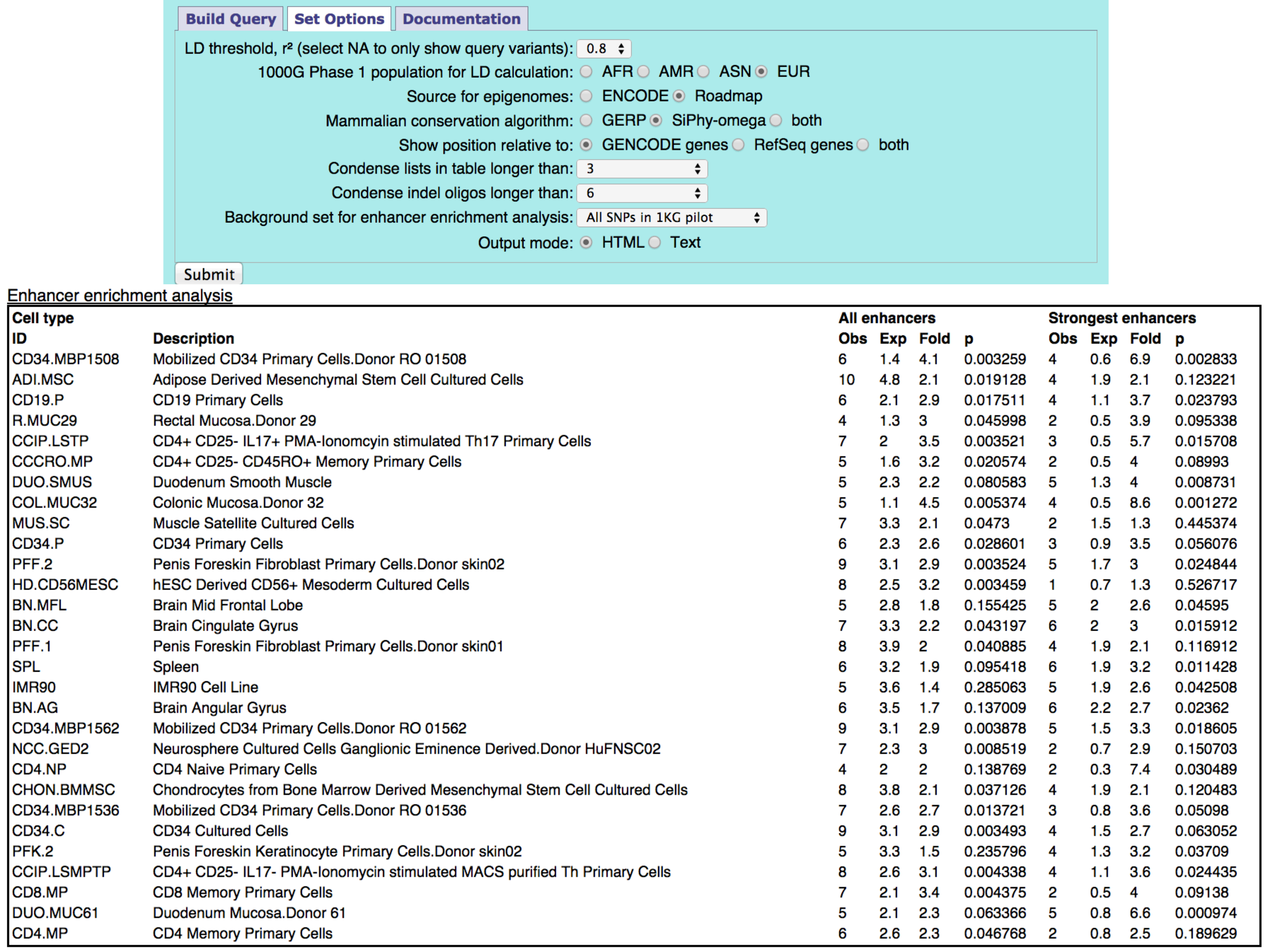
Output mode: ● HTML ○ Text

Enhancer enrichment analysis

Cell type		All enhancers				Strongest enhancers			
ID	**Description**	**Obs**	**Exp**	**Fold**	**p**	**Obs**	**Exp**	**Fold**	**p**
H1	H1 Cell Line	2	2.7	0.7	0.752295	2	0.2	8.9	0.021544
HepG2	hepatocellular carcinoma	4	2.4	1.7	0.210316	4	0.8	5.1	0.007989
Huvec	umbilical vein endothelial cells	10	2.9	3.4	0.000582	3	1.5	2.1	0.178031
K562	leukemia	6	3	2	0.074745	4	1	4.2	0.015601
GM12878	B-lymphocyte, lymphoblastoid	7	3	2.3	0.029501	3	1.1	2.8	0.095176

DNase enrichment analysis

Cell type				DNase			
ID	**Description**	**Treatment**	**Production center**	**Obs**	**Exp**	**Fold**	**p**
WI-38	embryonic lung fibroblast cells	None	UW	3	0.6	5.1	0.021678
GM06990	B-lymphocyte, lymphoblastoid	None	UW	2	0.3	6.1	0.042304
Melano	epidermal melanocytes	None	Duke	4	1	3.9	0.0202
HMVEC-LBl	blood microvascular endothelial cells, lung-derived	None	UW	3	0.6	4.9	0.02325
SAEC	small airway epithelial cells	None	UW	3	0.7	4.1	0.036289
HRCEpiC	renal cortical epithelial cells	None	UW	3	0.7	4.3	0.032306
HCPEpiC	choroid plexus epithelial cells	None	UW	3	0.8	4	0.039914
HIPEpiC	iris pigment epithelial cells	None	UW	3	0.8	3.7	0.046374



