ENCODE: Getting Started

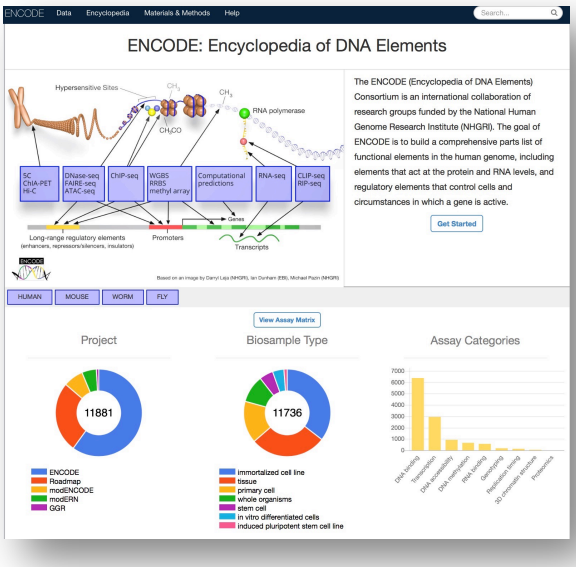
https://www.encodeproject.org; https://www.genome.gov/encode/

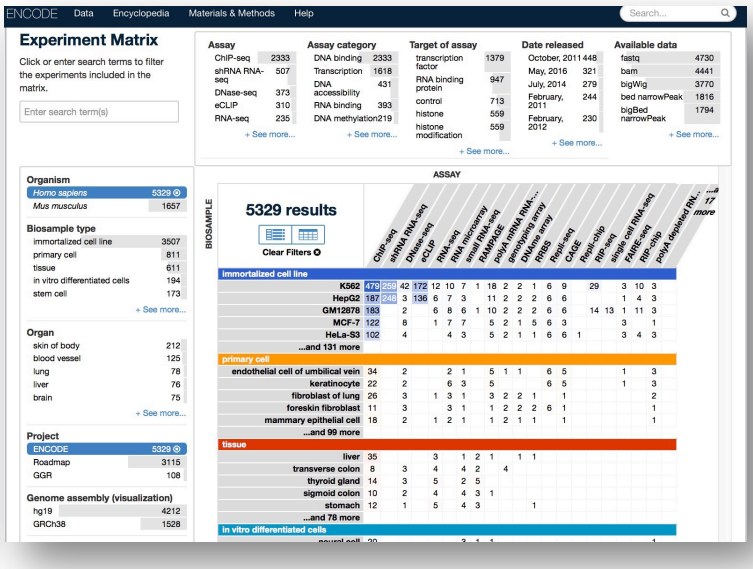
ENCODE Project:

The goals of the NHGRI ENCODE project are to identify all candidate functional elements in the genome, and to make a catalog of those elements freely available

ENCODE Portal (https://encodeproject.org)

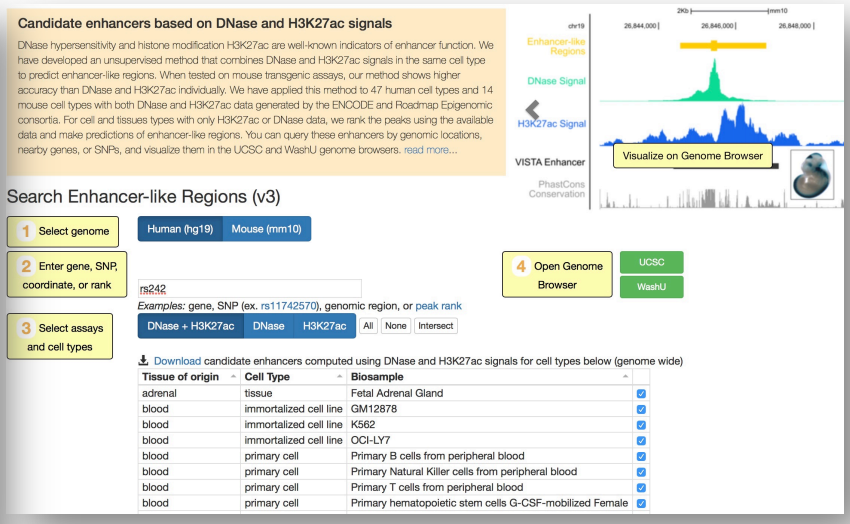
The ENCODE portal has data from over 5000 human and 1500 mouse ENCODE experiments, as well as resources from other projects such as the NIH Common Fund Epigenomics Program. Searches can be refined through facets, and data can be downloaded or displayed through genome browsers. ENCODE metadata is also available programmatically via web services.





ENCODE Encyclopedia (https://www.encodeproject.org/data/annotations/)

The ENCODE Encyclopedia organizes the most salient analysis products into annotations, and provides tools to search and visualize them. Ground level annotations, such as the human/mouse gene expression visualizer shown below left, are typically derived directly from the experimental data. Middle level annotations, such as the human/mouse enhancer visualizer shown below right, integrate multiple types of experimental data and multiple ground level annotations. Top level annotations, such as the tools RegulomeDB, HaploReg, and FunSeq shown on the next page, integrate a broad range of experimental data and ground and middle level annotations.

