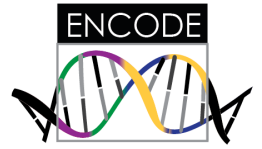


ENCODE: Getting Started

<https://www.encodeproject.org/>

<https://www.genome.gov/encode/>

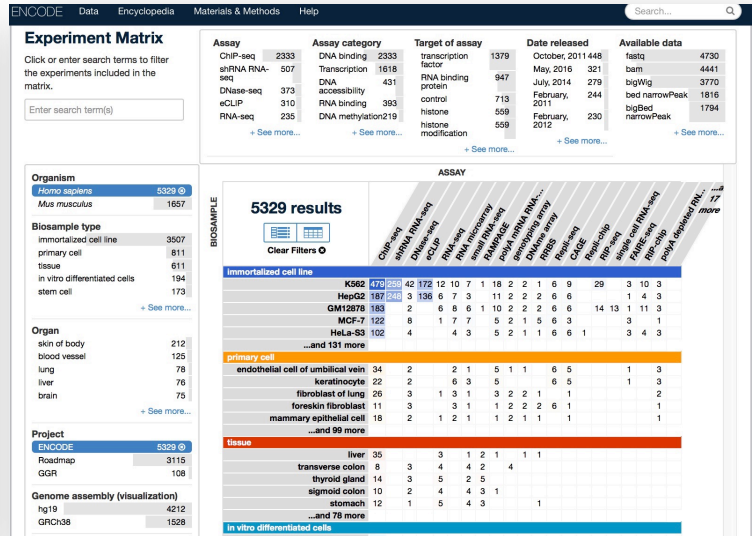
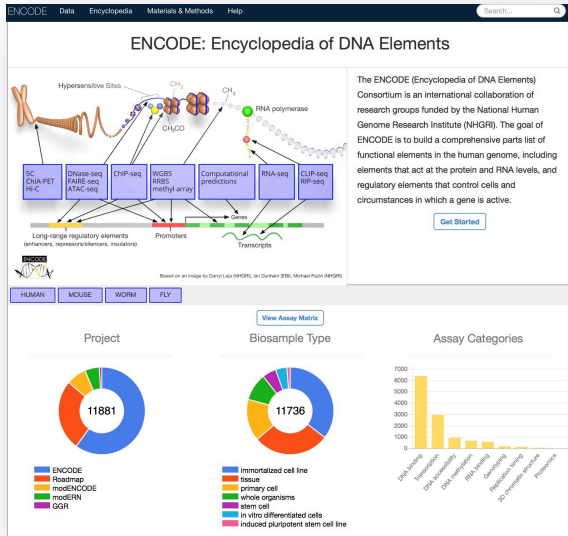


ENCODE Project:

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

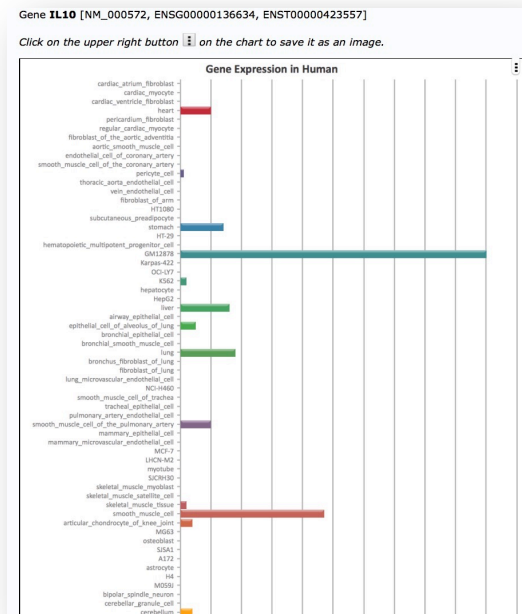
ENCODE Portal (<https://encodeproject.org>)

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.



Candidate enhancers based on DNase and H3K27ac signals

DNase hypersensitivity and histone modification H3K27ac are well-known indicators of enhancer function. We have developed an unsupervised method that combines DNase and H3K27ac signals in the same cell type to predict enhancer-like regions. When tested on mouse transgenic assays, our method shows higher accuracy than DNase and H3K27ac individually. We have applied this method to 47 human cell types and 14 mouse cell types with both DNase and H3K27ac data generated by the ENCODE and Roadmap Epigenomic consortia. For cell and tissue types with only H3K27ac or DNase data, we rank the peaks using the available data and make predictions of enhancer-like regions. You can query these enhancers by genomic locations, nearby genes, or SNPs, and visualize them in the UCSC and WashU genome browsers. [read more...](#)

