



Using ENCODE Data To Interpret Disease-associated Genetic Variation

Mike Pazin

National Human Genome Research Institute, NIH

ENCODE Users Meeting

June 8, 2016





Welcome



National Human Genome
Research Institute

- Objectives
 - We want to tell the community about the ENCODE resource
 - We want to hear community experiences and suggestions



Elise Feingold



Dan Gilchrist



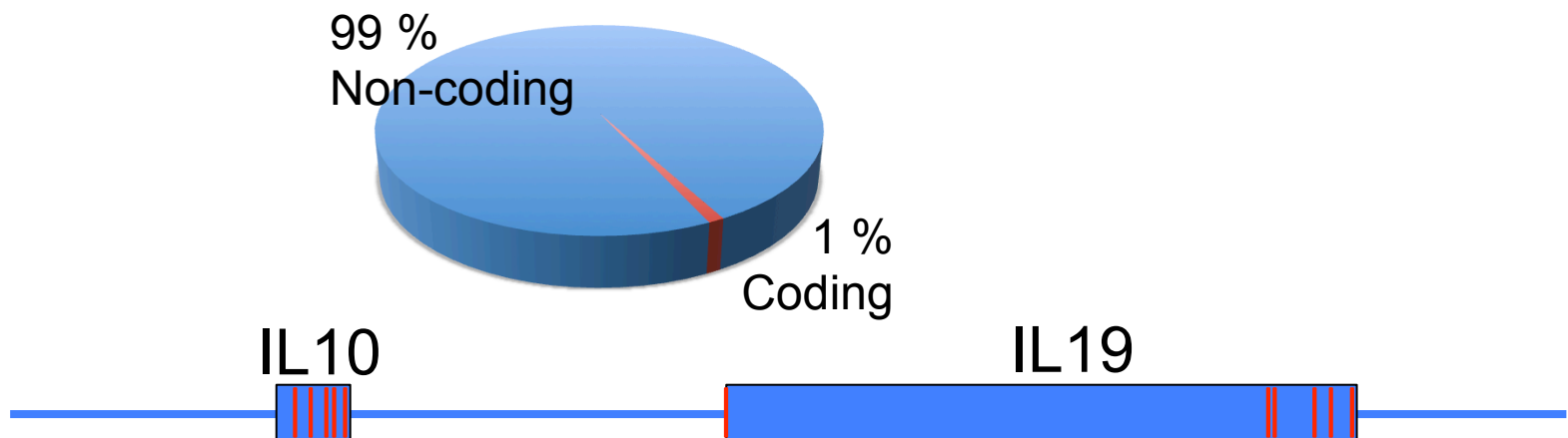
Overview

- The ENCODE Resource
- Use of ENCODE to illuminate the role of genetic variation in human disease
- Accessing ENCODE materials



Reading The Human Genome Is Difficult

- Genetic code very powerful for 1% of the human genome
 - No correspondingly powerful regulatory code
 - Sequence conservation can identify some candidate functional elements (but not when or where they act)
 - Regulatory regions aren't always in the same order as gene targets
- Need unbiased experimental investigation





Non-coding DNA Is Important For Disease And Gene Regulation

- Vast majority of common disease associations and heritability lie outside of protein-coding regions
- Non-coding DNA variants are known to cause human diseases and alter human traits (FXS, ALS)

Functional information is needed to interpret the role of genetic variation in human disease, and to apply genomics in the clinic.

PMID: 22955828, PMID: 25439723, PMID: 23128226
PMID: 17477822, PMID: 25679767





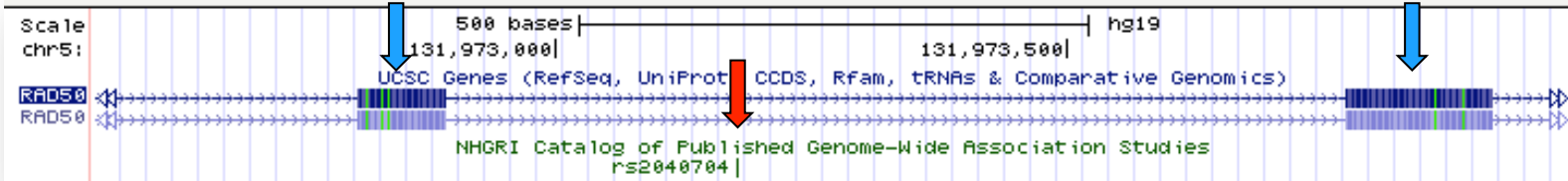
1,500 Letters Of Our 3 Billion Letter Genome

agccaagcagcaaagtttgctgctgttatTTTTgtagctctactatattctacttttaccattgaaaatattgaggaagtatt
tatatttctatTTTTatataattataattttatgtattttaataactattacacataattatTTTTatataatgaagtaccaatg
acttcttttccagagcaataatgaaatttcacagtatgaaaatggaagaaatcaataaaattatacgtgacctgtggcgaa
gtacctatcgtggacaaggtgagtaccatgggtgatcacaatgctcttccaaagccctctccgcagctcttccccttatga
cctctcatcatgccagcattacctcctggacccttctaagcatgtctttgagattttctaagaattcttatcttggcaacatc
ttgtagcaagaaaatgtaaagtttctgttccagagcctaacaggacttacatattgactgcagtaggcattatatttagctg
atgacataataggttctgtcatagtgtagatagggataagccaaaatgcaataagaaaaacctccagagggaaactctttt
TTTTcttttcttttttttttccagatggagtctcgcacttctctgtcaccgggctggagcgcagtggtgcaatcttggctca
ctgcaacctccacctcctgggttcaggtgattctcccacctcagcctcccgagtagtagctggaattacaggtgcgcgctccc
acacctggctaattTTTTgtattcttagtagagatggggtttcccatgttggccaggctggtctcaaactcctgccctcaggtg
atctgccaccttggcctcccagtggttgggttacaggcgtgagccaccgcgctggcctggaggaaactcttaacagggaa
actaagaaagagttgaggctgaggaactggggcatctgggttgcttctggccagaccaccaggctcttgaatcctcccagc
cagagaaagagttccacaccagccattgtttcctctggtaatgtcagcctcatctgttgttctaggcttacttgatatgtttg
taaatagacaaaaggctacagagcataggttctctaaaatattcttcttctgtgtcagatattgaatacatagaaatcggg
ctgatgccgatgaaaatgtatcagcttctgataaaaggcgggaattataactaccgagtggtgatgctgaaggagacacag
ccttggatatgcgaggacgatgcagtgctggacaaaaggcaggtatctcaaaagcctggggagccaactcacccaagtaa
ctgaaagagagaaacaaacatcagtgagtggaagcaccaaggctacacctgaatggtgggaagctctttgctgctata
taaataaatcaggctcagctactattattacactctctgaagctaaccaacatttctgcaacattatgtagactt



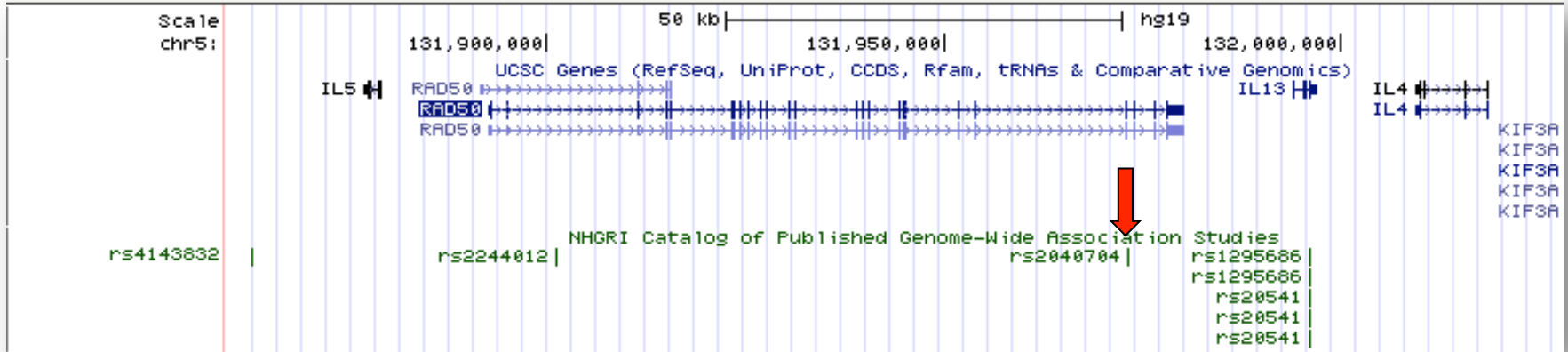
Maps And Annotation Help Us To Understand The Sequence

agccaagcagcaaagtttggctgctgttatTTTTgtagcttactatattctactttaccattgaaaatattgaggaagtatt
 tatatttctatttttatatattataattttatgtattttaataactattacacataattttttatataatgaagtaccaatg
 acttcctttccagagcaataatgaaatttcacagtatgaaaatggaagaaatcaataaaattatacgtgacctgtggcgaa
 gtacctatcgtggacaaggtgagtaccatgggtgatcacaatgctctttccaaagccctctccgcagctctccccttatga
 cctctcatcatgccagcattacctcctggaccctttctaagcatgtctttgagattttctaagaattcttatctggcaacatc
 ttgtagcaagaaaatgtaaagtttctgttccagagcctaacaggacttacatattgactgcagtaggcattatatttagctg
 atgacataataggttctgtcatagtgtagatagggataagccaaatgcaataagaaaaacctccagaggaaactctttt
 tttttctttttctttttttttccagatggagtctcgcacttctctgtcaccgggctggagcgcagtggtgcaatcttggctca
 ctgcaacctccacctcctgggttcaggtgattctccacctcagcctcccagtagtagctggaattacaggtgcgcgctccc
 acacctggctaatttttgtattcttagtagagatggggtttcccatgttggccaggctggtctcaaactcctgccctcaggtg
 atctgccaccttggcctcccagtggttgggttacaggcgtgagccaccgcgctggcctggaggaaactcttaacagggaa
 actaagaaagagttgaggctgaggaactggggcatctgggttgcttctggccagaccaccaggctcttgaatcctcccagc
 cagagaaagagttccacaccagccattgtttcctctggtaatgtcagcctcatctgttgttcttaggcttacttgatatgttg
 taaatgacaaaaggctacagagcataggttcctctaaaatattcttcttctgtgtcagatattgaatacatagaaatcggg
 ctgatgccgatgaaaatgtatcagcttctgataaaaggcgggaattataactaccgagtggtgatgctgaagggagacacag
 cttggatatgagaggacgatgcagtctggacaaaaggcaggtatctcaaagcctggggagccaactcacccaagtaa
 ctgaaagagagaaacaaacatcagtgagtggaagcaccaaggctacacctgaatggtggaagctctttgctgctata
 taaatgaatcaggctcagctactattattacactctctgaagctaaccaacatttctgcaacattatgtagactt



