



How to Display and Download ENCODE Data

Mike Pazin
NHGRI, NIH



Goals of ENCODE

- Catalog all functional elements in the genome
- Freely available resource for all biologists
- Human as well as other species
- Project components:
 - Data generation
 - Data analysis
 - Data repository



How Can ENCODE Data Be Displayed?



Encyclopedia of

Human

Integrative Analysis

Experiment Matrix

Experiment List

Search

Downloads

Genome Browser (hg19)

Session Gallery

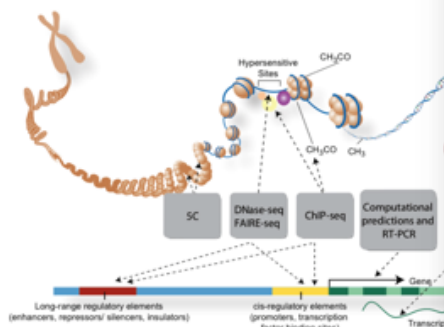
Cell Types

Mouse

Experiment Matrix

About ENCODE Data

The [Encyclopedia of DNA Elements](#) (ENCODE) is a project of the National Human Genome Research Institute (NHGRI). The goal is to identify all the functional elements that act at the protein and RNA levels, and



[Click to enlarge](#)

To search for ENCODE data related to (e.g., a specific gene or features). The [Experiment List](#) (Human) you would like to receive notifications of new information about how to access this data.

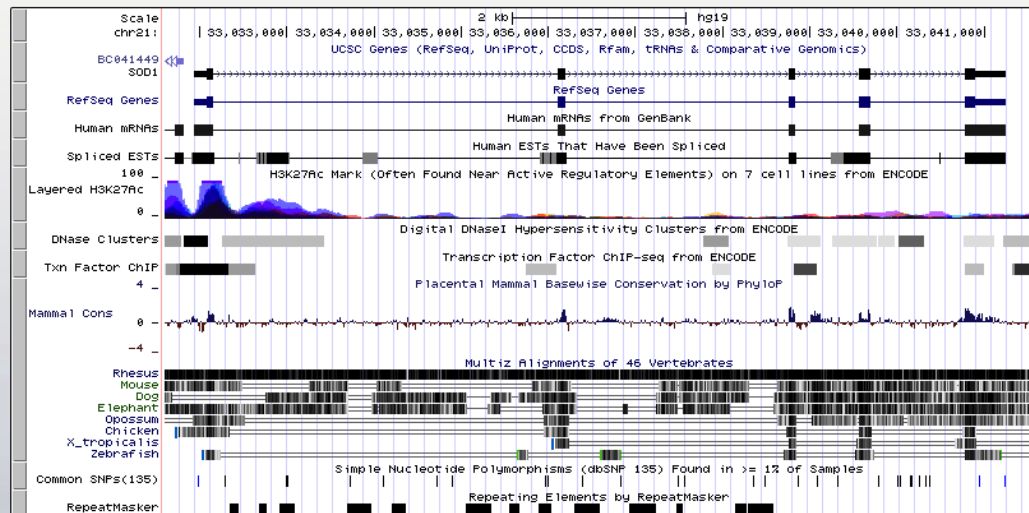
All ENCODE data is freely available for

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

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

chr21:33,031,597-33,041,570 9,974 bp. enter position, gene symbol or search terms go

chr21 (q22.11) 21p13 21p12 21p11.2 21q21.1 21q21.2 21q21.3 21q22.11 22q2.2 21q22.3



move start < 2.0 > move end < 2.0 >
Click on a feature for details. Click or drag in the base position track to zoom in. Click side bars for track options. Drag side bars or labels up or down to reorder tracks. Drag tracks left or right to new position.

track search default tracks default order hide all add custom tracks track hubs configure reverse resize refresh

collapse all Use drop-down controls below and press refresh to alter tracks displayed. Tracks with lots of items will automatically be displayed in more compact modes. expand all

Mapping and Sequencing Tracks refresh

Base Position Chromosome Band STS Markers FISH Clones Recomb Rate deCODE Recomb
dense hide hide hide hide hide

Handout

<http://encodeproject.org>



How Can The Display Be Configured?



chr21:33,031,597-33,041,570 9,974 bp. enter position, gene symbol or search terms go

chr21 (q22.11)

Scale chr21: | 33,000

BC041449 SOD1

RefSeq Genes

Human mRNAs

Spliced ESTs

Layered H3K27Ac

DNase Clusters

Txn Factor ChIP

Mammal Cons

Rhesus

Mouse

Dog

Elephant

Opossum

Chicken

X_tropicalis

Zebrafish

Common SNPs(135)

RepeatMasker

move start < 2.0 > Click to zoom up

track search default tracks

collapse all Use default tracks

Regulation

<input checked="" type="checkbox"/> ENCODE Regulation... show	<input checked="" type="checkbox"/> CD34 Dnase1 hide	<input checked="" type="checkbox"/> CpG Islands hide	<input checked="" type="checkbox"/> ENC Chromatin... hide
<input checked="" type="checkbox"/> ENC Histone... hide	<input checked="" type="checkbox"/> ENC RNA Binding... hide	<input checked="" type="checkbox"/> ENC TF Binding... hide	<input checked="" type="checkbox"/> FSU Repli-chip hide

Phenotype and Disease Associations

<input checked="" type="checkbox"/> GAD View hide	<input checked="" type="checkbox"/> DECIPHER hide	<input checked="" type="checkbox"/> OMIM AV SNPs hide	<input checked="" type="checkbox"/> OMIM Genes hide
<input checked="" type="checkbox"/> GWAS Catalog full	<input checked="" type="checkbox"/> ISCA hide	<input checked="" type="checkbox"/> RGD Human QTL hide	<input checked="" type="checkbox"/> RGD Rat QTL hide