Build Query | Set Options | Documentation

LD threshold, r^2 (select NA to only show query variants): 0.8

1000G Phase 1 population for LD calculation: ○ AFR ○ AMR ○ ASN ● EUR

Source for epigenomes: ○ ENCODE ● Roadmap

Mammalian conservation algorithm: ○ GERP ● SiPhy-omega ○ both

Show position relative to: ● GENCODE genes ○ RefSeq genes ○ both

Condense lists in table longer than: 3

Condense indel oligos longer than: 6

Background set for enhancer enrichment analysis: All SNPs in 1KG pilot

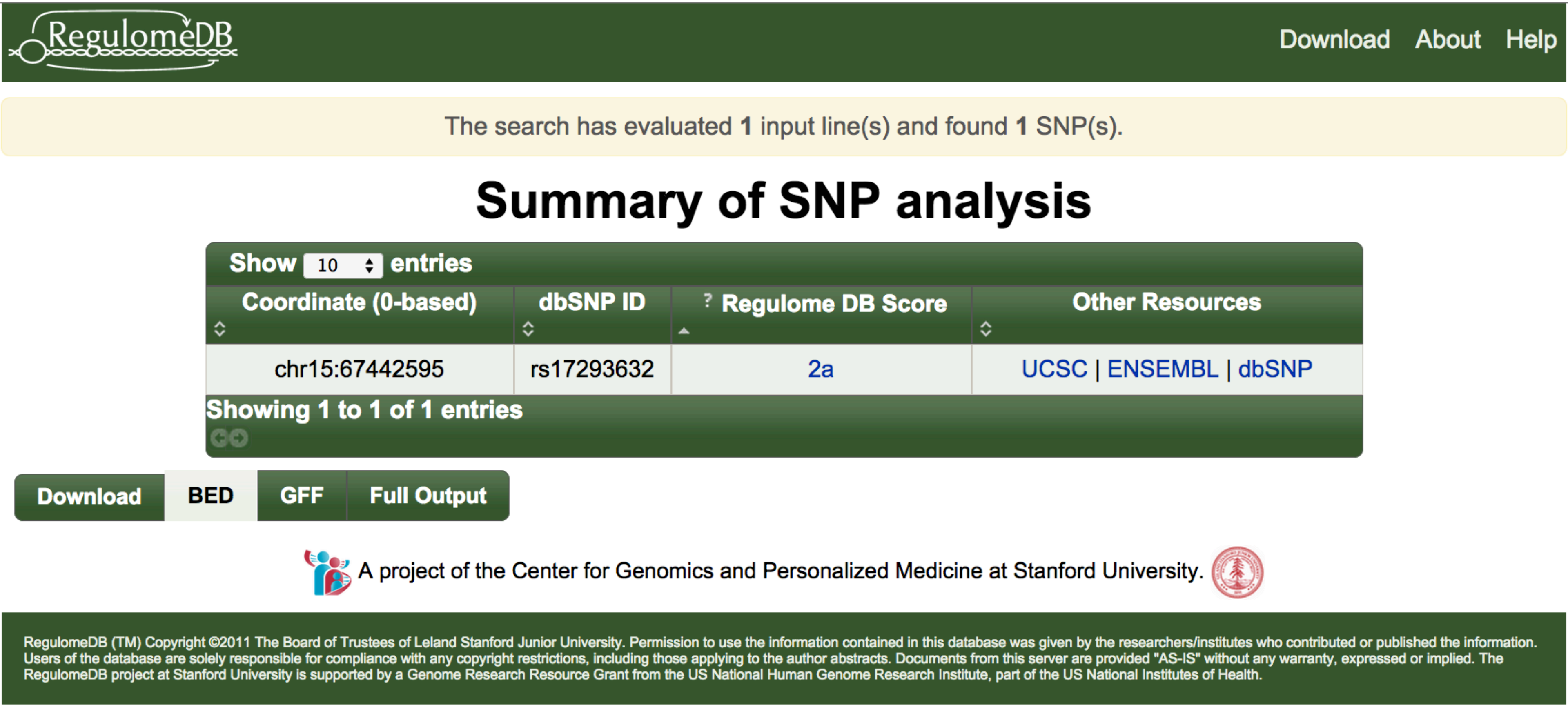
Output mode: ● HTML ○ Text

Submit

Enhancer enrichment analysis

Cell type		All enhancers				Strongest enhancers			
ID	Description	Obs	Exp	Fold	p	Obs	Exp	Fold	p
CD34.MBP1508	Mobilized CD34 Primary Cells.Donor RO 01508	6	1.4	4.1	0.003259	4	0.6	6.9	0.002833
ADI.MSC	Adipose Derived Mesenchymal Stem Cell Cultured Cells	10	4.8	2.1	0.019128	4	1.9	2.1	0.123221
CD19.P	CD19 Primary Cells	6	2.1	2.9	0.017511	4	1.1	3.7	0.023793
R.MUC29	Rectal Mucosa.Donor 29	4	1.3	3	0.045998	2	0.5	3.9	0.095338
CCIP.LSTP	CD4+ CD25- IL17+ PMA-Ionomcyin stimulated Th17 Primary Cells	7	2	3.5	0.003521	3	0.5	5.7	0.015708
CCCRO.MP	CD4+ CD25- CD45RO+ Memory Primary Cells	5	1.6	3.2	0.020574	2	0.5	4	0.08993
DUO.SMUS	Duodenum Smooth Muscle	5	2.3	2.2	0.080583	5	1.3	4	0.008731
COL.MUC32	Colonic Mucosa.Donor 32	5	1.1	4.5	0.005374	4	0.5	8.6	0.001272
MUS.SC	Muscle Satellite Cultured Cells	7	3.3	2.1	0.0473	2	1.5	1.3	0.445374
CD34.P	CD34 Primary Cells	6	2.3	2.6	0.028601	3	0.9	3.5	0.056076
PFF.2	Penis Foreskin Fibroblast Primary Cells.Donor skin02	9	3.1	2.9	0.003524	5	1.7	3	0.024844
HD.CD56MESC	hESC Derived CD56+ Mesoderm Cultured Cells	8	2.5	3.2	0.003459	1	0.7	1.3	0.526717
BN.MFL	Brain Mid Frontal Lobe	5	2.8	1.8	0.155425	5	2	2.6	0.04595
BN.CC	Brain Cingulate Gyrus	7	3.3	2.2	0.043197	6	2	3	0.015912
PFF.1	Penis Foreskin Fibroblast Primary Cells.Donor skin01	8	3.9	2	0.040885	4	1.9	2.1	0.116912
SPL	Spleen	6	3.2	1.9	0.095418	6	1.9	3.2	0.011428
IMR90	IMR90 Cell Line	5	3.6	1.4	0.285063	5	1.9	2.6	0.042508
BN.AG	Brain Angular Gyrus	6	3.5	1.7	0.137009	6	2.2	2.7	0.02362
CD34.MBP1562	Mobilized CD34 Primary Cells.Donor RO 01562	9	3.1	2.9	0.003878	5	1.5	3.3	0.018605
NCC.GED2	Neurosphere Cultured Cells Ganglionic Eminence Derived.Donor HuFNSC02	7	2.3	3	0.008519	2	0.7	2.9	0.150703
CD4.NP	CD4 Naive Primary Cells	4	2	2	0.138769	2	0.3	7.4	0.030489
CHON.BMMSC	Chondrocytes from Bone Marrow Derived Mesenchymal Stem Cell Cultured Cells	8	3.8	2.1	0.037126	4	1.9	2.1	0.120483
CD34.MBP1536	Mobilized CD34 Primary Cells.Donor RO 01536	7	2.6	2.7	0.013721	3	0.8	3.6	0.05098
CD34.C	CD34 Cultured Cells	9	3.1	2.9	0.003493	4	1.5	2.7	0.063052
PFK.2	Penis Foreskin Keratinocyte Primary Cells.Donor skin02	5	3.3	1.5	0.235796	4	1.3	3.2	0.03709
CCIP.LSMPTP	CD4+ CD25- IL17- PMA-Ionomycin stimulated MACS purified Th Primary Cells	8	2.6	3.1	0.004338	4	1.1	3.6	0.024435
CD8.MP	CD8 Memory Primary Cells	7	2.1	3.4	0.004375	2	0.5	4	0.09138
DUO.MUC61	Duodenum Mucosa.Donor 61	5	2.1	2.3	0.063366	5	0.8	6.6	0.000974
CD4.MP	CD4 Memory Primary Cells	6	2.6	2.3	0.046768	2	0.8	2.5	0.189629