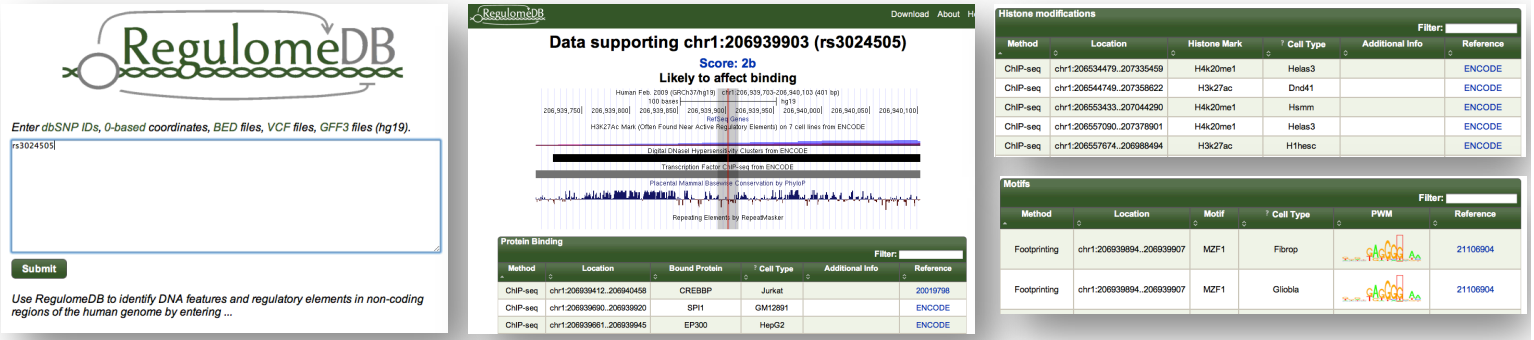


ENCODE: Getting Started (Page 2)

https://www.encodeproject.org; https://www.genome.gov/encode/

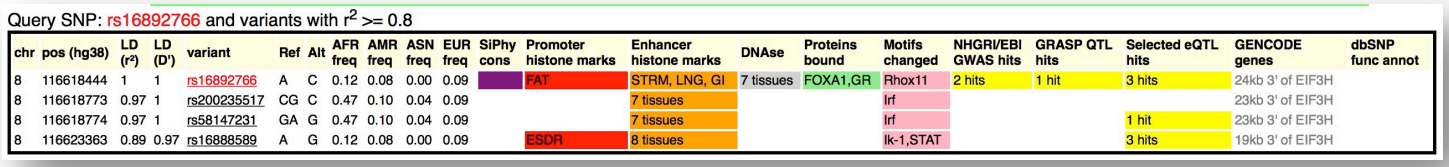
RegulomeDB (http://www.regulomedb.org)

Identifies DNA features and regulatory elements in non-coding regions of the human genome. Uses data from ENCODE, Roadmap Epigenomics, GTEx, GWAS catalog, and other projects. Accepts SNP IDs, genomic coordinates, BED files, VCF files, GFF3 files (hg19) as input, and returns information about the regulatory potential and cell specificity of the input regions. PMID: 22955989



HaploReg (http://archive.broadinstitute.org/mammals/haploreg/haploreg.php)

Explores annotations of the noncoding genome at variants on haplotype blocks, such as candidate regulatory SNPs at disease-associated loci. Uses data from ENCODE, Roadmap Epigenomics, GTEx, GWAS catalog, GENCODE, 1000 Genomes, and other projects. Accepts SNP IDs and genomic coordinates as input, and returns information about the regulatory potential and cell specificity of the input regions. PMID: 26657631



Query SNP: rs16892766 and variants with r^2 >= 0.8

chr	pos (hg38)	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	Motifs changed	NHGRI/EBI GWAS hits	GRASP QTL hits	Selected eQTL hits	GENCODE genes	dbSNP func annot
8	116618444	1	1	rs16892766	A	C	0.12	0.08	0.00	0.09		FAT	STRM, LNG, GI	7 tissues	FOXA1,GR	Rhox11	2 hits	1 hit	3 hits	24kb 3' of EIF3H	
8	116618773	0.97	1	rs200235517	CG	C	0.47	0.10	0.04	0.09			7 tissues			Irf				23kb 3' of EIF3H	
8	116618774	0.97	1	rs58147231	GA	G	0.47	0.10	0.04	0.09			7 tissues			Irf			1 hit	23kb 3' of EIF3H	
8	116623363	0.89	0.97	rs16888589	A	G	0.12	0.08	0.00	0.09		ESDR	8 tissues			Ik-1,STAT			3 hits	19kb 3' of EIF3H	

FunSeq (http://funseq.gersteinlab.org)

Annotates germline and somatic variants, particularly in the noncoding regions of cancer genomes. Uses data from ENCODE, Roadmap Epigenomics, GENCODE, and 1000 Genomes, integrating all the evidence into a single impact score for a variant. Accepts BED or VCF files as input. PMID: 25273974