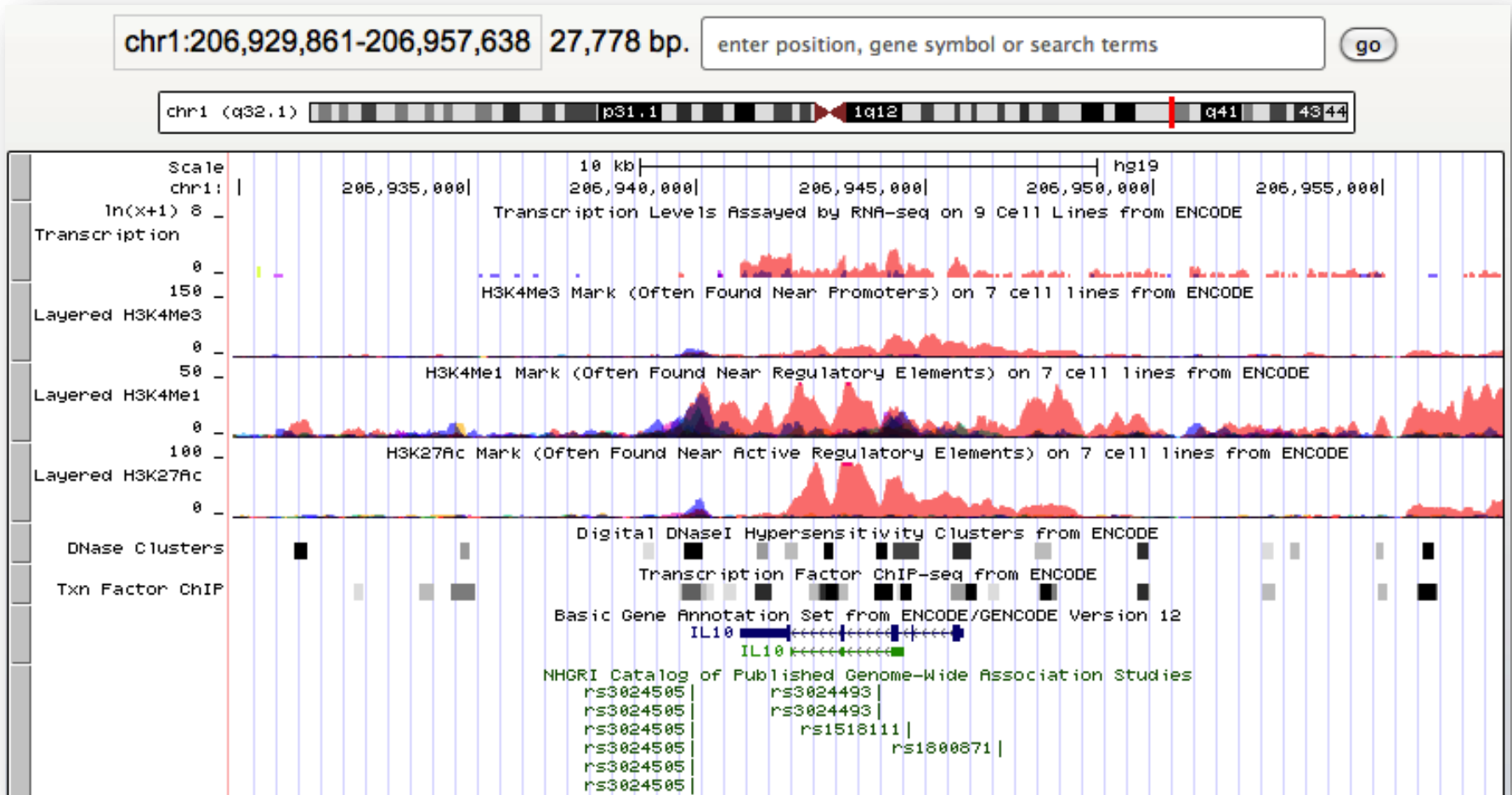
Mapping and Sequencing Tracks

<input checked="" type="checkbox"/> Base Position dense	<input checked="" type="checkbox"/> Chromosome Band hide	<input checked="" type="checkbox"/> STS Markers hide	<input checked="" type="checkbox"/> FISH Clones hide	<input checked="" type="checkbox"/> Recomb Rate hide	<input checked="" type="checkbox"/> deCODE Recomb hide
--	---	---	---	---	---

<http://encodeproject.org>



ENCODE Browser, Locus of Interest



<http://encodeproject.org>



ENCODE Data From HaploReg



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

Ward and Kellis, Nucleic Acids Research 40-D930, 2011

HaploReg



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 in nine cell types, 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.

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 (refSNP ID(s), comma-delimited): ← 1

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

or, select a GWAS:

← 2

Query SNP: **rs3024505** and variants with $r^2 \geq 1$ ← 3

chr	pos (hg19)	LD	variant	Ref	Alt	ASN freq	CEU freq	YRI freq	GERP cons	SiPhy cons	Promoter histone marks	Enhancer histone marks	DNAse	Proteins bound	eQTL tissues	Motifs changed	GENCODE genes	RefSeq genes	dbSNP func annot
1	206939904	1	rs3024505	G	A	0.02	0.19	0.03			GM12878	K562, NHLF	74 cell types	22 bound proteins			1kb 3' of IL10	1kb 3' of IL10	
1	206942413	1	rs3024495	C	T	0	0.19	0.02				GM12878					IL10	IL10	intronic