Analyzing rs17293632 with RegulomeDB



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

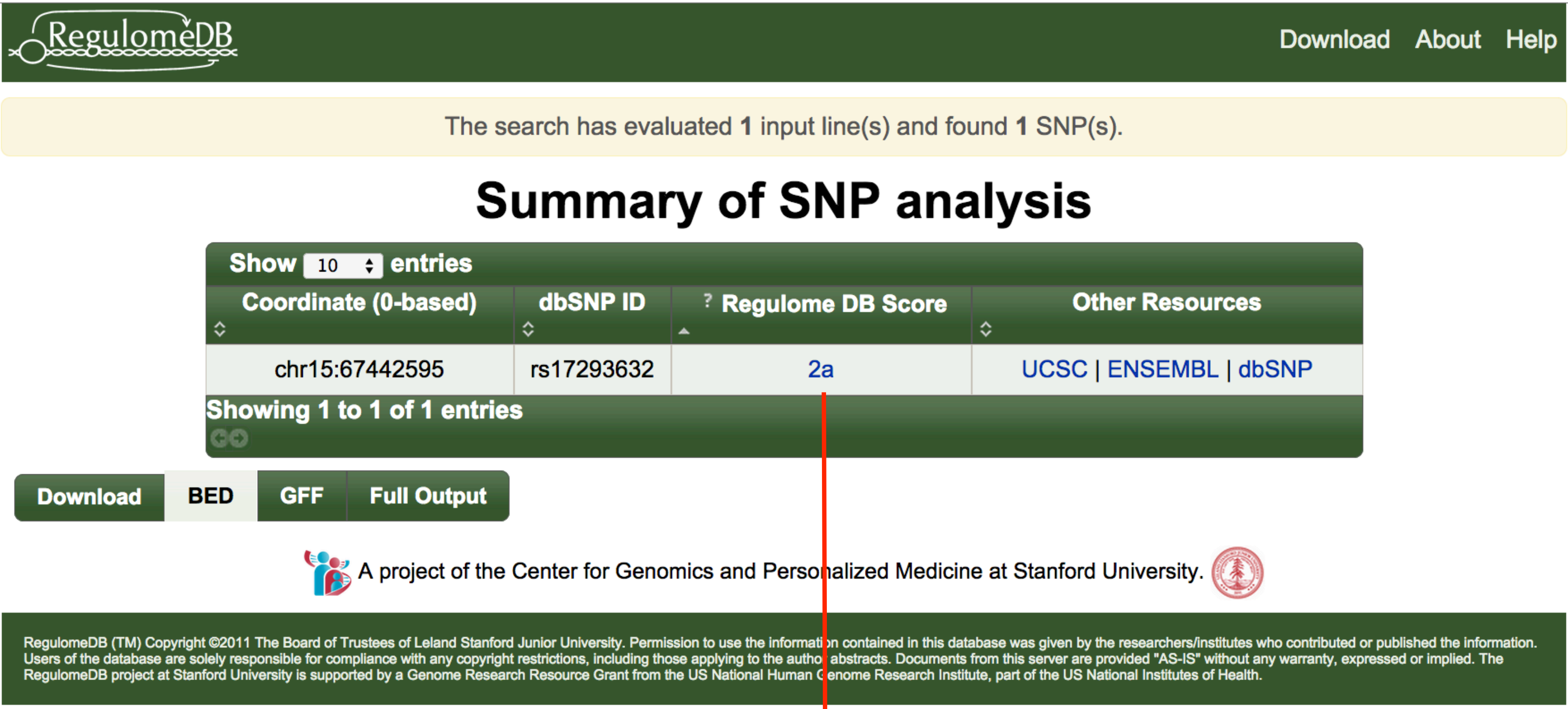
Summary of SNP analysis

Coordinate (0-based)	dbSNP ID	? Regulome DB Score	Other Resources
chr15:67442595	rs17293632	2a	UCSC | ENSEMBL | dbSNP

Showing 1 to 1 of 1 entries

A project of the Center for Genomics and Personalized Medicine at Stanford University.

Analyzing rs17293632 with RegulomeDB



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

Summary of SNP analysis

Show 10 entries

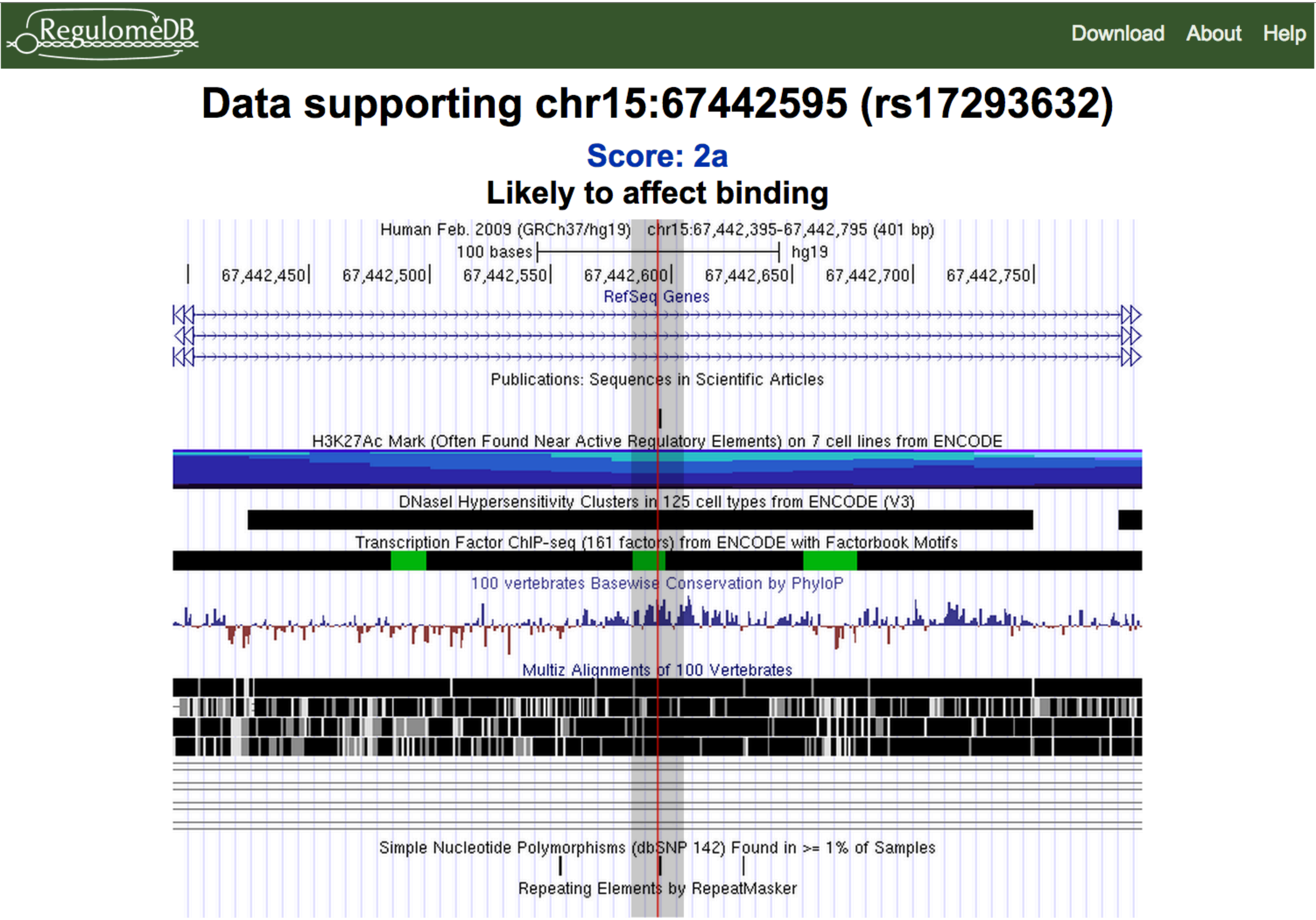
Coordinate (0-based)	dbSNP ID	? Regulome DB Score	Other Resources
chr15:67442595	rs17293632	2a	UCSC | ENSEMBL | dbSNP

Showing 1 to 1 of 1 entries

A project of the Center for Genomics and Personalized Medicine at Stanford University.