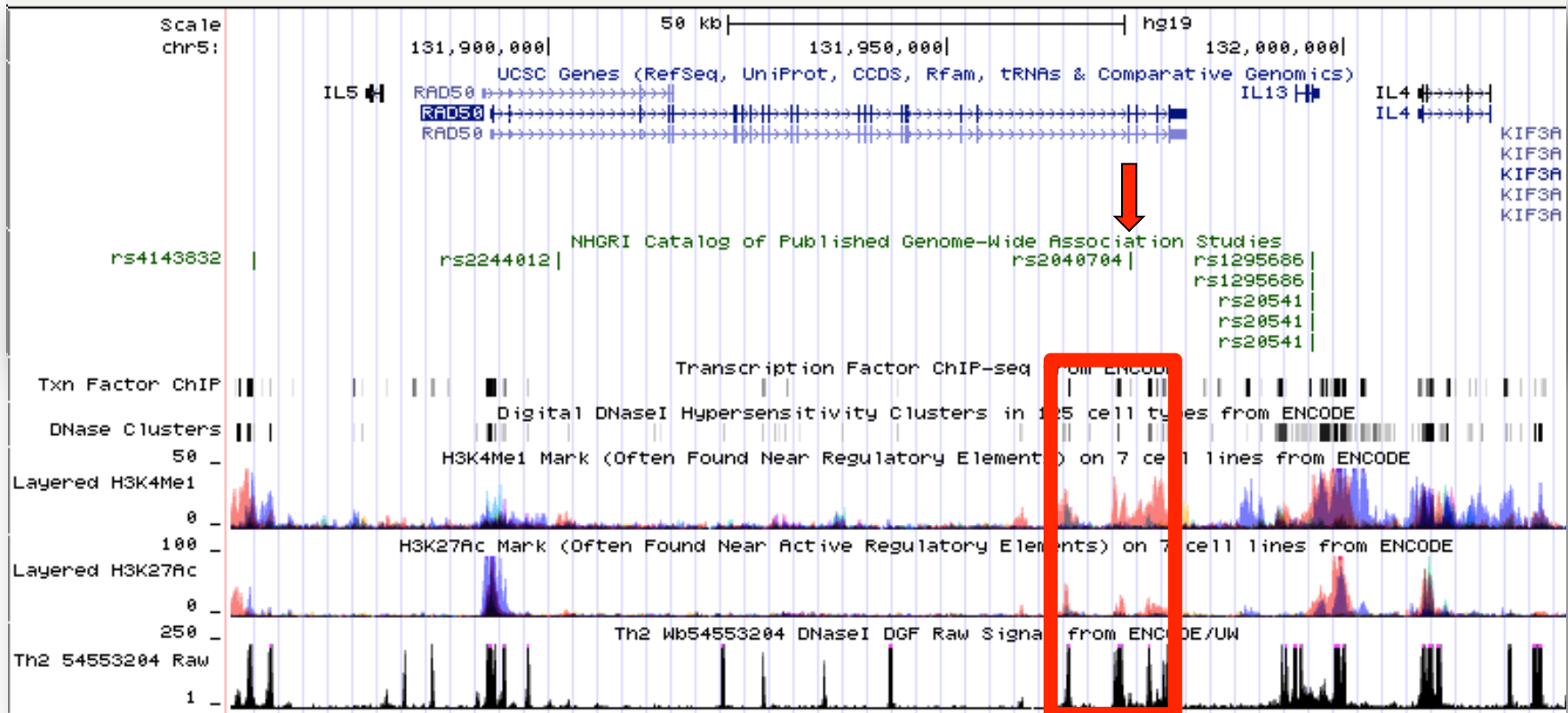


Richer Maps Provide More Information



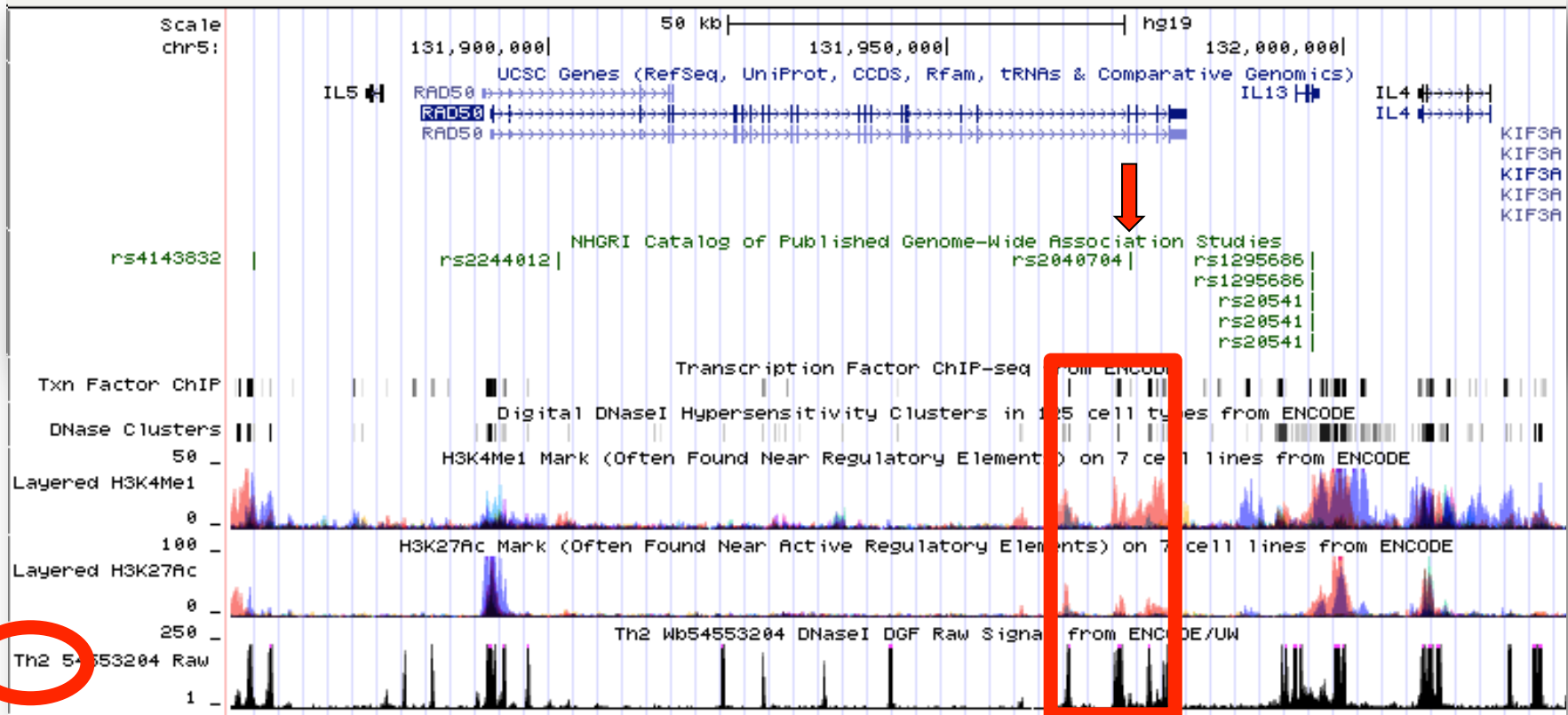


Richer Maps Provide More Information





Richer Maps Provide More Information





Richer Maps Provide More Information



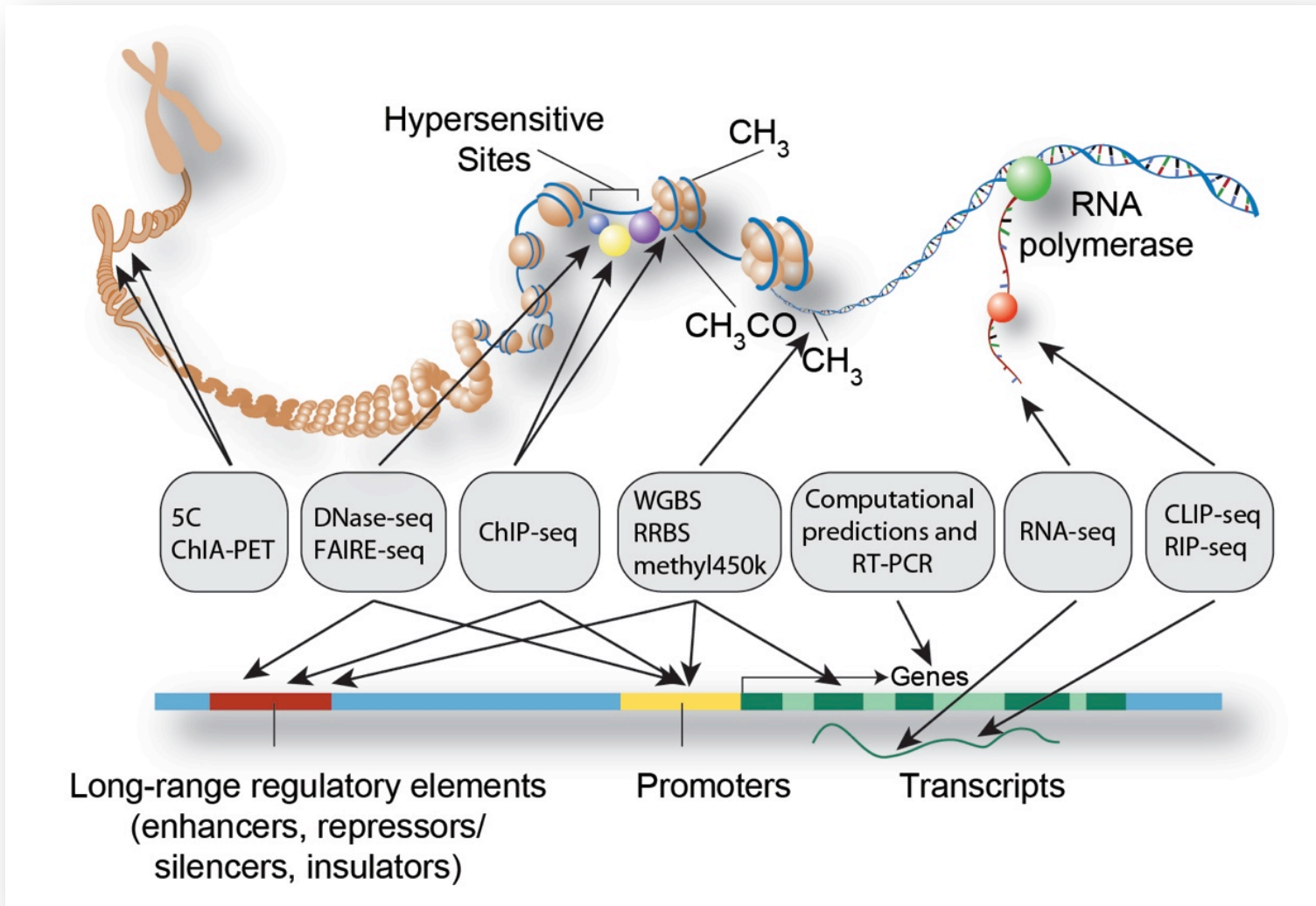


ENCODE: Encyclopedia Of DNA Elements

- Identify all candidate functional elements in the genome
- Make resource freely available to community
 - genetic basis of disease
 - gene regulation



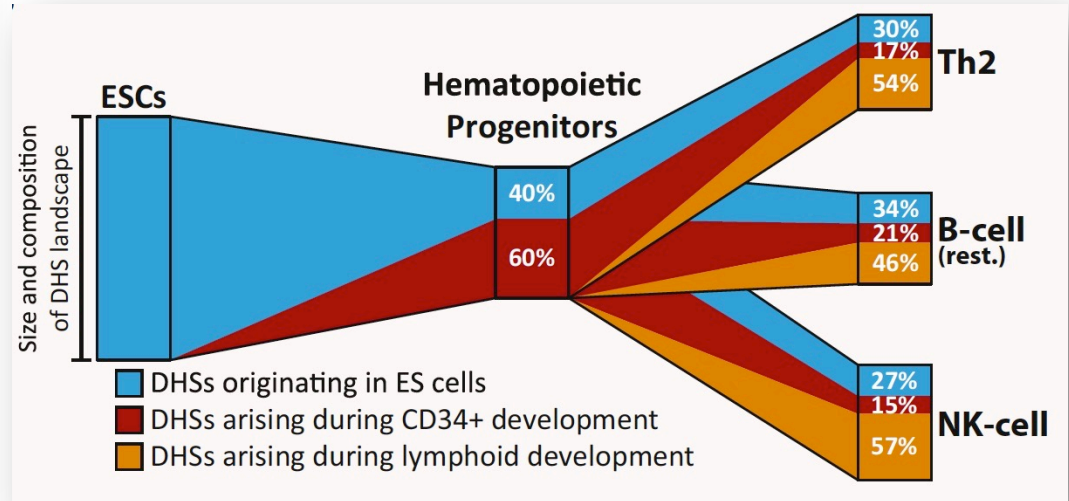
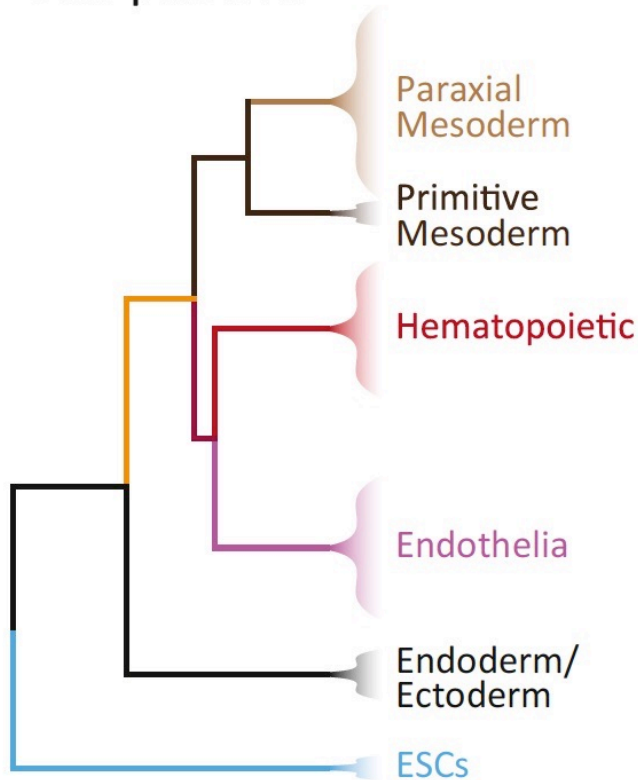
ENCODE Data Types





ENCODE Data Are Cell-Type Specific

Clustering based on DHS patterns





ENCODE Accomplishments

- Sharing 1000s of datasets
 - No embargo
 - Unrestricted access
 - High quality
 - Uniformly processed
- Sharing software
- Data interoperability