ENCODE Data From RegulomeDB



<http://regulome.stanford.edu/> ; Boyle...Snyder, Genome Research 22-1790,2012

RegulomeDB

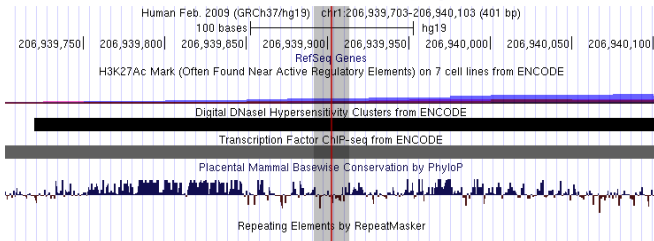
Enter dbSNP IDs, 0-based coordinates, BED files, VCF files, GFF3 files (hg19).

rs3024505

Submit

Data supporting chr1:206939903 (rs3024505)

Score: 2b
Likely to affect binding



Method	Location	Bound Protein	Cell Type	Additional Info	Reference
ChIP-seq	chr1:206939412..206940458	CREBBP	Jurkat		20019798
ChIP-seq	chr1:206939690..206939920	SPI1	GM12891		ENCODE
ChIP-seq	chr1:206939664..206939945	EP300	HepG2		ENCODE

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

Method	Location	Cell Type	Additional Info	Reference
DNase-seq	chr1:206939662..206940103	Htr8		ENCODE
DNase-seq	chr1:206939677..206940116	Panisd		ENCODE
DNase-seq	chr1:206939685..206940094	Aosmc	Serum	ENCODE
DNase-seq	chr1:206939684..206940110	Fibrop		ENCODE

Method	Location	Histone Mark	Cell Type	Additional Info	Reference
ChIP-seq	chr1:206534479..207335459	H4k20me1	Helas3		ENCODE
ChIP-seq	chr1:206544749..207358622	H3k27ac	Dnd41		ENCODE
ChIP-seq	chr1:206553433..207044290	H4k20me1	Hsimm		ENCODE
ChIP-seq	chr1:206557090..207378901	H4k20me1	Helas3		ENCODE
ChIP-seq	chr1:206557874..206988194	H3k27ac	H1esc		ENCODE

Handout



RegulomeDB Disease Database



<http://regulome.stanford.edu/GWAS>; Schaub...Snyder, Genome Research 22-1748,2012

RegulomeDB About Help

Linking Disease Associations with Regulatory Information in the Human Genome

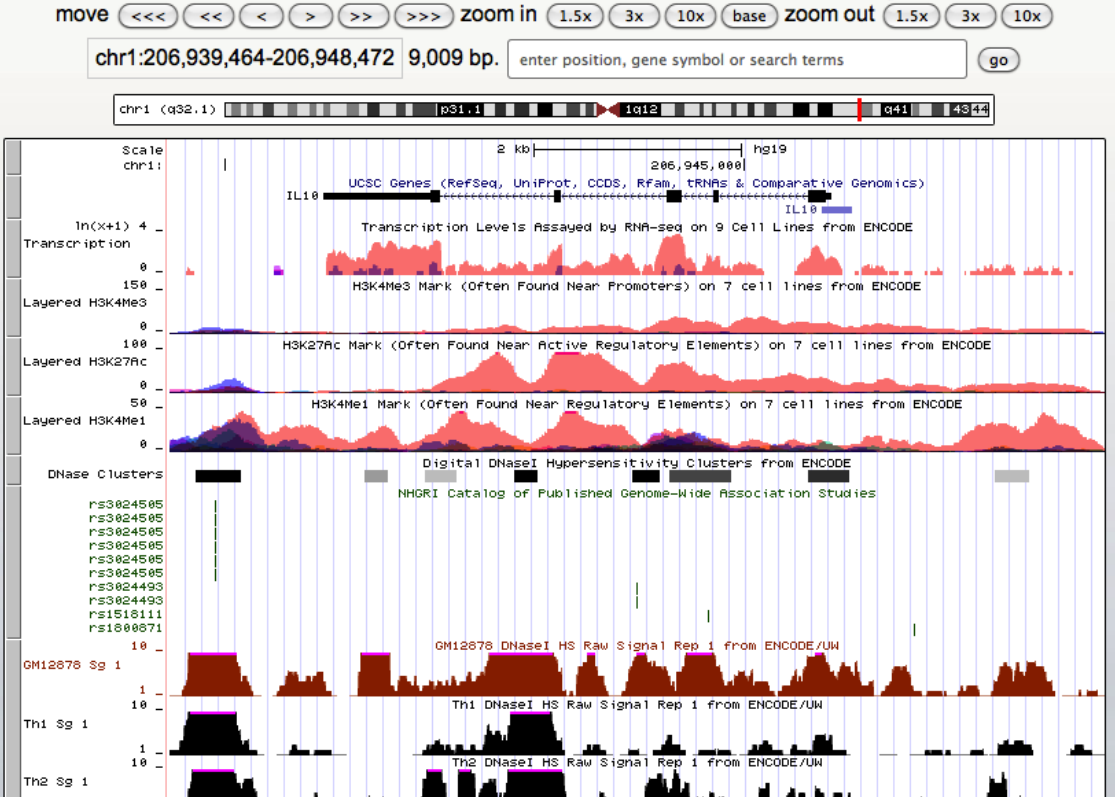
Companion website

Marc A. Schaub
Access the list of GWAS associations

- List of all associated SNPs
- By phenotype:
 - 5-HTT brain serotonin transporter
 - AB1-42
 - AIDS
 - AIDS progression
 - Abdominal aortic aneurysm
 - Acenocoumarol maintenance dose
 - Activated partial thromboplastin time
 - Acute lymphoblastic leukemia (ALL)
 - Adiponectin levels
 - Adiposity
 - Adverse response to aromatase inhibitor
 - Adverse response to carbamazepine
 - Age-related macular degeneration
 - Age-related macular degeneration
 - Aging
 - Aging traits
 - Alcohol consumption
 - Alcohol dependence
 - Alcoholism (12-month weekly alcohol consumption)
 - Alcoholism (alcohol dependence)
 - Alcoholism (alcohol use disorder)
 - Alcoholism (heaviness of drinking)
 - Alopecia areata
 - Alzheimer's disease
 - Alzheimer's disease (late onset)