"Likely to Affect Binding"

TF binding + matched TF motif + matched DNase footprint + DNase peak



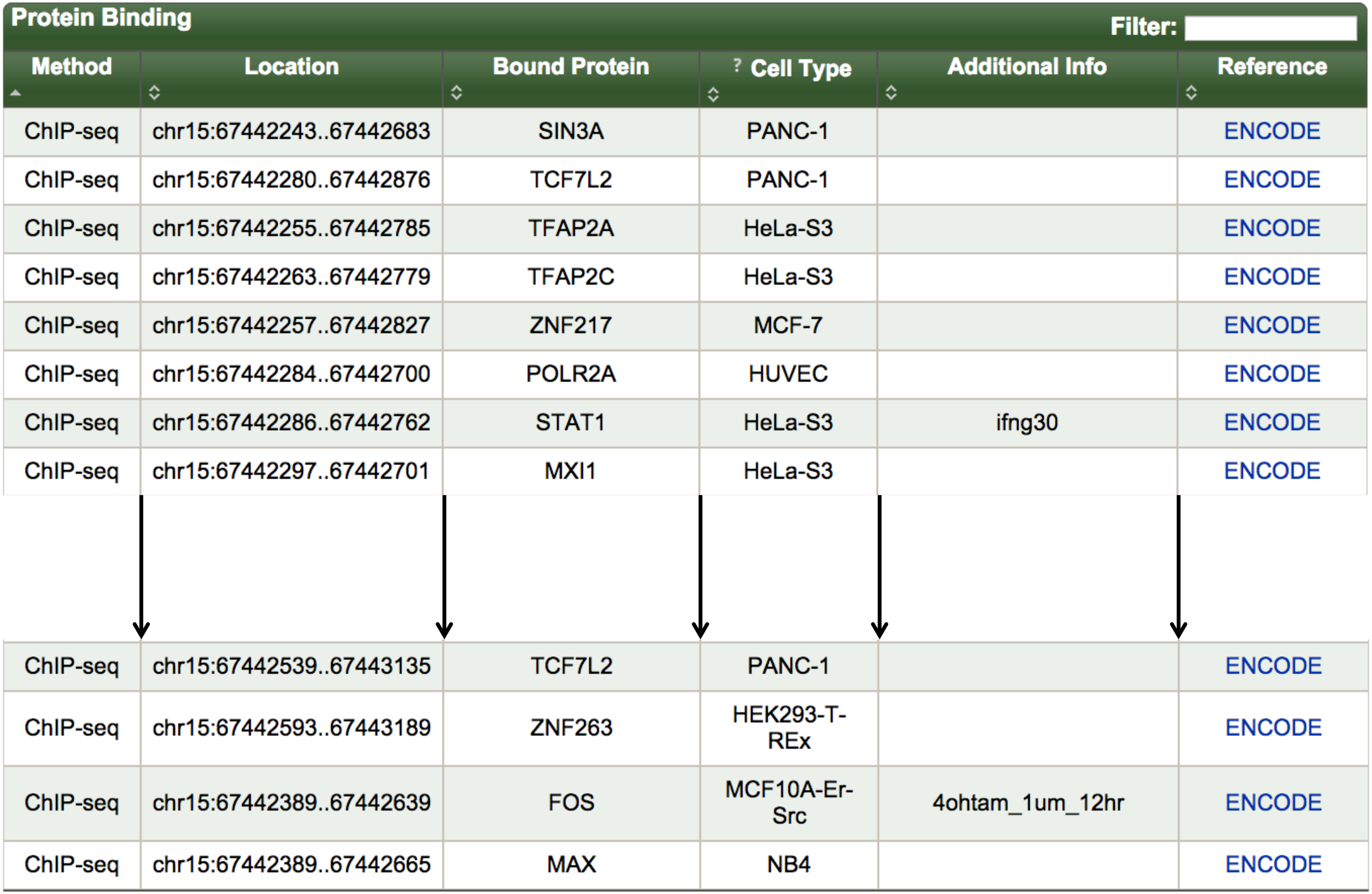
RegulomeDB Download About Help

Data supporting chr15:67442595 (rs17293632)

Score: 2a

Likely to affect binding

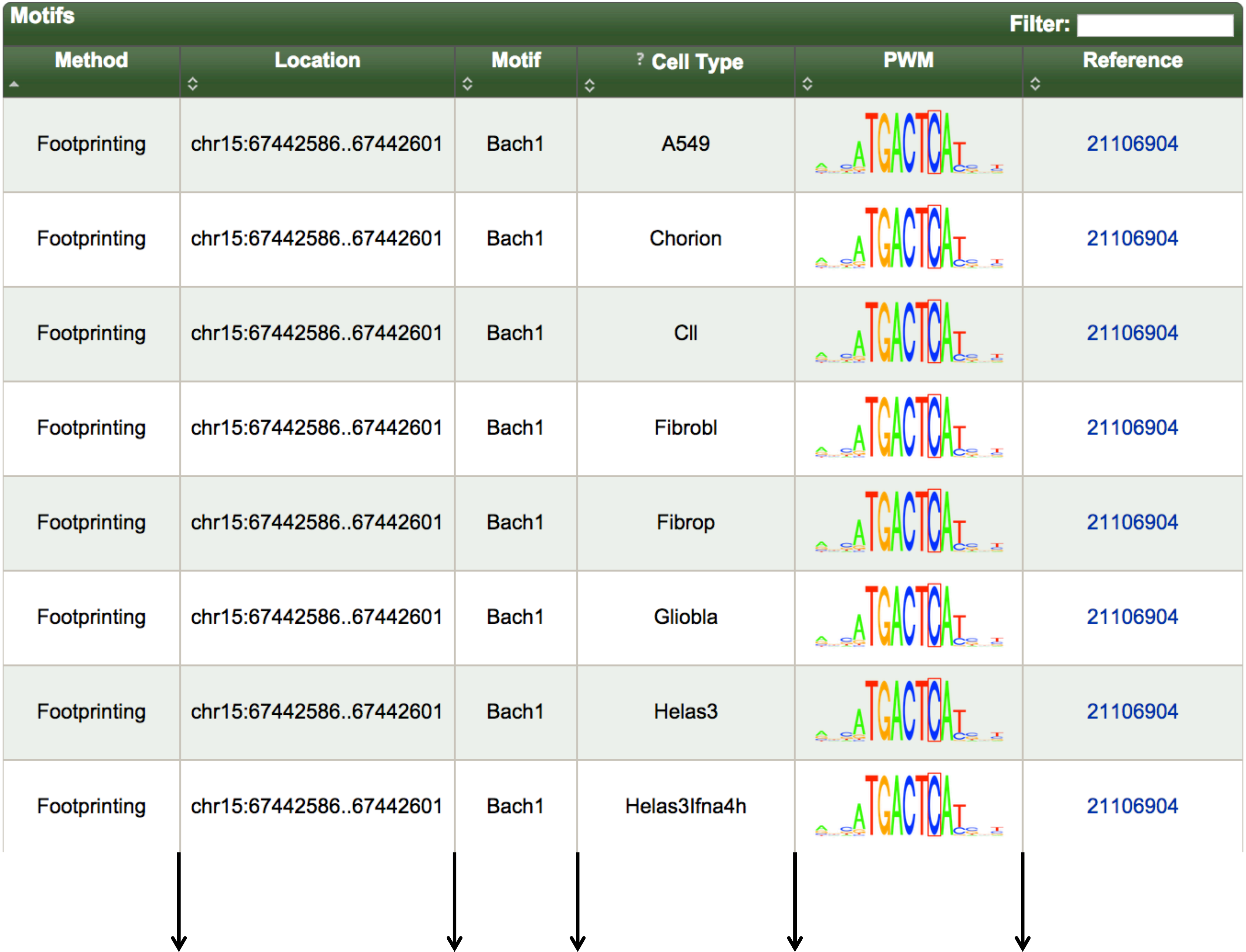
Human Feb. 2009 (GRCh37/hg19) chr15:67,442,395-67,442,795 (401 bp)
100 bases
hg19
67,442,450
67,442,500
67,442,550
67,442,600
67,442,650
67,442,700
67,442,750
RefSeq Genes
Publications: Sequences in Scientific Articles
H3K27Ac Mark (Often Found Near Active Regulatory Elements) on 7 cell lines from ENCODE
DNaseI Hypersensitivity Clusters in 125 cell types from ENCODE (V3)
Transcription Factor ChIP-seq (161 factors) from ENCODE with Factorbook Motifs
100 vertebrates Basewise Conservation by PhyloP
Multiz Alignments of 100 Vertebrates
Simple Nucleotide Polymorphisms (dbSNP 142) Found in >= 1% of Samples
Repeating Elements by RepeatMasker