Search Enhancer-like Regions (v3)

- Select genome: Human (hg19) | Mouse (mm10)
- Enter gene, SNP, coordinate, or rank: rs242
- Select assays and cell types: DNase + H3K27ac | DNase | H3K27ac | All | None | Intersect
- Open Genome Browser: UCSC | WashU

Download candidate enhancers computed using DNase and H3K27ac signals for cell types below (genome wide)

Tissue of origin	Cell Type	Biosample
adrenal	tissue	Fetal Adrenal Gland
blood	immortalized cell line	GM12878
blood	immortalized cell line	K562
blood	immortalized cell line	OCI-LY7
blood	primary cell	Primary B cells from peripheral blood
blood	primary cell	Primary Natural Killer cells from peripheral blood
blood	primary cell	Primary T cells from peripheral blood
blood	primary cell	Primary hematopoietic stem cells G-CSF-mobilized Female

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](#)

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

rs3024503

Submit

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

Data supporting chr1:206939903 (rs3024505)
Score: 2b
Likely to affect binding

Human Feb. 2009 (GRCh37/hg19) chr1:206,939,700-206,940,100 (401 bp)
100 bases hg19

206,939,700 206,939,800 206,939,900 206,939,950 206,940,000 206,940,050 206,940,100

H3K27Ac MAF (One Found Near Active Regulatory Elements) on 7 cell lines from ENCODE

Transcription Factor ChIP-seq from ENCODE

Transcription Factor ChIP-seq from ENCODE

Repeating Elements by RepeatMasker

Protein Binding

Method	Location	Bound Protein	Cell Type	Additional Info	Reference
ChIP-seq	chr1:206939412..206940458	CREBBP	Jurkat		2019798
ChIP-seq	chr1:206939906..206939920	SP1	GM12891		ENCODE
ChIP-seq	chr1:206939861..206939945	EP300	HepG2		ENCODE

Histone modifications

Method	Location	Histone Mark	Cell Type	Additional Info	Reference
ChIP-seq	chr1:206534479..207335459	H4K20me1	HeLaS3		ENCODE
ChIP-seq	chr1:206544749..207335822	H3K27ac	Dnd41		ENCODE
ChIP-seq	chr1:206553433..207044290	H4K20me1	Hsmm		ENCODE
ChIP-seq	chr1:206557090..207378901	H4K20me1	HeLaS3		ENCODE
ChIP-seq	chr1:206557674..206968494	H3K27ac	H1hesc		ENCODE

Motifs

Method	Location	Motif	Cell Type	PWM	Reference
Footprinting	chr1:206939894..206939907	MZF1	Fibrop		21106904
Footprinting	chr1:206939894..206939907	MZF1	Globla		21106904

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 $r^2 \geq 0.8$

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	NHGR/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						Irf			1 hit	23kb 3' of EIF3H	
8	116618774	0.97	1	rs58147231	GA	G	0.47	0.10	0.04	0.09						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](#)

FunSeq2 - A flexible framework to prioritize regulatory mutations from cancer genome sequencing

Analysis Results Downloads Documentation FAQ

Overview

This tool is specialized to prioritize somatic variants from cancer whole genome sequencing. It contains two components: 1) building data context from various resources; 2) variants prioritization. We provided downloadable scripts for users to customize the data context (found under 'Downloads'). The variants prioritization step is downloadable, and also implemented as web server (Right Panel), with pre-processed data context.

Instructions

- Input File - BED or VCF formatted. Click "green" button to add multiple files. With multiple files, the tool will do recurrent analysis. (Note: for BED format, user can put variants from multiple genomes in one file, see [Sample input file](#).)
- Recurrence DB - User can choose particular cancer type from the database. The DB will continue to be updated with newly available WGS data.
- Gene List - Option to analyze variants associated with particular set of genes. Note: Please use Gene Symbols, one row per gene.
- Differential Gene Expression Analysis - Option to detect differentially expressed genes in RNA-Seq data. Two files needed: expression file & class label file. Please refer to [Expression input files](#) for instructions to prepare those files.

Note: In addition to on-site calculation, we also provide scores for all possible noncoding SNVs of GRCh37/hg19 under 'Downloads' (without annotation and recurrence analysis).

Input File: (only for hg19 SNVs)
Choose File: No file chosen
BED or VCF files as input. Sample input file

Output Format:
bed

MAF:
0
Minor allele frequency threshold to filter polymorphisms from 1KG (value 0-1)

Cancer Type from Recurrence DB: Summary table
All Cancer Types

Add a gene list (Optional)
Add differential gene expression analysis (Optional)

Upload