Lead SNP
rs3024505
Position: chr1:206,939,904 ([Open in UCSC](#))
Distance to nearest TSS: 18,466 bp
GENCODE v7 location: Intergenic region
RegulomeDB Score: 2b - ChIP-seq
Linkage disequilibrium
Linkage disequilibrium threshold:
- In all HapMap 2 populations: $r^2 \geq 0.8$
- In the HapMap 2 CEU population: $r^2 \geq 0.8$
SNPs in the linkage disequilibrium
rs3024493
Position: chr1:206,943,968 ([Open in UCSC](#))
Distance to lead SNP: 4,064 bp
Distance to nearest TSS: 22,530 bp
GENCODE v7 location: Intron
RegulomeDB Score: 2b - ChIP-seq
Linkage disequilibrium with Lead SNP
rs3024493
Position: chr1:206,942,413 ([Open in UCSC](#))
Distance to lead SNP: 2,509 bp
Distance to nearest TSS: 20,975 bp
GENCODE v7 location: Intron
RegulomeDB Score: 5a - ChIP-seq
Linkage disequilibrium with Lead SNP

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





How Can The Display Be Configured?



Home Browser Tools Mirrors Downloads My Data About Us View Help

UCSC Genome Browser on hg19 (chr37/hg19) Assembly

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

chr1:206,939,569-206,946,986 7,418 bp. Search terms go

chr1 (q32.1) pS1.1 lq12 q41 4344

Scale 2 kb hg19
chr1: 206,941,000 | 206,942,000 | 206,943,000 | 206,944,000 | 206,945,000 | 206,946,000

UCSC Genes (RefSeq, UniProt, CCDS, Rfam, tRNAs & Comparative Genomics)
IL18 IL18

Transcription Levels Assayed by RNA-seq on 9 Cell Lines from ENCODE

H3K4Me3 Mark (Often Found Near Promoters) on 7 cell lines from ENCODE

H3K4Me1 Mark (Often Found Near Regulatory Elements) on 7 cell lines from ENCODE

H3K27Ac Mark (Often Found Near Active Regulatory Elements) on 7 cell lines from ENCODE

Digital DNaseI Hypersensitivity Clusters from ENCODE

Transcription Factor ChIP-seq from ENCODE

NHGRI Catalog of Published Genome-Wide Association Studies

Chromatin State Segmentation by HMM from ENCODE/Broad

GM12878 ChromHMM
H1-hESC ChromHMM
K562 ChromHMM
HepG2 ChromHMM
HMEC ChromHMM
HSMH ChromHMM
HUVEC ChromHMM
NHEK ChromHMM
NHLF ChromHMM

Common SNPs (135)
Simple Nucleotide Polymorphisms (dbSNP 135) Found in >= 1% of Samples

DNaseI Digital Genomic Footprinting from ENCODE/University of Washington

Th1 Sig 1
Th2 Sig
Th17 Sig
Treg Sig

move start Click on a feature for details. Click or drag in the base position track to zoom in. Click side bars for track options. Drag side bars or labels up or down to reorder tracks. Drag tracks left or right to new position. move end

< 2.0 >

track search default tracks default order hide all add custom tracks track hubs configure reverse resize refresh

collapse all Use drop-down controls below and press refresh to alter tracks displayed. Tracks with lots of items will automatically be displayed in more compact modes. expand all

Handout



ENCODE Experiment Matrix



Encyclopedia of DNA Elements

		DNA Methylation		Open Chromatin			RNA Binding Proteins				RNA Profiling					TFBS & Histones		Other							
		Methyl Array	Methyl RRBS	DNase-DGF	DNase-seq	FAIRE-seq	RIP Gene ST	RIP Tiling Array	RIP Validation	RIP-seq	CAGE	Exon Array	RNA-chip	RNA-PET	RNA-seq	Small RNA-seq	ChIP-seq	view matrix	5C	ChIA-PET	Combined	DNA-PET	Genotype		
Human	Integrative Analysis	search for: <input type="radio"/> tracks <input type="radio"/> files																							
	Experiment Matrix																								
	Experiment List	Cell Types																							
	Search	Tier 1																							
	Downloads	GM12878	1	1		2	1	7	4		4	6	2	6	2	12	5	133		2		2	3	1	
		H1-hESC	1	1		2	1	3				4	1		1	10	3	91		1		2		1	
		K562	1	1	3	16	3	6	4		4	9	7	9	6	17	7	224		2	2	2	3	1	
		Tier 2																							
		A549	1	1	1	2	1					3	2		3	10	9	87						1	
		CD20+										1				2	1	4							
	CD20+_RO01778			1	1												2								
	CD20+_RO01794				1												5								
	H1-neurons																								
	H1-neurons														3		4								
Mouse	HeLa-S3	1	1		3	3	4				6	4		3	8	3	93		1	1	2		1		
	HepG2	1	1	1	2	1	4				6	2	5	2	8	3	114		1		2		1		
	HUVEC	1		1	2	1					5	2		2	8	1	36				2		1		
	IMR90	1	1		1						3			3	4	9	11						1		
	LHCN-M2			2	2										2		7								

<http://encodeproject.org>



How Can ENCODE Data Be Downloaded?