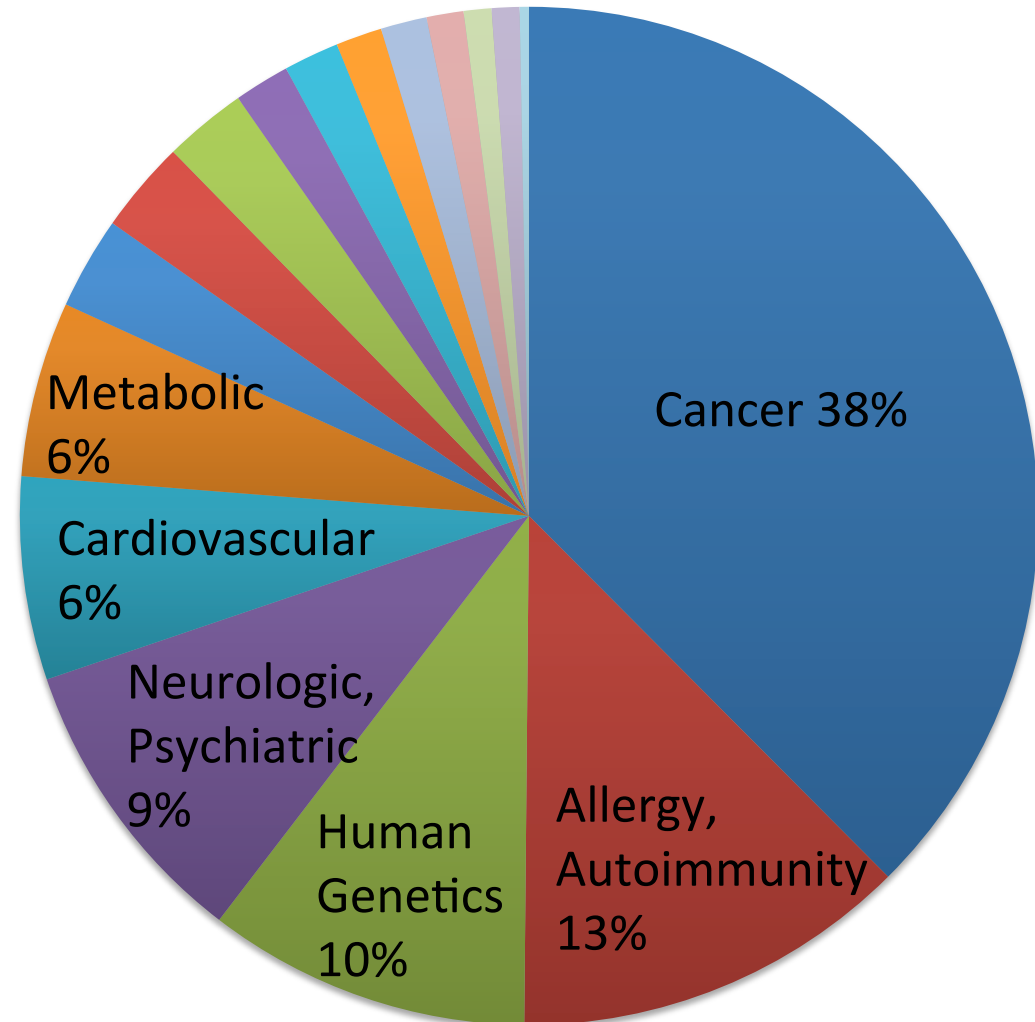


Publications Using ENCODE Data

Hundreds of Consortium publications

~1500 community publications using ENCODE data:

~675 Human Disease
~600 Basic Biology
~225 Methods/Software Development





Summary- ENCODE Resource

- Freely shared catalog of genomic data and candidate genomic functional elements
- ENCODE is built upon established techniques and interpretations developed for the study of gene regulation
- ENCODE maps can be used to make predictions about genome function



Overview

- The ENCODE Resource
- Use of ENCODE to illuminate the role of genetic variation in human disease
- Accessing ENCODE materials



Standard ENCODE Use Cases: Hypothesis Generation

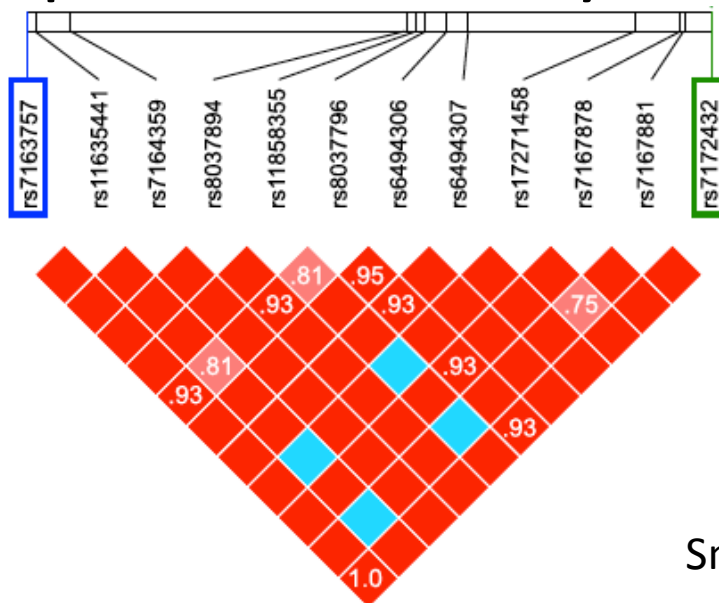
Major use: Hypothesis generation and refinement

- Prediction of causal variants/regulatory elements
- Prediction of target genes
- Prediction of target cell types
- Prediction of upstream regulators



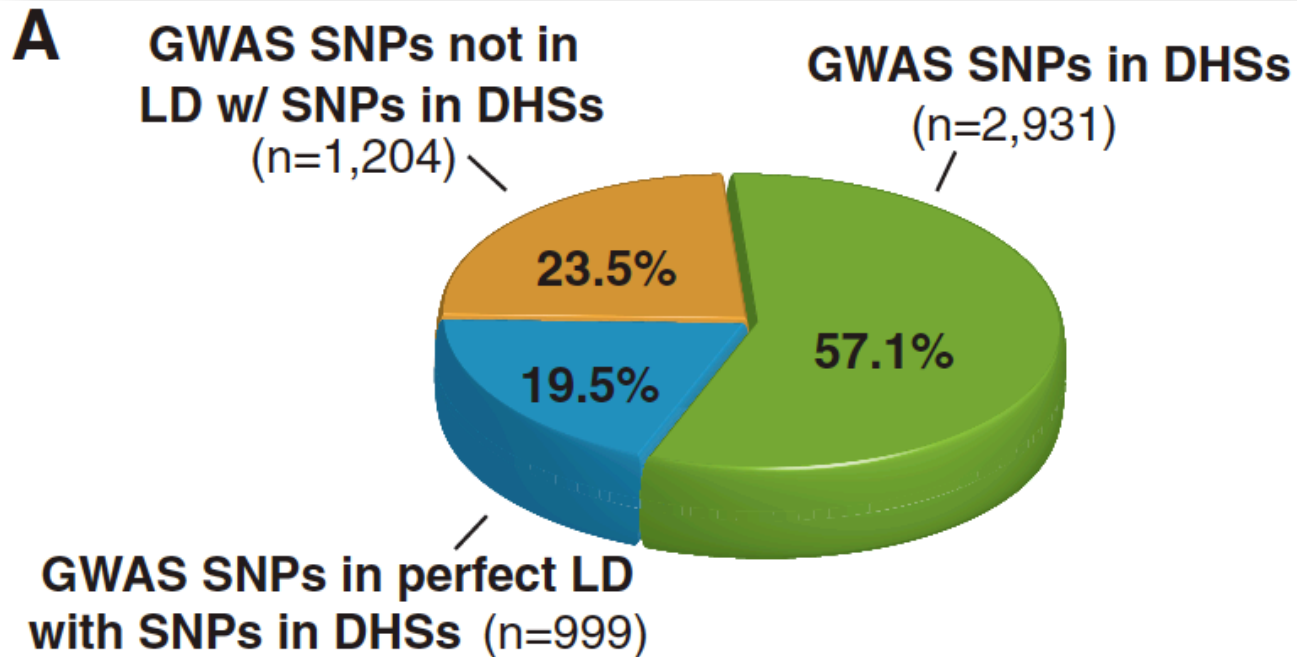
Prediction of Causal Variants

- Multiple variants may be in linkage disequilibrium
- The causal variant may not have been tested during data collection
- Multiple variants may be causal





Many GWAS Associations Lie In Regions Annotated By ENCODE And CF Epigenomics Data





ENCODE/Epigenomics Data From HaploReg

HaploReg v4.1



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 chromatin state and protein binding annotation from the Roadmap Epigenomics and ENCODE projects, sequence conservation across mammals, the effect of SNPs on regulatory motifs, and the effect of SNPs on expression from eQTL studies. HaploReg is designed for researchers developing mechanistic hypotheses of the impact of non-coding variants on clinical phenotypes and normal variation.

Update 2015.11.05: Version 4.1 GWAS and eQTL have been updated; a simpler pruning strategy is applied when combining GWAS; and links out to other NHGRI/EBI GWAS hits and GRASP QTL hits are provided.

Update 2015.09.15: Version 4.0 now includes many recent eQTL results including the GTEx pilot, four different options for defining enhancers using Roadmap Epigenomics data, and a complete set of source files for download and local analysis. Older versions available: [v3](#), [v2](#), [v1](#).

[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): **1**
 or, upload a text file (one refSNP ID per line): no file selected

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 | 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 | | | lrf | | | | 23kb 3' of EIF3H | |
| 8 | 116618774 | 0.97 | 1 | rs58147231 | GA | G | 0.47 | 0.10 | 0.04 | 0.09 | | | 7 tissues | | | lrf | | | 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 | |

www.broadinstitute.org/mammals/haploreg/



ENCODE Data From RegulomeDB

RegulomeDB

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

rs16892766 ← 1

Submit ← 2

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

Summary of SNP analysis

Show 10 entries

| Coordinate (0-based) | dbSNP ID | Regulome DB Score |
|----------------------|------------|-------------------|
| chr8:117630682 | rs16892766 | 2b ← 3 |

Showing 1 to 1 of 1 entries

Download BED GFF Full Output

A project of the Center for Genomics and Personalized Medicine at

RegulomeDB (TM) Copyright ©2011 The Board of Trustees of Leland Stanford Junior University. Permission to use the information contained in this database information. Users of the database are solely responsible for compliance with any copyright restrictions, including those applying to the author abstracts, expressed or implied. The RegulomeDB project at Stanford University is supported by a Genome Research Resource Grant from the US National Human G

Protein Binding

| Method | Location | Bound Protein | Cell Type | Additional Info | Reference |
|----------|---------------------------|---------------|-----------|-----------------|-----------|
| ChIP-seq | chr8:117630539..117630739 | FOXA1 | ECC-1 | DMSO_0.02pct | ENCODE |
| ChIP-seq | chr8:117630626..117630842 | NR3C1 | ECC-1 | DEX_100nM | ENCODE |

Chromatin structure

| Method | Location | Cell Type | Additional Info | Reference |
|-----------|---------------------------|-----------|-----------------|-----------|
| DNase-seq | chr8:117630480..117630690 | Rptec | | ENCODE |
| DNase-seq | chr8:117630480..117630730 | Nhlf | | ENCODE |
| DNase-seq | chr8:117630500..117630710 | Nha | | ENCODE |
| DNase-seq | chr8:117630500..117630770 | Hah | | ENCODE |
| DNase-seq | chr8:117630510..117630704 | Aosmc | Serum | ENCODE |
| DNase-seq | chr8:117630520..117630790 | Hvmf | | ENCODE |
| DNase-seq | chr8:117630625..117631002 | Htr8 | | ENCODE |
| FAIRE | chr8:117630519..117630761 | Medullo | | ENCODE |

Histone modifications

| Method | Location | Histone Mark | Cell Type | Additional Info | Reference |
|----------|---------------------------|--------------|-----------|-----------------|-----------|
| ChIP-seq | chr8:110578383..117647033 | H3k09me3 | Dnd41 | | ENCODE |
| ChIP-seq | chr8:116009496..120997897 | H2az | Hepg2 | | ENCODE |
| ChIP-seq | chr8:117409399..118413945 | H4k20me1 | Hmec | | ENCODE |
| ChIP-seq | chr8:117555446..118475451 | H4k20me1 | Nhlf | | ENCODE |
| ChIP-seq | chr8:117384499..117650386 | H2az | Dnd41 | | ENCODE |

<http://regulomedb.org/>



ENCODE cis-element Browser

PENNSSTATE
1855

Welcome to YUE Lab
Computational and Functional Genomics/Epigenomics

HOME MOUSE HUMAN DOWNLOAD LINKS CONTACT

Query human ENCODE data!

Option 1: Search gene expression across ~ 60 human cell types (total 108 datasets)

human (hg19) Gene name(Sox2, Nanog ...) submit!