Protein Binding

Filter:

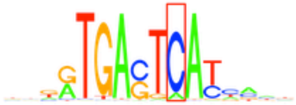
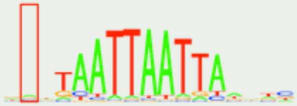
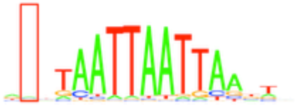
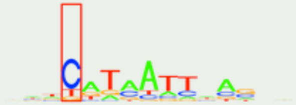
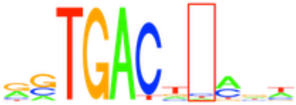
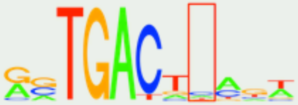
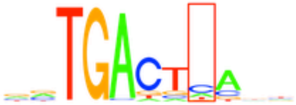
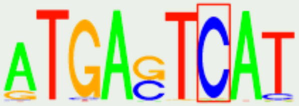
Method	Location	Bound Protein	? Cell Type	Additional Info	Reference
ChIP-seq	chr15:67442243..67442683	SIN3A	PANC-1		ENCODE
ChIP-seq	chr15:67442280..67442876	TCF7L2	PANC-1		ENCODE
ChIP-seq	chr15:67442255..67442785	TFAP2A	HeLa-S3		ENCODE
ChIP-seq	chr15:67442263..67442779	TFAP2C	HeLa-S3		ENCODE
ChIP-seq	chr15:67442257..67442827	ZNF217	MCF-7		ENCODE
ChIP-seq	chr15:67442284..67442700	POLR2A	HUVEC		ENCODE
ChIP-seq	chr15:67442286..67442762	STAT1	HeLa-S3	ifng30	ENCODE
ChIP-seq	chr15:67442297..67442701	MXI1	HeLa-S3		ENCODE
ChIP-seq	chr15:67442539..67443135	TCF7L2	PANC-1		ENCODE
ChIP-seq	chr15:67442593..67443189	ZNF263	HEK293-T-REx		ENCODE
ChIP-seq	chr15:67442389..67442639	FOS	MCF10A-Er-Src	4ohtam_1um_12hr	ENCODE
ChIP-seq	chr15:67442389..67442665	MAX	NB4		ENCODE

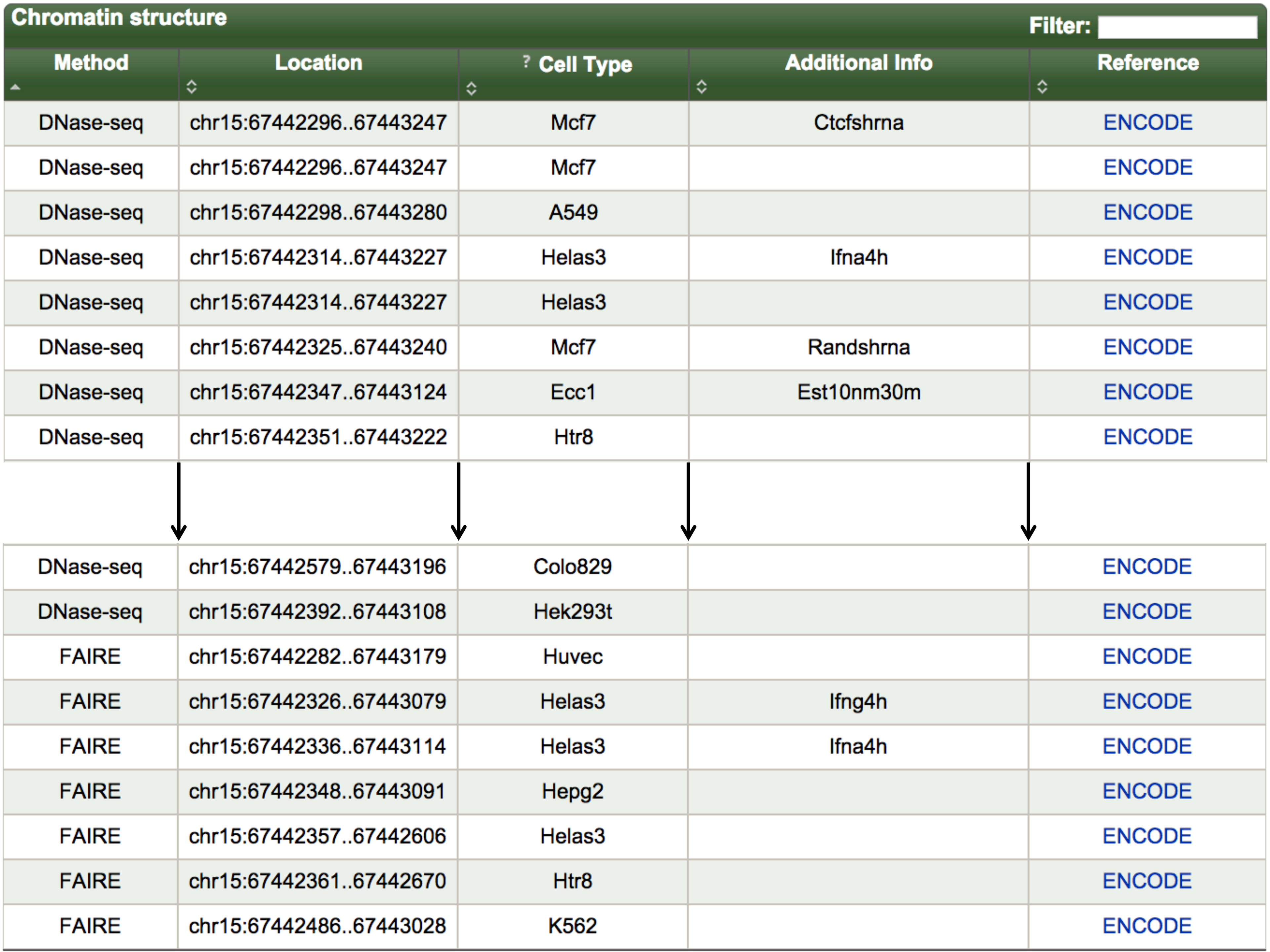


Motifs

Filter:

Method	Location	Motif	? Cell Type	PWM	Reference
Footprinting	chr15:67442586..67442601	Bach1	A549	ATGACTCAT	21106904
Footprinting	chr15:67442586..67442601	Bach1	Chorion	ATGACTCAT	21106904
Footprinting	chr15:67442586..67442601	Bach1	CII	ATGACTCAT	21106904
Footprinting	chr15:67442586..67442601	Bach1	Fibrobl	ATGACTCAT	21106904
Footprinting	chr15:67442586..67442601	Bach1	Fibrop	ATGACTCAT	21106904
Footprinting	chr15:67442586..67442601	Bach1	Gliobla	ATGACTCAT	21106904
Footprinting	chr15:67442586..67442601	Bach1	Helas3	ATGACTCAT	21106904
Footprinting	chr15:67442586..67442601	Bach1	Helas3Ifna4h	ATGACTCAT	21106904

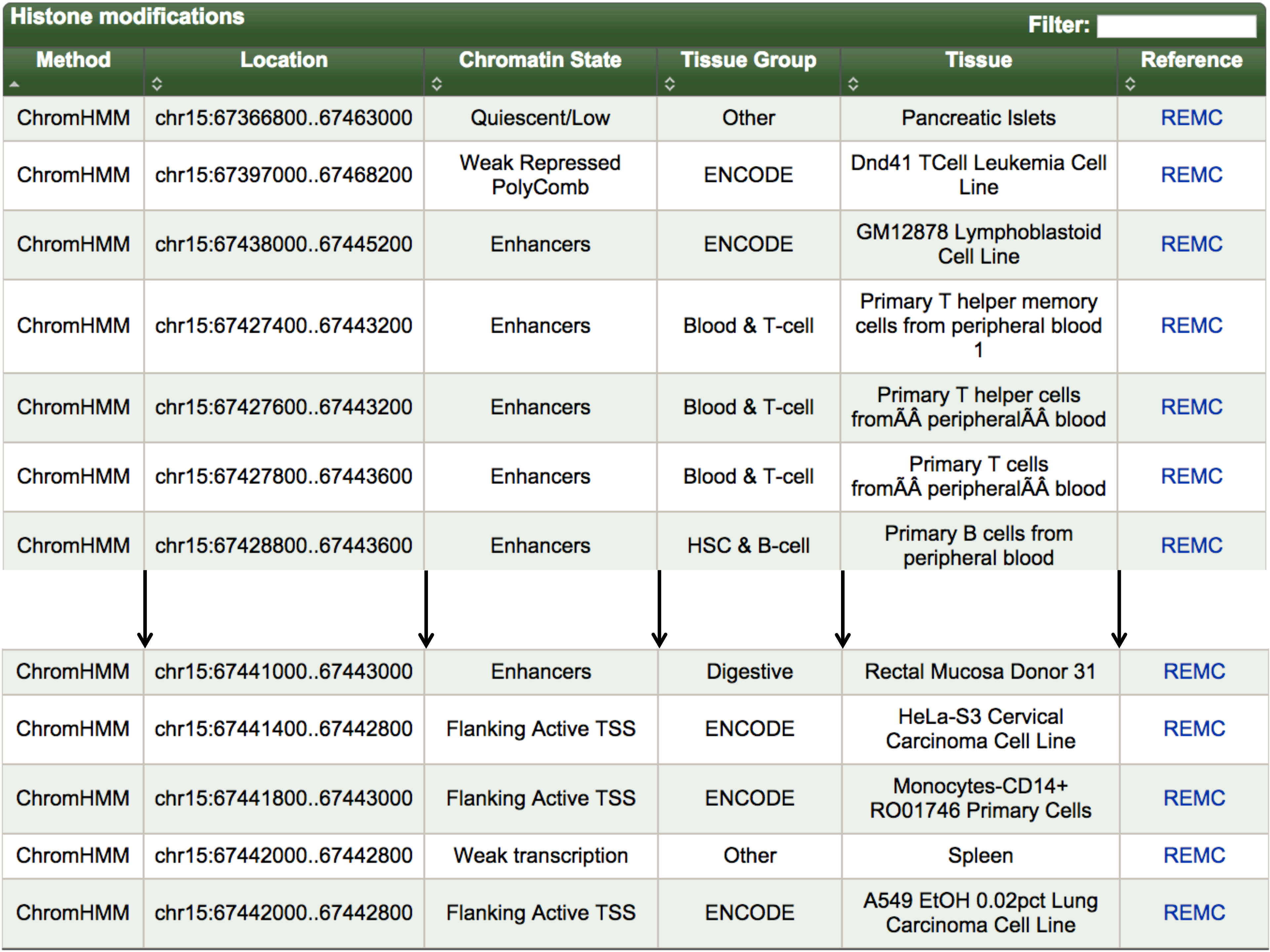
PWM	chr15:67442586..67442602	Jundm2			19443739
PWM	chr15:67442594..67442611	Pou1f1			18585359
PWM	chr15:67442594..67442611	Pou3f1			18585359
PWM	chr15:67442592..67442607	Sox5			19443739
PWM	chr15:67442588..67442599	AP-1			16381825
PWM	chr15:67442588..67442599	AP-1			16381825
PWM	chr15:67442588..67442599	AP-1			16381825
PWM	chr15:67442589..67442597	JDP2			23332764



Chromatin structure

Filter:

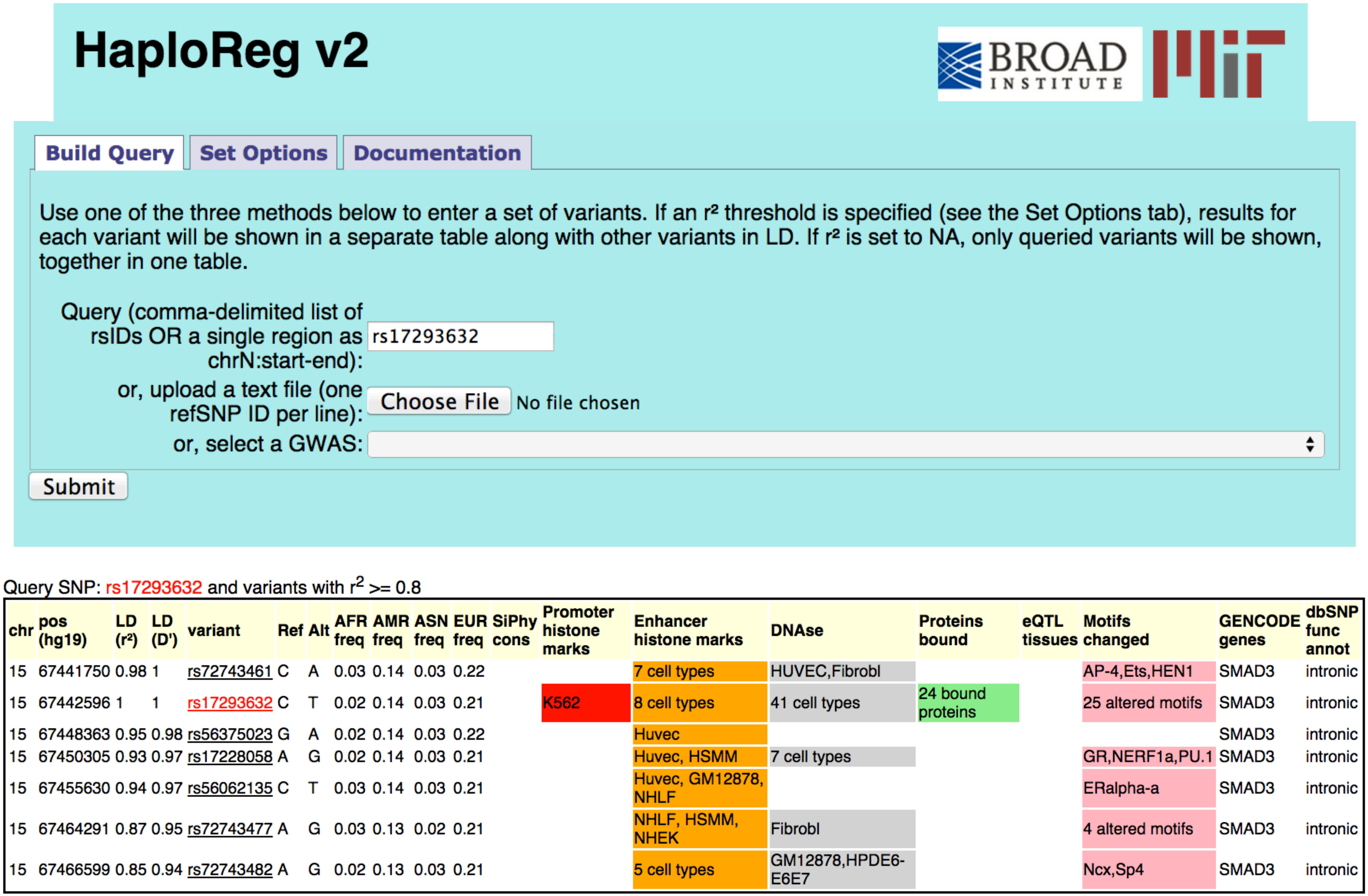
Method	Location	? Cell Type	Additional Info	Reference
DNase-seq	chr15:67442296..67443247	Mcf7	Ctcfshrna	ENCODE
DNase-seq	chr15:67442296..67443247	Mcf7		ENCODE
DNase-seq	chr15:67442298..67443280	A549		ENCODE
DNase-seq	chr15:67442314..67443227	Helas3	Ifna4h	ENCODE
DNase-seq	chr15:67442314..67443227	Helas3		ENCODE
DNase-seq	chr15:67442325..67443240	Mcf7	Randshrna	ENCODE
DNase-seq	chr15:67442347..67443124	Ecc1	Est10nm30m	ENCODE
DNase-seq	chr15:67442351..67443222	Htr8		ENCODE
DNase-seq	chr15:67442579..67443196	Colo829		ENCODE
DNase-seq	chr15:67442392..67443108	Hek293t		ENCODE
FAIRE	chr15:67442282..67443179	Huvec		ENCODE
FAIRE	chr15:67442326..67443079	Helas3	Ifng4h	ENCODE
FAIRE	chr15:67442336..67443114	Helas3	Ifna4h	ENCODE
FAIRE	chr15:67442348..67443091	Hepg2		ENCODE
FAIRE	chr15:67442357..67442606	Helas3		ENCODE
FAIRE	chr15:67442361..67442670	Htr8		ENCODE
FAIRE	chr15:67442486..67443028	K562		ENCODE



Histone modifications

Filter:

Method	Location	Chromatin State	Tissue Group	Tissue	Reference
ChromHMM	chr15:67366800..67463000	Quiescent/Low	Other	Pancreatic Islets	REMC
ChromHMM	chr15:67397000..67468200	Weak Repressed PolyComb	ENCODE	Dnd41 TCell Leukemia Cell Line	REMC
ChromHMM	chr15:67438000..67445200	Enhancers	ENCODE	GM12878 Lymphoblastoid Cell Line	REMC
ChromHMM	chr15:67427400..67443200	Enhancers	Blood & T-cell	Primary T helper memory cells from peripheral blood 1	REMC
ChromHMM	chr15:67427600..67443200	Enhancers	Blood & T-cell	Primary T helper cells fromÃÂ peripheralÃÂ blood	REMC
ChromHMM	chr15:67427800..67443600	Enhancers	Blood & T-cell	Primary T cells fromÃÂ peripheralÃÂ blood	REMC
ChromHMM	chr15:67428800..67443600	Enhancers	HSC & B-cell	Primary B cells from peripheral blood	REMC
ChromHMM	chr15:67441000..67443000	Enhancers	Digestive	Rectal Mucosa Donor 31	REMC
ChromHMM	chr15:67441400..67442800	Flanking Active TSS	ENCODE	HeLa-S3 Cervical Carcinoma Cell Line	REMC
ChromHMM	chr15:67441800..67443000	Flanking Active TSS	ENCODE	Monocytes-CD14+ RO01746 Primary Cells	REMC
ChromHMM	chr15:67442000..67442800	Weak transcription	Other	Spleen	REMC
ChromHMM	chr15:67442000..67442800	Flanking Active TSS	ENCODE	A549 EtOH 0.02pct Lung Carcinoma Cell Line	REMC



Use one of the three methods below to enter a set of variants. If an r² 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 r² is set to NA, only queried variants will be shown, together in one table.