ENCODE Data Coordination Center at

[Home](#) - [Downloads](#) - [Data Policy](#) - [Help](#)

ENCODE Downloads

This page contains links to directories containing raw and processed data for ENCODE data release (present). Formats and data release policy. it has not yet been r

Human Genom

- All-Exon Arrays**
 - Duke Affy Exon
 - UW Affy Exon
- Chromatin Access**
 - Duke DNaseI HS
 - GIS ChIA-PET
 - Open Chrom Synt
 - UNC FAIRE
 - UW DNaseI DGF
 - UW DNaseI HS
- Chromatin Interac**
 - UMass 5C
 - UW 5C
- DNA Methylation**
 - HAIB Methyl RRBS
 - HAIR Methyl450

Parent Directory

files.txt	21-Sep-2012 10:04	1.8M
md5sum.txt	15-Aug-2012 16:44	274K
supplemental/	17-Aug-2012 16:10	-
wgEncodeSydhTfbsA549Bhlhe40IqgrabAlnRepl.bam	02-Feb-2012 20:37	585M
wgEncodeSydhTfbsA549Bhlhe40IqgrabAlnRepl.bam.bai	03-Feb-2012 20:07	6.0M
wgEncodeSydhTfbsA549Bhlhe40IqgrabAlnRep2.bam	02-Feb-2012 19:46	489M
wgEncodeSydhTfbsA549Bhlhe40IqgrabAlnRep2.bam.bai	03-Feb-2012 20:10	5.9M
wgEncodeSydhTfbsA549Bhlhe40IqgrabPk.narrowPeak.gz	03-Feb-2012 20:12	112K
wgEncodeSydhTfbsA549Bhlhe40IqgrabRawDataRepl.fastq.tgz	03-Feb-2012 20:20	1.0G
wgEncodeSydhTfbsA549Bhlhe40IqgrabRawDataRep2.fastq.tgz	03-Feb-2012 20:17	860M
wgEncodeSydhTfbsA549Bhlhe40IqgrabSig.bigWig	02-Feb-2012 19:09	384M
wgEncodeSydhTfbsA549CebpbIqgrabAlnRepl.bam	02-Feb-2012 21:03	589M
wgEncodeSydhTfbsA549CebpbIqgrabAlnRepl.bam.bai	03-Feb-2012 20:06	6.0M
wgEncodeSydhTfbsA549CebpbIqgrabAlnRep2.bam	02-Feb-2012 20:00	450M
wgEncodeSydhTfbsA549CebpbIqgrabAlnRep2.bam.bai	03-Feb-2012 20:10	5.9M
wgEncodeSydhTfbsA549CebpbIqgrabPk.narrowPeak.gz	03-Feb-2012 20:17	1.7M
wgEncodeSydhTfbsA549CebpbIqgrabRawDataRepl.fastq.tgz	03-Feb-2012 20:17	1.0G
wgEncodeSydhTfbsA549CebpbIqgrabRawDataRep2.fastq.tgz	03-Feb-2012 20:20	781M
wgEncodeSydhTfbsA549CebpbIqgrabSig.bigWig	02-Feb-2012 19:18	387M
wgEncodeSydhTfbsA549CmycIqgrabAlnRepl.bam	31-Jul-2012 13:40	1.1G
wgEncodeSydhTfbsA549CmycIqgrabAlnRepl.bam.bai	31-Jul-2012 20:19	6.1M
wgEncodeSydhTfbsA549CmycIqgrabAlnRep2.bam	31-Jul-2012 13:41	1.3G
wgEncodeSydhTfbsA549CmycIqgrabAlnRep2.bam.bai	31-Jul-2012 20:26	6.2M
wgEncodeSydhTfbsA549CmycIqgrabPk.narrowPeak.gz	31-Jul-2012 20:43	127K
wgEncodeSydhTfbsA549CmycIqgrabRawDataRepl.fastq.tgz	31-Jul-2012 20:43	891M
wgEncodeSydhTfbsA549CmycIqgrabRawDataRep2.fastq.tgz	31-Jul-2012 20:36	1.0G
wgEncodeSydhTfbsA549CmycIqgrabSig.bigWig	31-Jul-2012 17:35	346M



Human

[Integrative Analysis](#)

[Experiment Matrix](#)

[Experiment List](#)

[Search](#)

Downloads

[Genome Browser \(hg19\)](#)

[Session Gallery](#)

[Cell Types](#)

Mouse

[Experiment Matrix](#)

<http://encodeproject.org>



Publications



General

Registered Variables

Antibodies

Education and Outreach

Release Log

Data Policy

Data Standards

File Formats

Software Tools

Publications

Contributors

Pilot Project

Contacts

To supplement the existing ENCODE 1 tutorial, OpenHelix has developed a second tutorial. This tutorial is freely available from the website at: <http://openhelix.com/ENCODE2>. Also available is a *Reference Card* that provides a handy summary of access features. For more information see the *ENCODE Reference Card*.

27 September 2012 - ENCODE data releases: UNC/BSU ProtGenc, UNC FAIRE (Rel 2), HAIB

One new track and three track updates were released on the human hg19 browser:

[Home](#) - [Downloads](#) - [Data Policy](#) - [Data Standards](#)



ENCODE-funded Publications

This page lists publications funded, at least in part, by ENCODE funds. The process of cataloging these publications is ongoing and updates to this page will be made periodically.

Publications from non-ENCODE Authors

In addition to tracking ENCODE-funded publications, the ENCODE project also tracks papers that were not published by ENCODE authors and were not funded by ENCODE, but that reference ENCODE data. This is being done in part to see how the resource is being used. The papers are grouped in [Human disease](#), [Basic biology](#), and [Tools/methods/databases/commentaries](#) sections. As there is no systematic way to search for these publications this list is likely incomplete.

Please contact [Mike Pazin at NHGRI](#) to suggest publications to add to this list.