Option 2: Search cis-elements in a given genomic region

human (hg19) chr8 start: 128390000 end: 128410000 submit!

Option 3: search cis-elements surrounding a gene

human (hg19) Gene name(Sox2, Nanog ...) submit!


Extended region (default +/- 100kb) kb submit!




ENCODE cis-element Browser

Candidate cis-elements in your queried region.

Human (hg19)
chr8:128390000-128410000

DNaseI Hypersensitive Sites: 

| Coordinate | Tissue/cell type |
|--------------------------|---|
| chr8:128394860-128395010 | NHDF-Ad |
| chr8:128395580-128395730 | HSMMtube,HSMM |
| chr8:128398205-128398355 | Osteobl |
| chr8:128398585-128398735 | Osteobl |
| chr8:128399500-128399650 | GM12878,NHDF-Ad,HSMM |
| chr8:128400960-128401110 | HSMM,HSMMtube |
| chr8:128402480-128402630 | HSMM |
| chr8:128403580-128403730 | HMEC,Osteobl,HSMM,NHDF-Ad,HSMMtube,NH-A,HeLa-S3,NHEK,NHLF |
| chr8:128404560-128404710 | HMEC |
| chr8:128404720-128404870 | HSMM |
| chr8:128405400-128405550 | HSMM |
| chr8:128407420-128407570 | HeLa-S3 |
| chr8:128407885-128408035 | HUVEC,Osteobl,NHDF-Ad |
| chr8:128408160-128408310 | HMEC |

TF binding Site: 

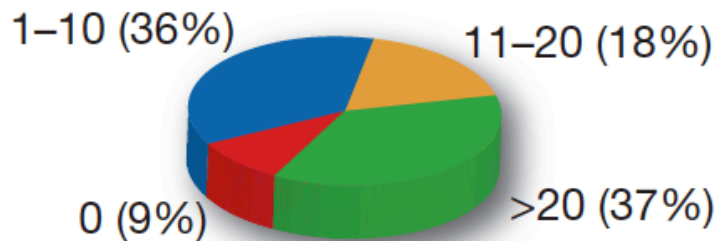
| Coordinate | TF | tissue |
|--------------------------|----------------|--|
| chr8:128398585-128398735 | USF1 | USF1(K562), USF1(SK-N-SH_RA) |
| chr8:128399500-128399650 | RUNX3, SPI1 | RUNX3(GM12878), SPI1(GM12878), SPI1(GM12891) |
| chr8:128403580-128403730 | multiple | CEBPB(HeLa-S3), CEBPB(IMR90), EP300(HeLa-S3), FOS(MCF10A-Er-Src), FOXA1(A549), GATA3(T-47D), JUN(HeLa-S3), JUND(HeLa-S3), MAX(HeLa-S3), MYC(MCF10A-Er-Src), NR3C1(A549), POLR2A(HeLa-S3), POLR2A(MCF10A-Er-Src), RCOR1(HeLa-S3), SMC3(HeLa-S3), STAT3(HeLa-S3), STAT3(MCF10A-Er- |



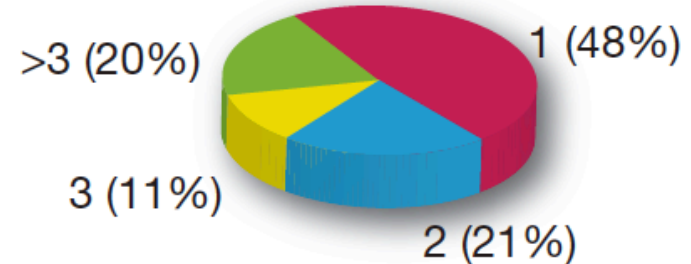
Prediction of Target Genes

- Regulatory regions can operate on multiple, distal genes
- The target gene could be a non-coding RNA

Distal DHSs connected
per promoter DHS
($n = 69,965$)

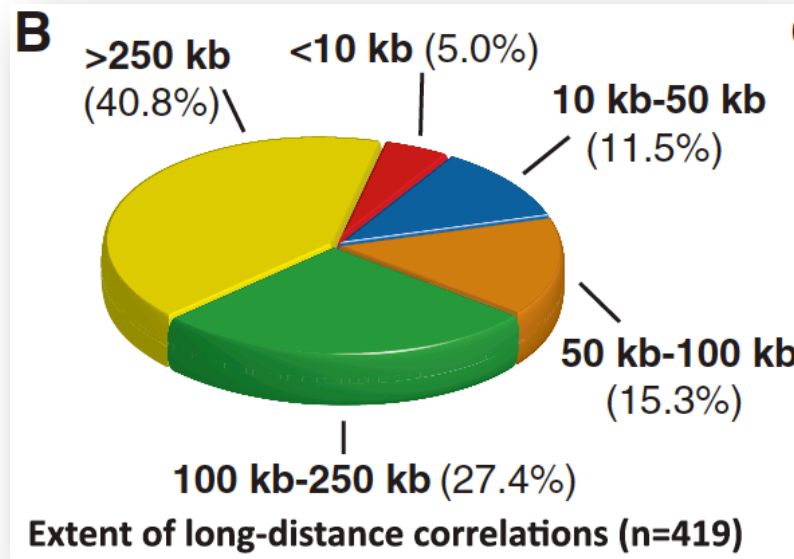


Promoter DHSs connected
per distal DHS
($n = 578,905$ of 1,454,901 total)



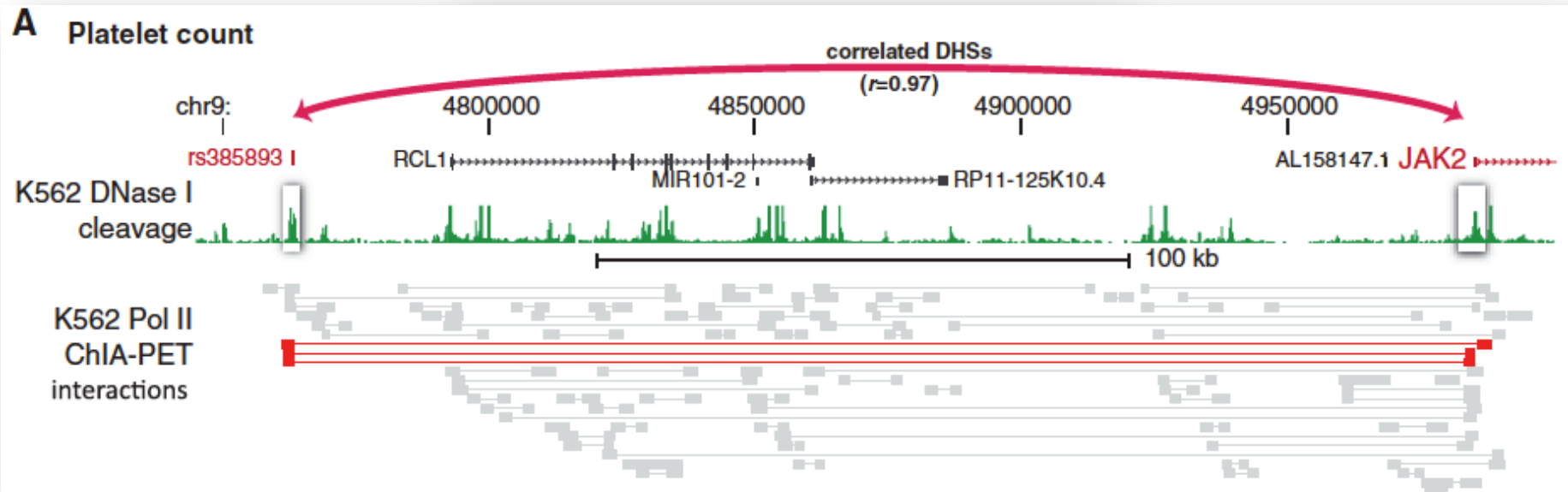


Many GWAS Associations Are Predicted To Be Linked To Distal Genes





Many GWAS Associations Are Predicted To Be Linked To Distal Genes





Prediction of Linkage Between Regulatory Elements and Genes

Regulatory Elements Database

This database provides a user interface to the results of the analysis presented in Sheffield et al. (2013). This is a signal in 112 human samples. Questions? contact [Nathan Sheffield](#). See also: [Supplemental Files](#)

There are 5 ways to explore the data:

- By **CELLTYPE** - Select cell-types to include or exclude.
- By **CLUSTER** - View promoter, CpG-island, and conserved element overlap for all clusters, and select individual clusters
 - Muscle-specific cluster: [cluster 1520](#)
 - Prostate and hepatocyte cluster: [cluster 910](#)
 - Prostate-only cluster: [cluster 2483](#)
 - Hematopoietic cluster: [cluster 25](#)
 - Pluripotency cluster: [cluster 104](#)
- By **GENE** - Search by gene of interest
 - [MyoG](#)
 - [RFX1](#)
 - [LIN28A](#)
 - [HBG1](#)
- By **COORDINATE**: Give chr, start, stop to find all regulatory elements in a region.
 - IRF2 regulator: [chr4: 185240845-185240995](#)
 - MyoD1 regulator: [chr11: 17828545-17828695](#)
 - Blood regulator: [chr3: 128166420-128166570](#)
- By **FACTOR**: find a specific TF of interest
 - [CTCF](#)
 - [Myf](#)
 - [AP1](#)
 - [Pou5f1 \(Oct-4\)](#)

<http://dnase.genome.duke.edu>



Prediction of Linkage Between Regulatory Elements and Genes

Regulatory Elements Database

Chromosome, start, stop:

5: 131960000-132000000

23 DHS sites found

| Site ID | Location | Cluster | Connections |
|---------|---------------------------|---------|---------------------------------------|
| 2174568 | chr5: 131988965-131989115 | 2001 | (0) - |
| 2174546 | chr5: 131967940-131968090 | 2110 | (1) - ENSG00000131437 |
| 2174547 | chr5: 131968265-131968415 | 1605 | (1) - ENSG00000169194 |
| 2174548 | chr5: 131970980-131971130 | 2072 | (1) - ENSG00000169194 |
| 2174550 | chr5: 131972960-131973110 | 2072 | (1) - ENSG00000169194 |
| 2174556 | chr5: 131977240-131977390 | 2168 | (1) - ENSG00000113520 |
| 2174557 | chr5: 131977640-131977790 | 2072 | (2) - ENSG00000169194,ENSG00000223442 |
| 2174560 | chr5: 131982700-131982850 | 860 | (1) - ENSG00000223442 |
| 2174564 | chr5: 131985020-131985170 | 1605 | (1) - ENSG00000169194 |
| 2174577 | chr5: 131992180-131992330 | 1259 | (1) - ENSG00000113520 |
| 2174578 | chr5: 131992545-131992695 | 1078 | (2) - ENSG00000223442,ENSG00000113520 |
| 2174582 | chr5: 131993765-131993915 | 1970 | (1) - ENSG00000113520 |
| 2174584 | chr5: 131994205-131994355 | 1970 | (1) - ENSG00000223442 |
| 2174587 | chr5: 131995065-131995215 | 1704 | (1) - ENSG00000113520 |
| 2174590 | chr5: 131995740-131995890 | 369 | (1) - ENSG00000113520 |
| 2174595 | chr5: 131997040-131997190 | 715 | (1) - ENSG00000113522 |
| 2174596 | chr5: 131997225-131997375 | 514 | (1) - ENSG00000113522 |
| 2174597 | chr5: 131997380-131997530 | 514 | (2) - ENSG00000113522,ENSG00000223442 |
| 2174600 | chr5: 131998145-131998295 | 2072 | (1) - ENSG00000113520 |
| 2174602 | chr5: 131998645-131998795 | 218 | (1) - ENSG00000113522 |
| 2174603 | chr5: 131998890-131999040 | 70 | (2) - ENSG00000113522,ENSG00000223442 |

o API

o Pou5f1 (Oct-4)

ed in Sheffield et al. (2013). This is c
Supplemental Files

lap for all clusters, and select indiv

a region.