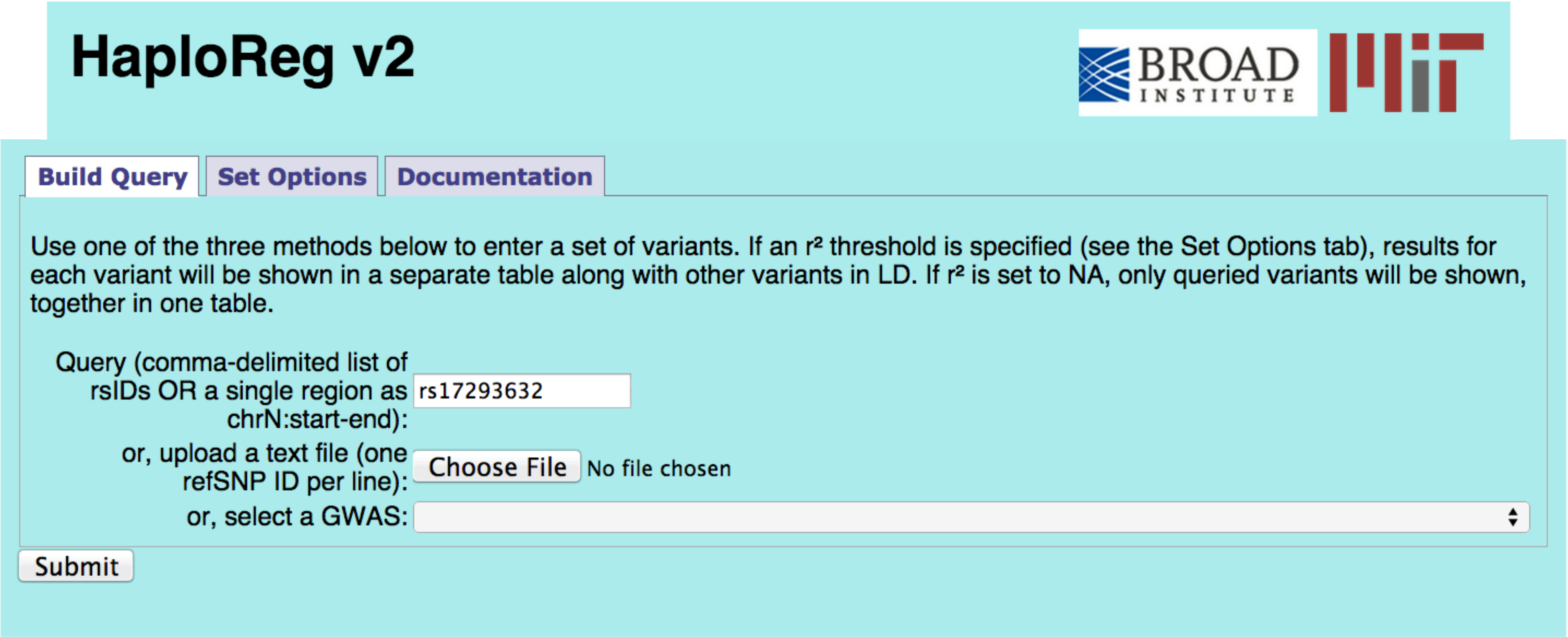
Query (comma-delimited list of rsIDs OR a single region as chrN:start-end): rs17293632

or, upload a text file (one refSNP ID per line): No file chosen

or, select a GWAS:

Query SNP: rs17293632 and variants with r^2 >= 0.8

chr	pos (hg19)	LD (r²)	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
15	67441750	0.98	1	rs72743461	C	A	0.03	0.14	0.03	0.22			7 cell types	HUVEC,Fibrobl			AP-4,Ets,HEN1	SMAD3	intronic
15	67442596	1	1	rs17293632	C	T	0.02	0.14	0.03	0.21		K562	8 cell types	41 cell types	24 bound proteins		25 altered motifs	SMAD3	intronic
15	67448363	0.95	0.98	rs56375023	G	A	0.02	0.14	0.03	0.22			Huvec					SMAD3	intronic
15	67450305	0.93	0.97	rs17228058	A	G	0.02	0.14	0.03	0.21			Huvec, HSMM	7 cell types			GR,NERF1a,PU.1	SMAD3	intronic
15	67455630	0.94	0.97	rs56062135	C	T	0.03	0.14	0.03	0.21			Huvec, GM12878, NHLF				ERalpha-a	SMAD3	intronic
15	67464291	0.87	0.95	rs72743477	A	G	0.03	0.13	0.02	0.21			NHLF, HSMM, NHEK	Fibrobl			4 altered motifs	SMAD3	intronic
15	67466599	0.85	0.94	rs72743482	A	G	0.02	0.13	0.03	0.21			5 cell types	GM12878,HPDE6-E6E7			Ncx,Sp4	SMAD3	intronic



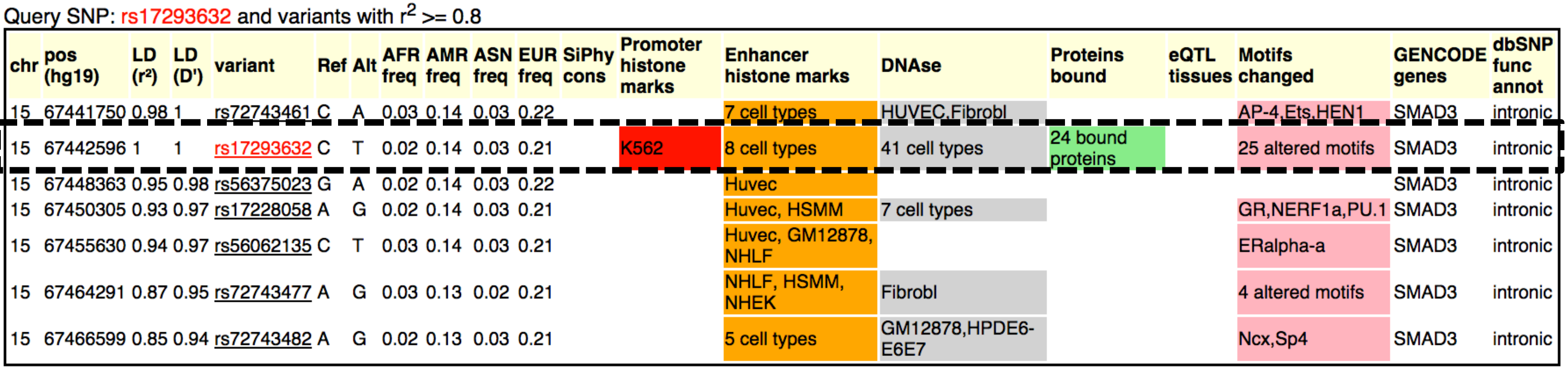
HaploReg v2

Use one of the three methods below to enter a set of variants. If an r² 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 r² 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): rs17293632

or, upload a text file (one refSNP ID per line): No file chosen

or, select a GWAS:



Query SNP: rs17293632 and variants with r^2 >= 0.8

chr	pos (hg19)	LD (r²)	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
15	67441750	0.98	1	rs72743461	C	A	0.03	0.14	0.03	0.22			7 cell types	HUVEC,Fibrobl			AP-4,Ets,HEN1	SMAD3	intronic
15	67442596	1	1	rs17293632	C	T	0.02	0.14	0.03	0.21		K562	8 cell types	41 cell types	24 bound proteins		25 altered motifs	SMAD3	intronic
15	67448363	0.95	0.98	rs56375023	G	A	0.02	0.14	0.03	0.22			Huvec					SMAD3	intronic
15	67450305	0.93	0.97	rs17228058	A	G	0.02	0.14	0.03	0.21			Huvec, HSMM	7 cell types			GR,NERF1a,PU.1	SMAD3	intronic
15	67455630	0.94	0.97	rs56062135	C	T	0.03	0.14	0.03	0.21			Huvec, GM12878, NHLF				ERalpha-a	SMAD3	intronic
15	67464291	0.87	0.95	rs72743477	A	G	0.03	0.13	0.02	0.21			NHLF, HSMM, NHEK	Fibrobl			4 altered motifs	SMAD3	intronic
15	67466599	0.85	0.94	rs72743482	A	G	0.02	0.13	0.03	0.21			5 cell types	GM12878,HPDE6-E6E7			Ncx,Sp4	SMAD3	intronic