Updated 24 August 2012

<http://encodeproject.org>



Over 170 Publications Using ENCODE Data

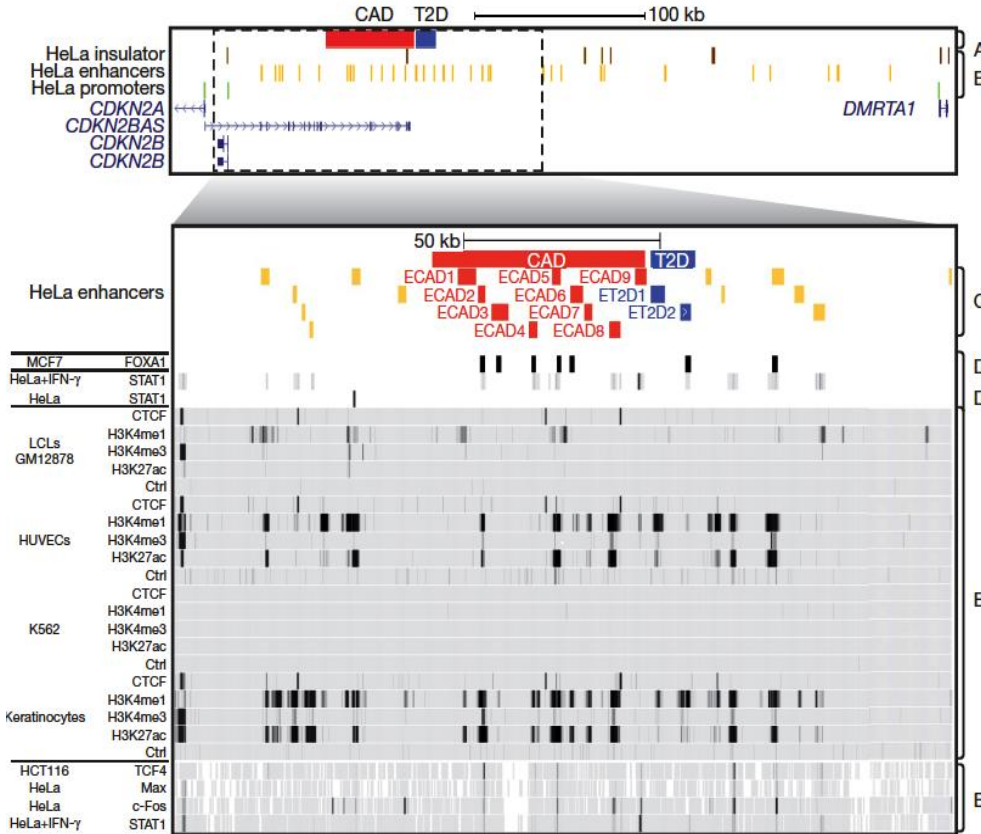


Table 1 PBC risk loci at genome-wide significance

Chr.	SNP ^a	RA ^b	RAF ^c	P ^d	OF	e)	Nearby gene(s) ^f	Functional annotation ^g
1p31	rs72678531	G	0.17	2.47 × 10 ⁻³⁸	1.61 ((31,842)	<i>IL12RB2</i>	OC
1q31	rs2488393	A	0.21	4.29 × 10 ⁻¹²	1.28 (:1 (438,818)	<i>DENND1B</i>	
2q32	rs3024921	A	0.06	2.59 × 10 ⁻¹⁸	1.62 (.7 (0)	<i>STAT1, STAT4</i>	
2q32	Second signal rs7574865	A	0.22	1.38 × 10 ⁻¹³	1.31 ('9 (29,292)	<i>STAT1, STAT4</i>	
3q13	rs2293370	G	0.8	6.84 × 10 ⁻¹⁶	1.39 ('8 (136,058)	<i>TMEM39A, POGUT1, TIMMD1, CD80</i>	NS
3q25	rs2366643	A	0.57	3.92 × 10 ⁻²²	1.35 ('0 (16,805)	<i>IL12A</i>	OC
3q25	Second signal rs62270414	G	0.15	5.74 × 10 ⁻¹⁷	1.41 ('6 (52,623)	<i>IL12A</i>	OC
3q25	Third signal rs668998	G	0.43	4.73 × 10 ⁻⁹	1.26 (15 (5,550)	<i>IL12A</i>	OC
3q25	Fourth signal rs80014155	A	0.004	2.64 × 10 ⁻¹¹	3.44 (17 (68,660)	<i>IL12A</i>	
4q24	rs7665090	G	0.52	8.48 × 10 ⁻¹⁴	1.26 (i1 (0)	<i>MANBA, NFKB1</i>	NS, eQTL
5p13	rs6871748	A	0.72	2.26 × 10 ⁻¹³	1.3 (1	(35,833)	<i>IL7R, CAPSL, SFEF2, UGT3A1</i>	NS
7q32	rs35188261	A	0.17	6.52 × 10 ⁻²²	1.52 (.0 (126,258)	<i>IRF5, TNPO3</i>	OC
7q32	Second signal rs3807307	G	0.47	4.12 × 10 ⁻⁹	1.22 (.6 (6,713)	<i>IRF5, TNPO3</i>	OC, eQTL
11q23	rs80065107	A	0.79	7.20 × 10 ⁻¹⁶	1.39 (i2 (133,223)	<i>DDX6</i>	OC
12p13	rs1800693	G	0.4	1.18 × 10 ⁻¹⁴	1.27 (!,802)	<i>TNFRSF1A, LTBR, SCNN1A</i>	OC
12p13	Second signal rs11064157	A	0.25	1.69 × 10 ⁻⁹	1.23 (<i>TNFRSF1A, LTBR, SCNN1A</i>	OC
12q24	rs11065979	A	0.44	2.87 × 10 ⁻⁹	1.2 (1	17 (726,106)	<i>ATXN2, BRAP, SH2B3</i>	NS
14q24	rs911263	A	0.71	9.95 × 10 ⁻¹¹	1.26 ((0)	<i>RAD51B</i>	
16p13	rs1646019	G	0.71	6.72 × 10 ⁻¹⁵	1.31 ((18,452)	<i>SOCS1, CLEC16A, PRM1, PRM2</i>	OC, eQTL
16p13	Second signal rs12708715	C	0.68	2.19 × 10 ⁻¹³	1.29 ((118,128)	<i>SOCS1, CLEC16A, PRM1, PRM2</i>	OC, eQTL
16p13	Third signal rs80073729	A	0.004	2.69 × 10 ⁻⁸	2.96 ((0)	<i>SOCS1, CLEC16A, PRM1, PRM2</i>	OC
16q24	rs11117433	G	0.77	1.41 × 10 ⁻⁹	1.26 ((0)	<i>IRFB</i>	OC
17q12	rs8067378	G	0.52	6.05 × 10 ⁻¹⁴	1.26 ((177,700)	<i>ORMDL3, ZPBP2, GSDMB, IKZF3</i>	NS, OC, eQTL
17q21	rs17564829	G	0.24	2.15 × 10 ⁻⁹	1.25 ((1,164,644)	<i>CRHR1, MAPT</i>	NS, OC, eQTL
19p12	rs34536443	G	0.95	1.23 × 10 ⁻¹²	1.91 ((0)	<i>TYK2</i>	NS
22q13	rs2267407	A	0.23	1.29 × 10 ⁻¹³	1.29 ((9,600)	<i>SYNGR1, PDGFB, RPL3</i>	OC, eQTL