<http://dnase.genome.duke.edu>



Prediction of Linkage Between Regulatory Elements and Genes

Regulatory Elements Database

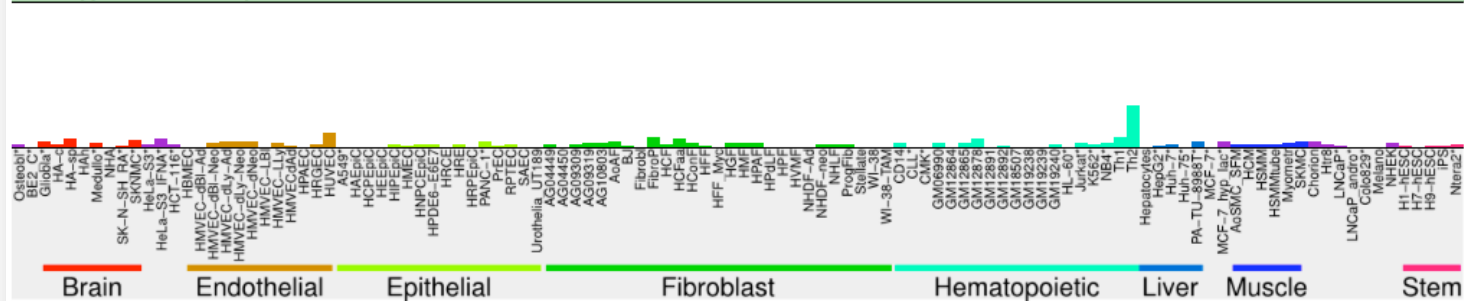
Chromosome, start, stop:

DHS: #2174550

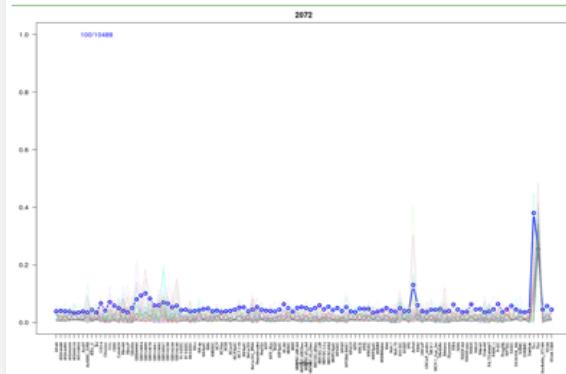
chr5: 131972960-131973110

Belongs to SOM cluster: 2072

[Site Hypersensitivity Profile](#)



Cluster Profile:



RESOURCES

Correlated Genes:

p-values indicate significant higher or lower correlation 1 genes found

Gene Pvalue

IL13 0.009


External Databases

[UCSC](#)
[Ensembl](#)

<http://dnase.genome.duke.edu>



ENCODE cis-element Browser

 **WELCOME TO YUE LAB**
Computational and Functional Genomics/Epigenomics

HOME **MOUSE** **HUMAN** **DOWNLOAD** **LINKS** **CONTACT**

Query human ENCODE data!

Option 1: Search gene expression across ~ 60 human cell types (total 108 datasets)

human (hg19) Gene name(Sox2, Nanog ...) submit!

Option 2: Search cis-elements in a given genomic region

human (hg19) chr1 start: end: submit!

Option 3: search cis-elements surrounding a gene

human (hg19) Gene name(Sox2, Nanog ...) kb submit!

Option 4: search cis-elements LINKED to a gene based on DNaseI HSS specificity

human (hg19) Gene name(Sox2, Nanog ...) IL13 submit!



ENCODE cis-element Browser

Cis-elements linked to your queried gene.

Human (hg19)

Gene **IL13** [NM_002188, ENSG00000169194, ENST00000304506]

Cis-element lined by DNaseI Hypersensitive Sites Linkage:

| Proximal DHS (TSS) | start | end | Gene | Distal DHS | start | end | correlation |
|--------------------|-----------|-----------|------|------------|-----------|-----------|-------------|
| chr5 | 131992140 | 131992290 | IL13 | chr5 | 131512800 | 131512950 | 0.743283 |
| chr5 | 131992140 | 131992290 | IL13 | chr5 | 131558440 | 131558590 | 0.761866 |
| chr5 | 131992140 | 131992290 | IL13 | chr5 | 131571820 | 131571970 | 0.782866 |
| chr5 | 131992140 | 131992290 | IL13 | chr5 | 131720440 | 131720590 | 0.766176 |
| chr5 | 131992140 | 131992290 | IL13 | chr5 | 131732540 | 131732690 | 0.739405 |
| chr5 | 131992140 | 131992290 | IL13 | chr5 | 131745200 | 131745350 | 0.765629 |
| chr5 | 131992140 | 131992290 | IL13 | chr5 | 131747860 | 131748010 | 0.749684 |
| chr5 | 131993580 | 131993730 | IL13 | chr5 | 131917860 | 131918010 | 0.797702 |
| chr5 | 131993580 | 131993730 | IL13 | chr5 | 131921920 | 131922070 | 0.800141 |
| chr5 | 131993580 | 131993730 | IL13 | chr5 | 131970980 | 131971130 | 0.772113 |
| chr5 | 131993580 | 131993730 | IL13 | chr5 | 131971640 | 131971790 | 0.763557 |
| chr5 | 131993580 | 131993730 | IL13 | chr5 | 131972960 | 131973110 | 0.797839 |
| chr5 | 131993580 | 131993730 | IL13 | chr5 | 131977060 | 131977210 | 0.848905 |
| chr5 | 131993580 | 131993730 | IL13 | chr5 | 131990420 | 131990570 | 0.855445 |
| chr5 | 131993580 | 131993730 | IL13 | chr5 | 131993880 | 131994030 | 0.769354 |
| chr5 | 131993580 | 131993730 | IL13 | chr5 | 132011780 | 132011930 | 0.756074 |
| chr5 | 131993580 | 131993730 | IL13 | chr5 | 132077740 | 132077890 | 0.820222 |
| chr5 | 131993580 | 131993730 | IL13 | chr5 | 132083520 | 132083670 | 0.770558 |
| chr5 | 131993580 | 131993730 | IL13 | chr5 | 132164240 | 132164390 | 0.837496 |
| chr5 | 131993580 | 131993730 | IL13 | chr5 | 132200200 | 132200350 | 0.752104 |



Prediction of Target Cell Types

- Some diseases are known to affect multiple cell types
- The defect may not be intrinsic to the cell type with obvious pathology
- The disease etiology may not be completely known



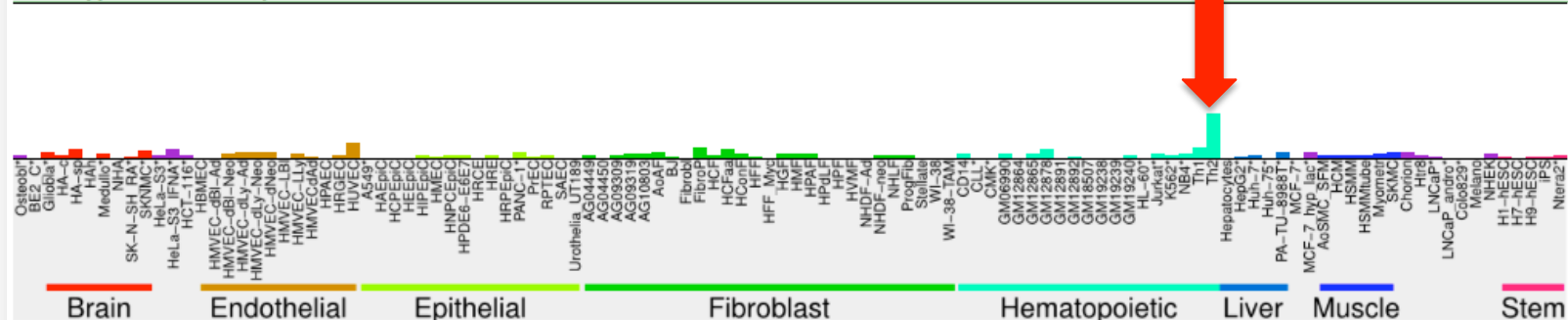
Prediction of Linkage Between Regulatory Elements and Cell Type

DHS: #2174550

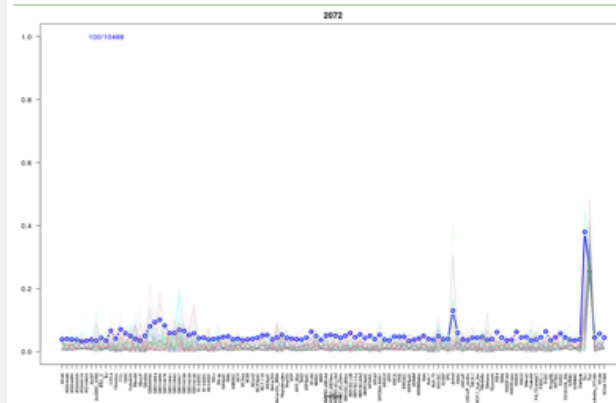
chr5: 131972960-131973110

Belongs to SOM cluster: 2072

Site Hypersensitivity Profile



Cluster Profile:



RESOURCES

Correlated Genes:

p-values indicate significant higher or lower correlation 1 genes found

Gene Pvalue

IL13 0.009

External Databases

UCSC

Ensembl

<http://dnase.genome.duke.edu>



Prediction of Linkage Between Regulatory Elements and Cell Type

HaploReg v4.1

HaploReg is a tool for exploring annotations of the noncoding genome at variants on haplotype blocks, such as candidate regulatory LD information from the 1000 Genomes Project, linked SNPs and small indels can be visualized along with chromatin state and prof. Epigenomics and ENCODE projects, sequence conservation across mammals, the effect of SNPs on regulatory motifs, and the effect studies. HaploReg is designed for researchers developing mechanistic hypotheses of the impact of non-coding variants on clinical p

Update 2015.11.05: Version 4.1 GWAS and eQTL have been updated; a simpler pruning strategy is applied when combining GWAS GWAS hits and GRASP QTL hits are provided.

Update 2015.09.15: Version 4.0 now includes many recent eQTL results including the GTEx pilot, four different options for defining data, and a complete set of source files for download and local analysis. Older versions available: [v3](#), [v2](#), [v1](#).

[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 ea 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): no file selected
 or, select a GWAS:

rs16892766 and variants with $r^2 \geq 0.8$

| ID | variant | Ref | Alt | AFR freq | AMR freq | ASN freq | EUR freq | SiPhy cons | Promoter histone marks | Enhancer histone marks | DNase-seq | Proteins bound | Motifs changed | NHGRIVEI GWAS hits | GRA hits |
|----|-------------------|-----|-----|----------|----------|----------|----------|------------|------------------------|------------------------|-----------|----------------|----------------|--------------------|----------|
| I | rs16892766 | A | C | 0.12 | 0.08 | 0.00 | 0.09 | | FAT | STRM, LNG, GI | 7 tissues | FOXA1,GR | Rhox11 | 2 hits | 1 hit |
| I | rs200235517 | CG | C | 0.47 | 0.10 | 0.04 | 0.09 | | | | 7 tissues | | | | |
| I | rs58147231 | GA | G | 0.47 | 0.10 | 0.04 | 0.09 | | | | 7 tissues | | | | |
| I | rs16888589 | A | G | 0.12 | 0.08 | 0.00 | 0.09 | | ESDR | | 8 tissues | | | | |

H1 Derived Mesenchymal Stem Cells, Fetal Lung, Fetal Stomach, Gastric, NH-A Astrocytes Primary Cells, NHDF-Ad Adult Dermal Fibroblast Primary Cells, NHLF Lung Fibroblast Primary Cells

Protein Binding

| Method | Location | Bound Protein | Cell Type | Additional Info | Reference |
|----------|---------------------------|---------------|-----------|-----------------|-----------|
| ChIP-seq | chr8:117630539..117630739 | FOXA1 | ECC-1 | 0.02pct | ENCODE |
| ChIP-seq | chr8:117630626..117630842 | NR3C1 | ECC-1 | DEX_100nM | ENCODE |

Chromatin structure

| Method | Location | Cell Type | Additional Info | Reference |
|-----------|---------------------------|-----------|-----------------|-----------|
| DNase-seq | chr8:117630480..117630690 | Rptec | | ENCODE |
| DNase-seq | chr8:117630480..117630730 | Nhlf | | ENCODE |

Histone modifications

| Method | Location | Chromatin State | Tissue Group | Tissue | Reference |
|----------|---------------------------|-------------------------|--------------|--|-----------|
| ChromHMM | chr8:117622200..117648800 | Quiescent/Low | Other | Pancreatic Islets | REMC |
| ChromHMM | chr8:117606000..117631800 | Weak Repressed PolyComb | ENCODE | GM12878 Lymphoblastoid Cell Line | REMC |
| ChromHMM | chr8:117611200..117638000 | Weak Repressed PolyComb | ENCODE | HepG2 Hepatocellular Carcinoma Cell Line | REMC |

Workshop Session 3

www.broadinstitute.org/mammals/haploreg/

<http://regulomedb.org/>



ENCODE cis-element Browser

PENNSSTATE
1855

Welcome to YUE Lab
Computational and Functional Genomics/Epigenomics

HOME MOUSE HUMAN DOWNLOAD LINKS CONTACT

Query mouse ENCODE data!

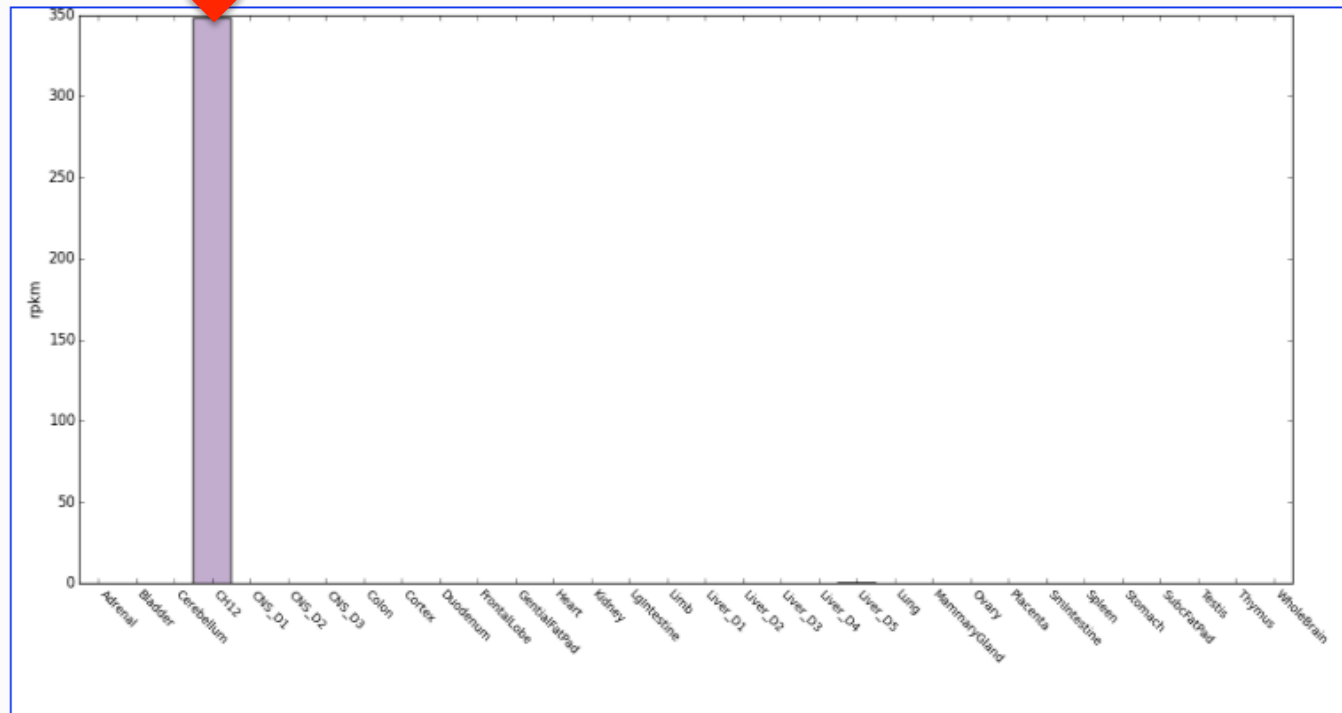
Option 1: Search gene expression across 32 mouse tissue/cell types

mouse (mm9) Gene name(Sox2, Nanog ...) I110 submit!



ENCODE cis-element Browser

Gene **II10** (mCC2645) [NM_010548, ENSMUSG00000016529, ENSMUST00000016673]

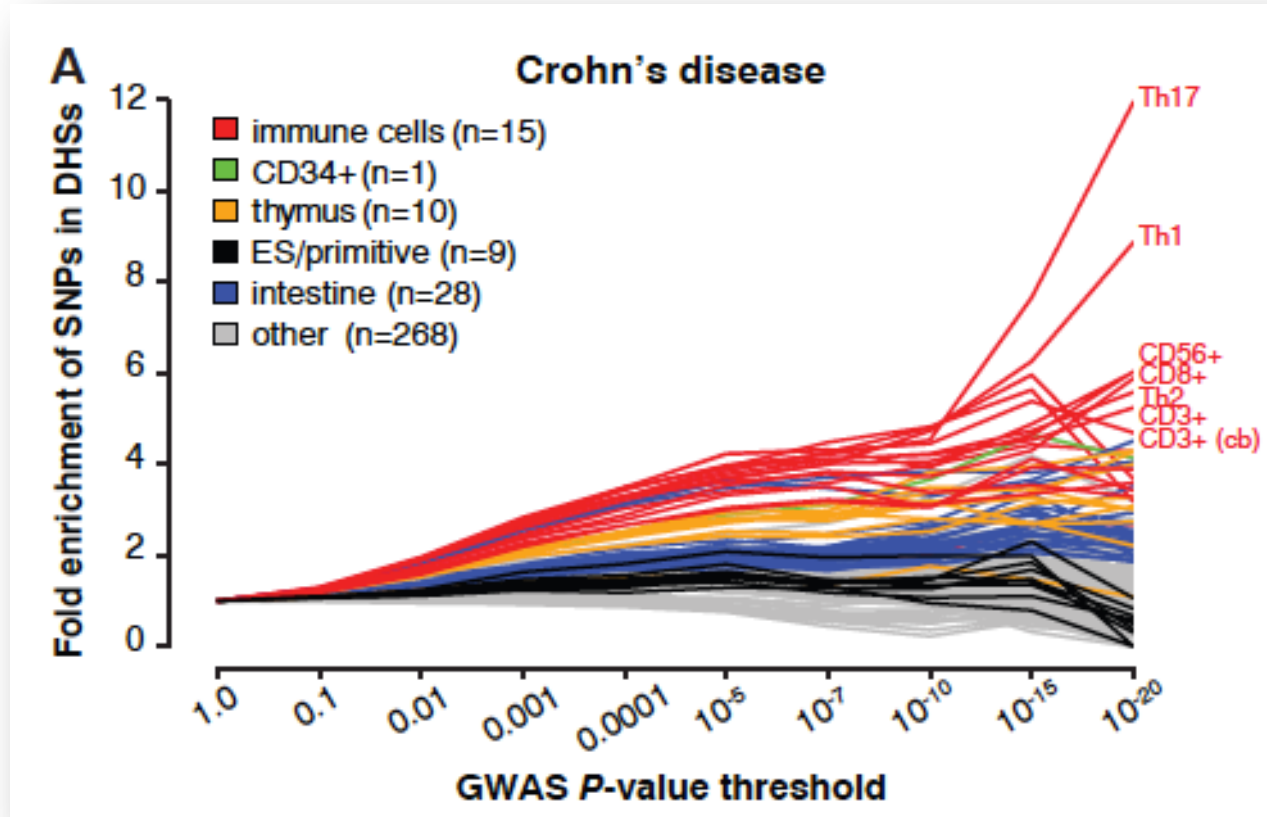


| | |
|------------|--------|
| Adrenal | 0 |
| Bladder | 0.07 |
| Cerebellum | 0 |
| CH12 | 348.42 |
| CNS_D1 | 0 |





ENCODE And Epigenomics Data Can Be Used To Predict Cell Types



https://github.com/mauranolab/GWAS_plots
Stamatoyannopoulos, Science 337:1190, 2012



Summary- ENCODE Use Cases

Major use: Hypothesis generation and refinement

- Prediction of causal variants/regulatory elements
 - Prediction of target genes
 - Prediction of target cell types
 - Prediction of upstream regulators
-
- Genetic v. epigenetic
 - Germline v. somatic



Overview

- The ENCODE Resource
- Use of ENCODE by the research community
- **Accessing ENCODE materials**

Workshop Session 2, 1, 5



ENCODE Data

ENCODE Encyclopedia of DNA Elements

The ENCODE (Encyclopedia of DNA Elements) Consortium is a collaboration of research groups funded by the National Institutes of Health (NIH). The goal of ENCODE is to build a comprehensive catalog of functional elements in the human genome, including enhancers, promoters, and RNA levels, and regulatory elements that control gene expression.

Image credits: Darryl Leja (NHGRI), Ian Dunham (EBI),

Quick Start **News** Follow @EncodeDCC



ENCODE Data

ENCODE

ENC

Quicl

Experiment Matrix

Click or enter search terms to filter the experiments included in the matrix.

Organism

- Homo sapiens* 4732
- Mus musculus* 1609

Biosample type

- immortalized cell line 3405
- tissue 1548
- primary cell 913
- in vitro differentiated cells 227
- stem cell 223

Organ

- brain 459
- skin of body 186
- liver 183
- heart 155
- lung 141

Project

- ENCODE** 6452
- Roadmap 3127
- modENCODE 883
- modERN 198

Assay

- ChIP-seq 3105
- shRNA RNA-seq 445
- DNase-seq 421
- RNA-seq 307
- eCLIP 270

+ See more...

Assay category

- DNA binding 3105
- Transcription 1823
- DNA accessibility 470
- RNA binding 460
- DNA methylation 310

+ See more...

Target of assay

- transcription factor 1367
- histone 1225
- histone modification 1225
- RNA binding protein 893
- control 866

+ See more...

ASSAY

6452 results

Clear Filters

| BIOSAMPLE | ASSAY | | | | | | | | | | | |
|---|----------|---------------|-----------|---------|-------|--------------------|---------------|---------------------|----------------|-----------|------|---|
| | ChIP-seq | shRNA RNA-seq | DNase-seq | RNA-seq | eCLIP | polyA mRNA RNA-seq | small RNA-seq | single cell RNA-seq | RNA Bind-n-Seq | DNase-seq | RRBS | |
| immortalized cell line | | | | | | | | | | | | |
| K562 | 374 | 227 | 21 | 13 | 156 | 18 | 12 | 8 | | 2 | 1 | 6 |
| HepG2 | 150 | 218 | 3 | 6 | 112 | 11 | 7 | 3 | | 2 | 2 | 6 |
| GM12878 | 174 | | 2 | 5 | 10 | 8 | 6 | 13 | | 2 | 2 | 6 |
| HeLa-S3 | 103 | | 4 | | 5 | 4 | 3 | | | 1 | 1 | 6 |
| A549 | 110 | | 2 | | 9 | 2 | 9 | | | 1 | 1 | |
| ...and 141 more | | | | | | | | | | | | |
| tissue | | | | | | | | | | | | |
| liver | 102 | | 5 | 11 | | 11 | 1 | 2 | | 1 | 1 | |
| heart | 84 | | 6 | 10 | | 6 | 1 | 1 | | | | |
| forebrain | 65 | | 2 | 8 | | | | 2 | | | | |
| hindbrain | 65 | | 3 | 8 | | | | 1 | | | | |
| midbrain | 65 | | 4 | 8 | | | | 1 | | | | |
| ...and 101 more | | | | | | | | | | | | |
| primary cell | | | | | | | | | | | | |
| endothelial cell of umbilical vein | 35 | | 2 | | 5 | 2 | 1 | | | 1 | | 6 |
| Purkinje cell | | | | 1 | | | | 61 | | | | |

43

Workshop Session 2

<https://www.encodeproject.org>



ENCODE Data

Data Use, Software, and Analysis Release Policies

The goal of the Encyclopedia of DNA Elements (ENCODE) Project is to build a comprehensive catalog of candidate functional elements in the genome. The catalog includes genes (protein-coding and non-protein coding), transcribed regions, and regulatory elements, as well as information about the tissues, cell types and conditions where they are found to be active. The current phase of ENCODE (2012-2016) greatly expands the number of cell types, data types and assays and includes the study of both the human and mouse genomes.

Like the Human Genome Project, the ENCODE Project seeks rapid data dissemination and use by the entire scientific community. Accordingly, to encourage the widest possible use of the datasets, all data produced will be available for unrestricted use immediately upon release to public databases, eliminating the nine-month moratorium previously used by ENCODE.

Data Use Policy for External Users

External data users may freely download, analyze and publish results based on any ENCODE data without restrictions as soon as they are released. This applies to all datasets, regardless of type or size, and includes no grace period for ENCODE data producers, either as individual members or as part of the Consortium. Researchers using unpublished ENCODE data are encouraged to contact the data producers to discuss possible coordinated publications; however, this is optional. The Consortium will continue to publish the results of its own analysis efforts in independent publications.

We request that researchers who use ENCODE datasets (published or unpublished) in publications and talks cite the ENCODE Consortium in all of the following ways:

1. Cite the Consortium's most recent integrative publication (PMID: [22955616](#); PMC: [PMC3439153](#));
2. Reference the ENCODE Data Coordination Center (DCC) or GEO accession numbers of the datasets (DCC accession: [ENCSR037HRJ](#); GEO accession: [GSE30567](#));
3. And acknowledge the ENCODE Consortium and the ENCODE production laboratory(s) generating the particular dataset(s)



ENCODE Data

ENCODE Data Encyclopedia Materials & Methods Help

Matrix
Search
Search by region
Publications

ENCODE Encyclopedia of DNA Elements

The ENCODE (Encyclopedia of DNA Elements) Consortium is an international collaboration of research groups funded by the National Institutes of Health (NIH). The goal of ENCODE is to build a comprehensive catalog of functional elements in the human genome, including enhancers and RNA levels, and regulatory elements that control when and where a gene is active.