Frazer lab, Nature 470-264,2011

Anderson lab, Nature Genetics 44-1137,2012



ENCODE Software Tools



General

Registered Variables

Antibodies

Education and Outreach

Release Log

Data Policy

Data Standards

File Formats

Software Tools

Publications

Contributors

Pilot Projects

Contacts

To supplement the existing ENCODE 1 tutorial, OpenHelix has developed a second tutorial. This tutorial is freely available from the website at: <http://openhelix.com/ENCODE2>. Also available is a *Reference Card* that provides a handy summary of access features. For more information see the *ENCODE Reference Card*.

27 October 2010 ENCODE 1 Tutorial - ENCODE 2 Tutorial - ENCODE Reference Card

[Home](#) - [Downloads](#) - [Data Policy](#) - [Data Standards](#)



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.

[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. Software for identifying functional elements, for integrated analysis of multiple data types, and for measuring the quality of the data are described.

[Software and Resources for 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. Two pieces of software from ENCODE researchers for annotating non-coding regions of the genome (such as SNPs) with features from ENCODE and other resources are presented, as well as a recent chapter in *Current Protocols in Human Genetics* explaining how to manually inspect genomic regions of interest for ENCODE annotations.

<http://encodeproject.org>



ENCODE Citation

General

Registered Variables

Antibodies

Education and Outreach

Release Log

Data Policy

Data Standards

File Formats

Software Tools

Publications

Contributors

Pilot Project

Contacts

To supplement the existing ENCODE 1.0 citations tutorial, OpenHelix has developed a second tutorial. This tutorial is freely available from the website at: <http://openhelix.com/ENCODE2>. Also available is a *Reference Card* that provides a handy summary of access features. For more information see the *Citation*

07/2011 - 0040 ENCODE Data Release Policy - UNCORRECTED - UNCORRECTED / RELEASE / 07/2011



ENCODE Data Coordination Center at UCSC

Home - Help

ENCODE Consortium Data Release Policy Summary

Please observe the following guidelines when using ENCODE data:

- Data users may freely download and analyze ENCODE data without restrictions. They may use ENCODE data in publications focused around individual genes; however, there is a narrow 9-month "moratorium" on the publication of global analysis of ENCODE data sets to allow the data producers the opportunity to publish first if they wish. Following the expiration of the moratorium period or publication by the data producers (whichever is first), publication of global analyses is unrestricted. Most released datasets are already outside the moratorium. See the [Data Summary](#) and [Mouse Data Summary](#) pages for the restriction times of individual data sets.
- The publication and presentation moratorium is expected to extend to all forms of public disclosure, including meeting abstracts, oral presentations, and formal electronic submissions to publicly accessible sites (e.g., public websites, web blogs).
- Resource users are expected to acknowledge the following in all oral or written presentations, disclosures, or publications of the analyses:
 - The resource producers
 - The following publication:
 - ENCODE Project Consortium, Myers RM, Stamatoyannopoulos J, Snyder M, Dunham I, Hardison RC, Bernstein BE, Gingeras TR, Kent WJ, Birney E *et al.* [A user's guide to the encyclopedia of DNA elements \(ENCODE\)](#). *PLoS Biol.* 2011 Apr;9(4):e1001046. Epub 2011 Apr 19. PMID: 21526222; PMCID: PMC3079585
 - The funding organization(s) that supported the work
 - The respective DCC
- Data users should properly acknowledge the ENCODE Project and resource producer(s) as the source of the data in any publication.
- See the full [ENCODE-modENCODE Data Release Policy \(2009-Present\)](#) document for further details, and the [ENCODE-modENCODE Data Release Policy](#) description page at NHGRI for background.