Image credits: Darryl Leja (NHGRI), Ian Dunham (EBI),



ENCODE Data

ENCODE Data Encyclopedia Materials & Methods Help

Matrix
Search
Search by region

ENCODE Encyclopedia of DNA Elements

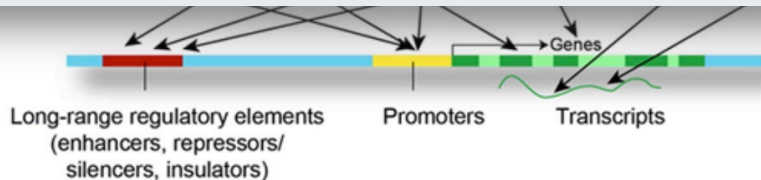
ENCODE Data Encyclopedia Materials & Methods Help

Region search

Enter any one of human Gene name, Symbol, Synonyms, Gene ID, HGNC ID, coordinates, rsid, Ensemble ID

Search

Please enter valid coordinates





ENCODE Data

ENCODE Data Encyclopedia Materials & Methods Help

Matrix
Search
Search by region

Encyclopedia of DNA Elements

Region search

rs16892766 chr8:117630683-117630683

Showing 25 of 31

| Assay | Count |
|-----------|-------|
| DNase-seq | 28 |
| ChIP-seq | 3 |

| Biosample term | Count |
|-----------------|-------|
| Ishikawa | 7 |
| skin fibroblast | 4 |
| A172 | 1 |
| Caki2 | 1 |
| HTR-8/SVneo | 1 |

[+ See more...](#)

| Target | Count |
|--------|-------|
| ESR1 | 1 |
| FOXA1 | 1 |
| NR3C1 | 1 |

| Organism | Count |
|---------------------|-------|
| <i>Homo sapiens</i> | 31 |

DNase-seq of cardiac mesoderm
Homo sapiens
Lab: John Stamatoyannopoulos, UW
Project: ENCODE

DNase-seq of skin fibroblast
Homo sapiens, adult 71 year
Lab: Gregory Crawford, Duke
Project: ENCODE

DNase-seq of HTR-8/SVneo
Homo sapiens
Lab: Gregory Crawford, Duke

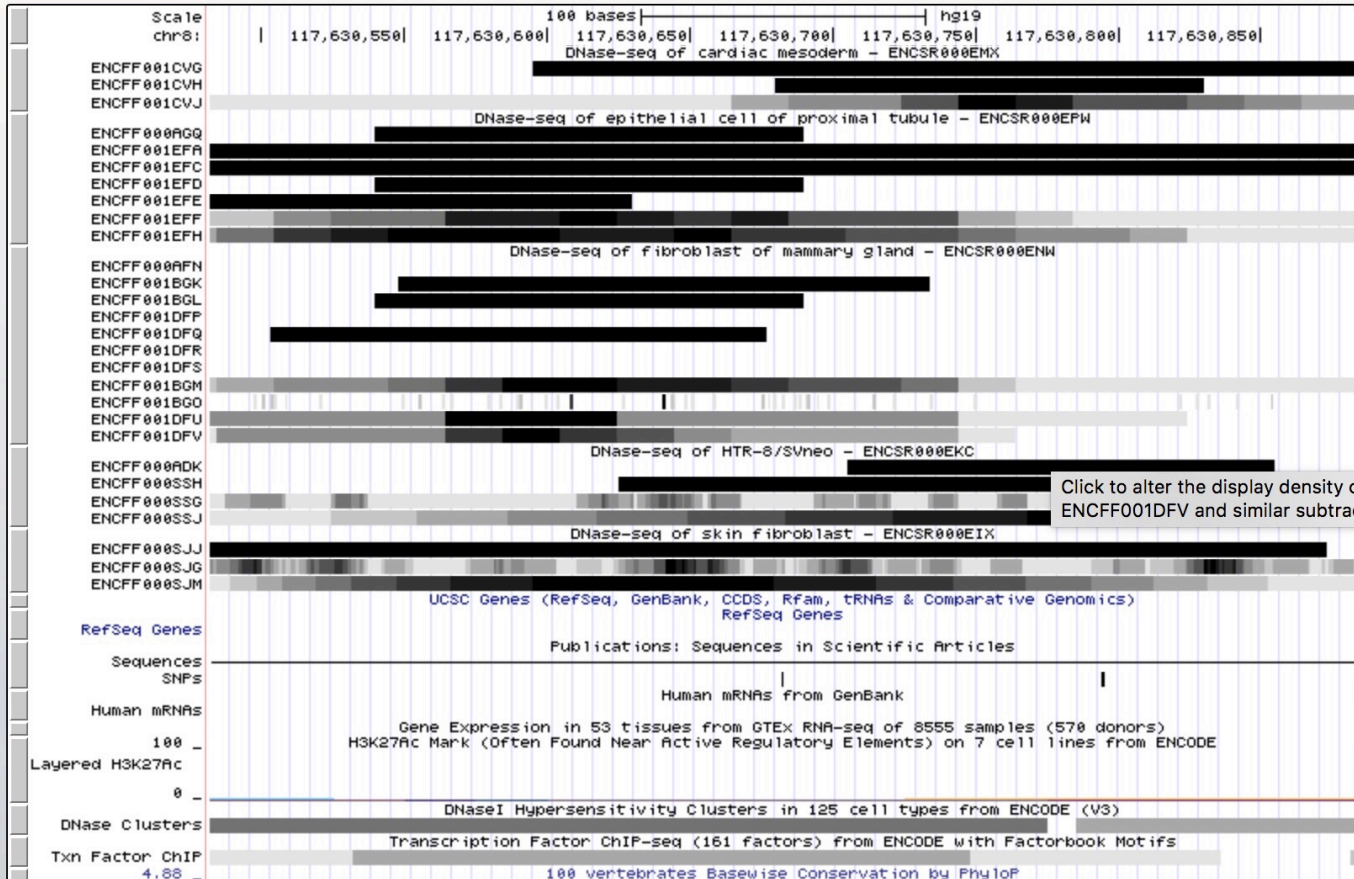


ENCODE Data

UCSC Genome Browser on Human Feb. 2009 (GRCh37/hg19) Assembly

move <<< << < > >> >>> zoom in 1.5x 3x 10x base zoom out 1.5x 3x 10x 100x

chr8:117,630,483-117,630,883 401 bp.



Click to alter the display density of ENCF001DFV and similar subtracks



ENCODE Encyclopedia

ENCODE Data Encyclopedia Materials & Methods Help

About Matrix Search

ENCODE Encyclopedia of DNA Elements

The ENCODE (Encyclopedia of DNA Elements) Consortium is a collaboration of research groups funded by the National Human Genome Research Institute (NHGRI). The goal of ENCODE is to build a comprehensive catalog of functional elements in the human genome, including enhancers, promoters, and RNA levels, and regulatory elements that control when and where a gene is active.

Image credits: Darryl Leja (NHGRI), Ian Dunham (EBI), et al.

Quick Start **News** [Follow @EncodeDCC](#)



ENCODE Encyclopedia

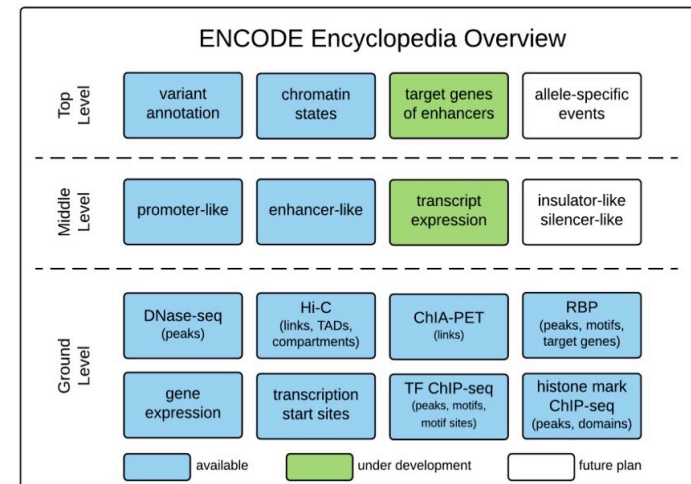


ENCODE Encyclopedia: Genomic annotations

Introduction

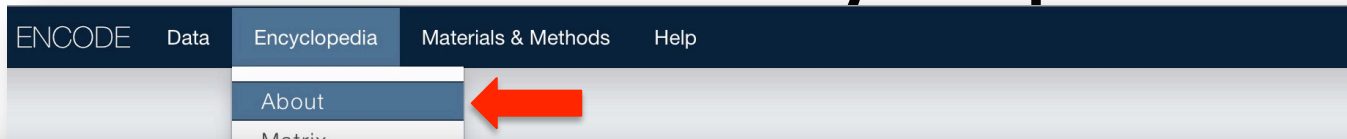
The ENCODE Consortium not only produces data, but also analyzes the data in an integrative fashion. The ENCODE Encyclopedia organizes the most salient analysis products into annotations, and provides tools to search and visualize them. The Encyclopedia has three levels of annotations:

- Ground level annotations are typically derived directly from the experimental data.
- Middle level annotations integrate multiple types of experimental data and multiple ground level annotations.
- Top level annotations integrate a broad range of experimental data and ground and middle level annotations.





ENCODE Encyclopedia



Ground Level Annotations

Gene expression (RNA-seq)

The expression levels of genes annotated by GENCODE 19 in over 100 human cell types and 70 mouse cell types.

[[Long RNA-seq Data](#) | [Query](#) | [Download](#) | [Method](#)]

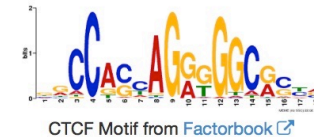


Transcription factor binding (TF ChIP-seq)

Peaks (enriched genomic regions) of TFs computed from ~900 human and mouse ChIP-seq experiments.

[[Raw Data](#) | [Peaks](#)]

Visualize sequence motifs and other information [[Factorbook](#)]



Histone mark enrichment (ChIP-seq)

Peaks of a variety of histone marks computed from ~600 ChIP-seq experiments.

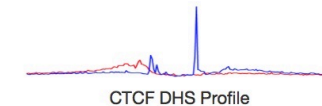
[[Raw Data](#) | [Peaks](#)]



Open chromatin (DNase-seq)

DNase I hypersensitive sites (also known as DNase-seq peaks) computed from ~300 human and mouse experiments.

[[Raw Data](#) | [Peaks](#)]



Topologically associating domains (TADs) and compartments (Hi-C)

TADs and A and B compartments computed from 12 human cell lines.

[[Raw Data](#) | [Visualize](#)]





ENCODE Encyclopedia

ENCODE

Data

Encyclopedia

Materials & Methods

Help

Middle Level Annotations

Gro

Promoter-like regions

Ger

ENC

The e

cell ty

[Lon

Trai

Peak

[Raw

Visua

DNase hypersensitivity and histone modification H3K4me3 are well-known indicators of promoter function. We have developed an unsupervised method that combines DNase and H3K4me3 signals in the same cell type to predict promoter-like regions. When used to predict ranked gene expression from RNA-seq data, our method shows higher accuracy than DNase and H3K4me3 individually. We have applied this method to 107 human cell types and 14 mouse cell types with both DNase and H3K4me3 data generated by the ENCODE and Roadmap Epigenomic consortia. For cell and tissues types with only H3K4me3 data, we centered predictions on H3K4me3 peaks and ranked them by H3K4me3 signals. You can query these promoter-like regions by genomic locations, nearby genes, or SNPs, and visualize them in the UCSC and WashU genome browsers.

[[Visualize](#) | | [Method](#)]

Enhancer-like regions

His

Peak

[Raw

Ope

DNas

exper

[Raw

Top

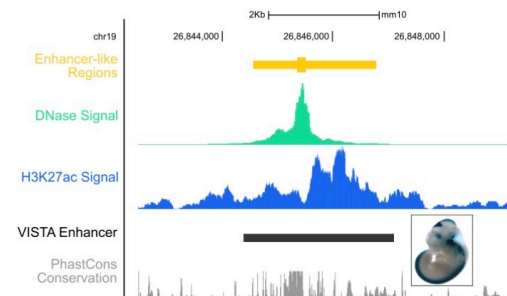
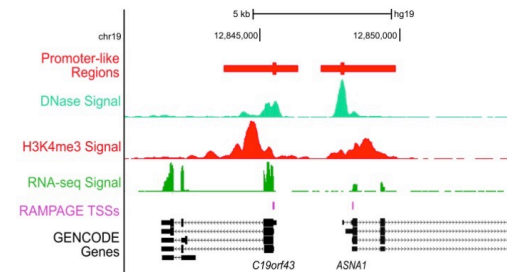
TADs

[Raw

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 tissues 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.

[[Visualize](#) | | [Method](#)]

Enhancer-like genomic regions were tested on VISTA experimentally-validated enhancer elements: [[VISTA](#)]





ENCODE Encyclopedia

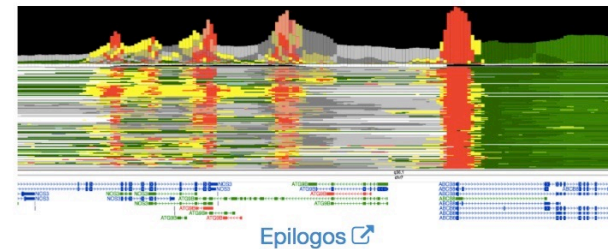
Top Level Annotations

Chromatin states



Semi-automated genomic annotation methods such as ChromHMM and Segway take as input a panel of epigenomic data (including histone mark ChIP-seq and DNase-seq) in a particular cell type and use machine learning methods to simultaneously partition the genome into segments and assign chromatin states to these segments; the states are assigned such that two segments with the same state exhibit similar epigenomic patterns. The procedure is "semi-automated" because states are then manually compared with known biological information in order to designate each state as an enhancer-like, promoter-like, gene body, etc.

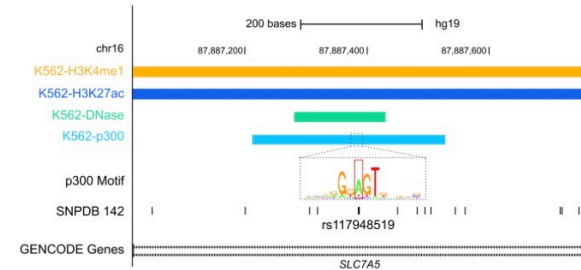
[[Search](#)]



Variant Annotation

Over the past decade, Genome Wide Association Studies (GWAS) have provided insights into how genetic variations contribute to human diseases. However, over 80% of the variants reported by GWAS are in noncoding regions of the genome and the mechanism of how they contribute to disease onset is unknown. By integrating data from the ENCODE project and other public sources, RegulomeDB and HaploReg are two resources developed by ENCODE labs to aid the research community in annotating GWAS variants. FunSeq2 is another ENCODE resource for annotating both germline and somatic variants, particularly in the noncoding regions of cancer genomes.

[[RegulomeDB](#) | [HaploReg](#) | [FunSeq2](#)]



TADs

[[Raw](#)

Enhancer-like genomic regions were tested on VISTA experimentally-validated enhancer elements: [[VISTA](#)]



Publications

ENCODE

Data

Encyclopedia

Materials & Methods

Help

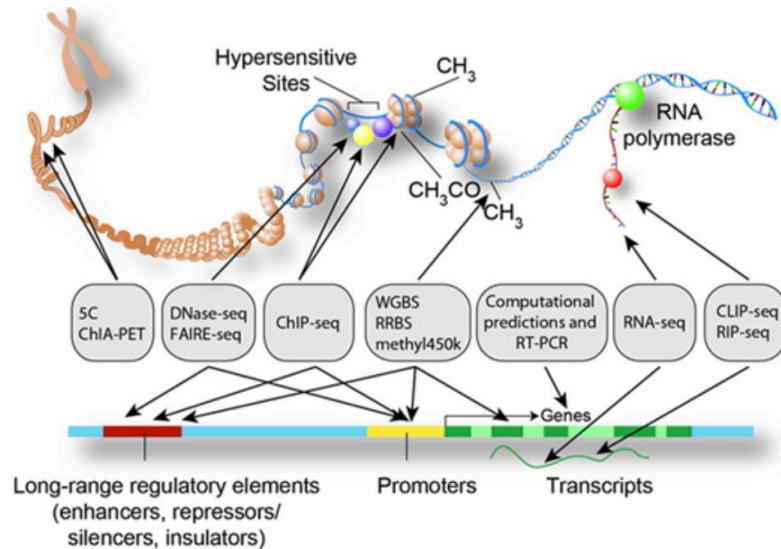
Matrix

Search

Search by region

Publications

ENCODE Encyclopedia of DNA Elements



The ENCODE (Encyclopedia of DNA Elements) Consortium is a collaboration of research groups funded by the National Institutes of Health (NIH). The goal of ENCODE is to build a catalog of functional elements in the human genome, including enhancers and RNA levels, and regulatory elements that control when and where a gene is active.

Image credits: Darryl Leja (NHGRI), Ian Dunham (EBI),



Publications

ENCODE

Data

Encyclopedia

Materials & Methods

Help

ENCODE

Matrix

Search

Search by region

Publications

Encyclopedia of DNA Elements

Publications using ENCODE data

ENCODE-funded publications

These are publications by members of the ENCODE, mouse ENCODE, and modENCODE consortia.

- [Key integrative publications by consortia members](#)
- [Human ENCODE publications](#)
- [Mouse ENCODE publications](#)
- [modENCODE publications](#)
- [Technology development publications](#)
- [ENCODE pilot project publications](#)

Community publications

These are publications that use ENCODE, published by authors not funded by ENCODE, as well as papers that use modENCODE data, published by authors not funded by modENCODE. The ENCODE project tracks these papers to assess impact of the resource and to provide examples of how the resource can be used. Please contact [Mike Pazin at NHGRI](#) to suggest publications to add to these lists.

- [Human disease publications](#)
- [Basic biology publications](#)
- [Software tools](#)
- [modENCODE community publications](#)



ENCODE Data Standards

The screenshot shows the ENCODE website interface. At the top, there is a navigation bar with links for 'ENCODE', 'Data', 'Encyclopedia', 'Materials & Methods', and 'Help'. A dropdown menu is open under 'Materials & Methods', listing various categories: 'Antibodies', 'Biosamples', 'Standards and guidelines' (highlighted with a red arrow), 'Ontologies', 'File formats', 'Software tools', 'Pipelines', 'Release policy', and 'Data access'. Below the navigation bar, the main content area features a large diagram illustrating genomic elements and associated technologies. The diagram shows a DNA strand with various elements: 'Long-range regulatory elements (enhancers, repressors/silencers, insulators)', 'Promoters', and 'Transcripts'. Above the DNA, there are 'Hypersensitive Sites' and 'Genes'. A red arrow points to the 'Standards and guidelines' menu item. Below the diagram, there are two buttons: 'Quick Start' and 'News Follow @EncodeDCC'. The text on the right side of the screenshot describes the ENCODE project as a collaboration of research groups funded by the National Human Genome Research Institute (NHGRI), with the goal of building a catalog of functional elements in the human genome.

ENCODE: Encyclopedia of DNA Elements

The ENCODE (Encyclopedia of DNA Elements) Consortium is an international collaboration of research groups funded by the National Human Genome Research Institute (NHGRI). The goal of ENCODE is to build a comprehensive catalog of functional elements in the human genome, including enhancers, promoters, and RNA levels, and regulatory elements that control gene expression.

Image credits: Darryl Leja (NHGRI), Ian Dunham (EBI), et al.



ENCODE Data Standards

ENCODE Data Encyclopedia Materials & Methods Help

Antibodies
Biosamples
Standards and guidelines ← Elements

ENCODE: Encyclopedia of Data Standards

Data standards

Overview

The ENCODE consortium analyzes the quality of the data produced using a variety of metrics. This page describes the data standards and metrics that are used to evaluate the data and what they appear to measure. These quality metrics will be updated on occasion to include analysis of more recent data.

It is important to note that quality metrics for evaluating epigenomic assays is an area of research, so standards are emerging as more metrics are used with more datasets and types of experiments. The typical values for a quality metric can be quite different with different assays, or even comparing different features in the same assays, such as different antibodies used in ChIP-seq experiments. Currently there is no single measurement that identifies all high-quality or low-quality samples. As with quality control for other types of experiments, multiple assessments (including manual inspection of tracks) are useful because they may capture different concerns. Comparisons within an experimental method (e.g., comparing replicates to each other, or comparing values for one antibody in several cell types, or the same antibody and cell type in different labs) can help identify possible stochastic error.

Experimental guidelines

The ENCODE Consortium has adopted uniform guidelines for the most common ENCODE experiments. The guidelines have evolved over time as technologies have changed. The current guidelines are informed by results gathered during the project. Previous versions of the standards are also available for reference.

- [Current experiment guidelines](#)
- [Antibody characterizations guidelines](#)
- [Terms and definitions](#)

Quality metrics



ENCODE Software Tools

ENCODE Data Encyclopedia Materials & Methods Help

- Antibodies
- Biosamples
- Standards and guidelines
- Ontologies
- File formats
- Software tools**
- Pipelines
- Release policy
- Data access

ENCODE: Encyclopedia of DNA Elements

Long-range regulatory elements (enhancers, repressors/silencers, insulators) Promoters Transcripts Genes

5C ChIA-PET DNase-seq FAIRE-seq ChIP-seq WGBS RRBS methyl450k Computational predictions and RT-PCR RNA-seq CLIP-seq RIP-seq

Hypersensitive Sites CH_3CO CH_3

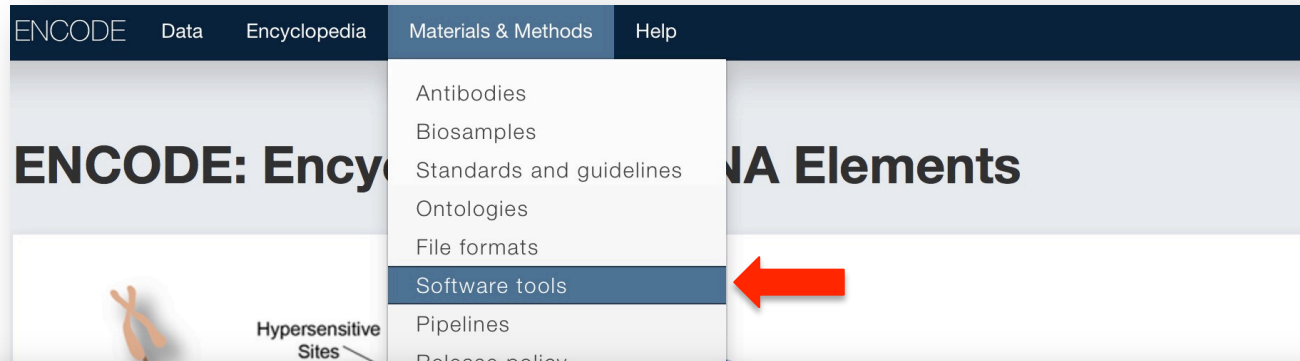
The ENCODE (Encyclopedia of DNA Elements) Consortium is a collaboration of research groups funded by the National Institutes of Health (NIH). The goal of ENCODE is to build a comprehensive catalog of functional elements in the human genome, including enhancers, promoters, and RNA levels, and regulatory elements that control when and where a gene is active.

Image credits: Darryl Leja (NHGRI), Ian Dunham (EBI),

Quick Start **News** Follow @EncodeDCC



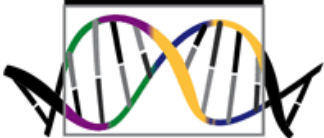
ENCODE Software Tools



Software Tools

The goal of the ENCODE project is to generate a comprehensive catalog of all functional elements. To facilitate this task, members of the consortium have developed and refined software tools. [View all software](#) used or developed by the ENCODE Consortium or select one of the following:

- [Software tools used to identify ENCODE elements](#): On this page are brief descriptions of some of the software used to identify ENCODE elements. Software for identification of functional elements, for integrated analysis of multiple data types, and for quality measurement of the data are described.
- [Software tools used to generate ENCODE quality metrics](#): On this page are brief descriptions of some of the software used to generate quality metrics for ENCODE datasets.
- [External software tools used to create the ENCODE resource](#): On this page are brief descriptions of some of the software used to create the ENCODE resource. This software was not funded by ENCODE, or developed by the consortium.
- [Software tools and resources for applying and analyzing ENCODE data](#): On this page are brief descriptions of software and resources that others might find useful for analyzing and using ENCODE data in their own research.



International Human Epigenome Consortium (IHEC)

- Data Portal: <http://epigenomesportal.ca/ihec/>
- Goal: Coordinate production of 1000 human epigenome maps for cellular states relevant to health and disease <http://ihec-epigenomes.org>
- Can view by consortium, by assay, by cell type
- Data from 8 consortia



IHEC
International Human Epigenome Consortium





Summary- Accessing ENCODE Resources

- ENCODE portal <https://www.encodeproject.org>
 - Display/download ENCODE and Roadmap Epigenomics data
 - Data Standards
 - Software tools
 - Publications
 - Encyclopedia prototype
- ENCODE Analysis Tools
 - RegulomeDB <http://regulomedb.org/>
 - HaploReg <http://www.broadinstitute.org/mammals/haploreg/>
 - Regulatory Elements Database <http://dnase.genome.duke.edu>
 - RegulomeDB GWAS Database <http://www.regulomedb.org/GWAS/>
- ENCODE Tutorials
 - <http://www.genome.gov/27553900>
 - <https://www.encodeproject.org/tutorials/>
<http://www.ncbi.nlm.nih.gov/pubmed/25762420>
- ENCODE mailing list :
 - <https://mailman.stanford.edu/mailman/listinfo/encode-announce>
- IHEC resources
 - IHEC Home Page <http://ihec-epigenomes.org>
 - IHEC Data Portal <http://epigenomesportal.ca/ihec/>



Goals Of ENCODE

- Catalog all functional elements in the genome
- Develop freely available resource for research community

ENCODE data are being used in the study of human disease and basic biology