ENCODE Consortium



Brad Bernstein (Eric Lander, Manolis Kellis/**Luke Ward**, Tony Kouzarides)

Ewan Birney (Jim Kent, Mark Gerstein, Bill Noble, Peter Bickel, Ross Hardison, Zhiping Weng)

Greg Crawford (Ewan Birney, Jason Lieb, Terry Furey, Vishy Iyer)

Jim Kent (David Haussler, **Kate Rosenbloom**)

John Stamatoyannopoulos (Evan Eichler, George Stamatoyannopoulos, Job Dekker, Maynard Olson, Michael Dorschner, Patrick Navas, Phil Green)

Mike Snyder (Kevin Struhl, Mark Gerstein, Peggy Farnham, Sherman Weissman)

Rick Myers (Barbara Wold)

Scott Tenenbaum (Luiz Penalva)

Tim Hubbard (Alexandre Reymond, Alfonso Valencia, David Haussler, Ewan Birney, Jim Kent, Manolis Kellis, Mark Gerstein, Michael Brent, Roderic Guigo)

Tom Gingeras (Alexandre Reymond, David Spector, Greg Hannon, Michael Brent, Roderic Guigo, Stylianos Antonarakis, Yijun Ruan, Yoshihide Hayashizaki)

Zhiping Weng (Nathan Trinklein, Rick Myers)

NHGRI: Elise Feingold, Peter Good, Laura Dillon, Rebecca Lowdon, Leslie Adams, Caroline Kelly, Shaila Chhibba, Sherry Zhou, Katya Vaydylevich

Additional ENCODE Participants: Elliott Marguiles, Eric Green, Job Dekker, Laura Elnitski, Len Pennachio, Jochen Wittbrodt

.. and many senior scientists, postdocs, students, technicians, computer scientists, statisticians and administrators in these groups

Questions? NHGRI booth (# 825) noon- 1 PM; pazinm@mail.nih.gov





What Do The Data Mean?



- Some standard interpretations-
 - RNA
 - Histone modifications
 - DNase
 - Transcription Factor ChIP



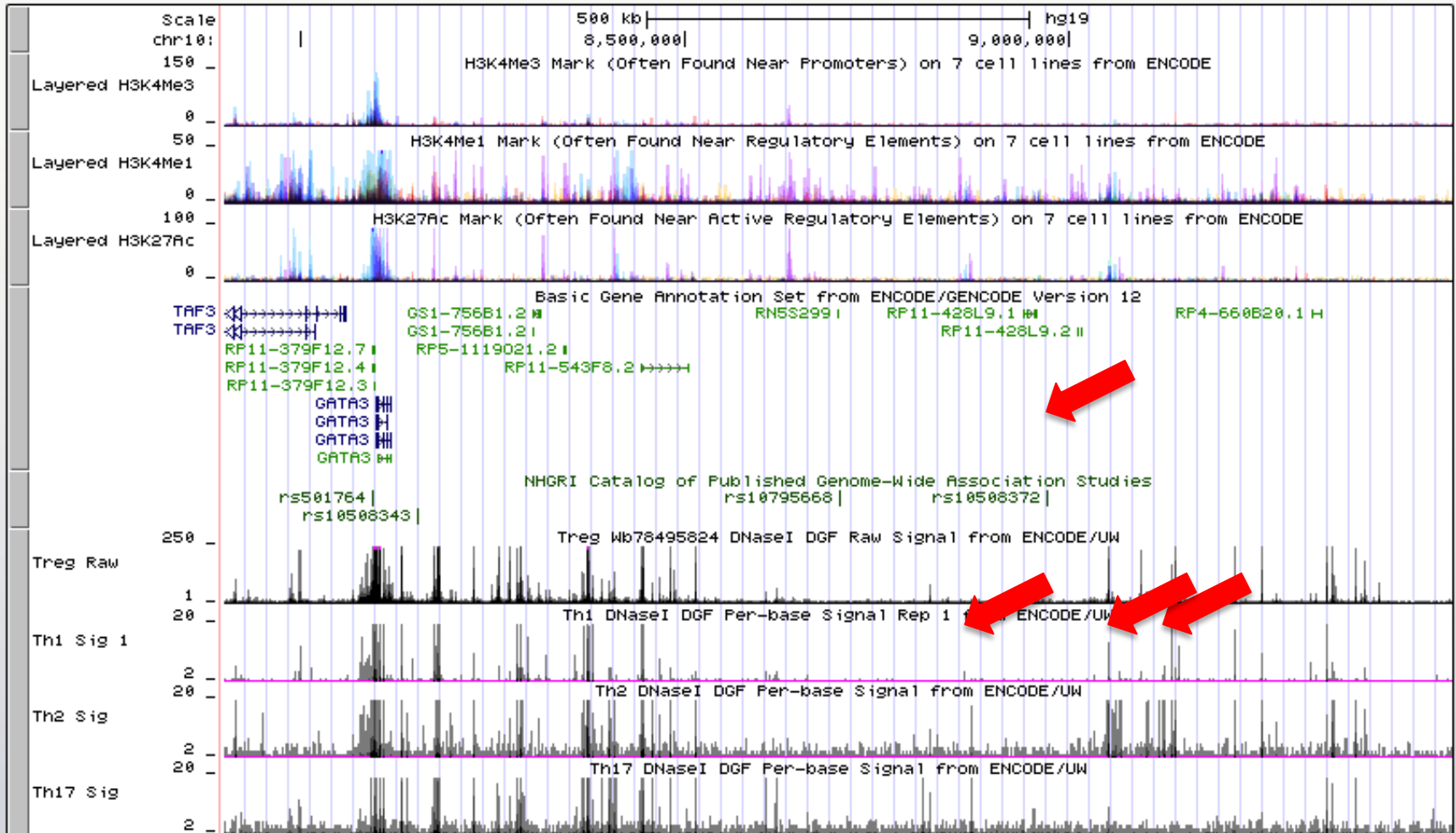
How Can ENCODE Data Be Used?



- A standard problem: many genetic findings for human disease map to non-protein coding regions of the human genome
 - What is the functional variant?
 - What is the target gene?
 - What is the target cell type?
 - What is the function of the variant?
- Standard Browser view for loci of interest
- HaploReg and RegulomeDB searches for loci of interest
- Search a cell type across all data types for loci of interest
- Search a data type across all cell types for loci of interest



ENCODE Browser, Cells of Interest



<http://encodeproject.org>



NIH Roadmap

Epigenomic Mapping Consortium



A public community resource of epigenomic data in primary human cells and tissues.

Range of cells/tissues covered:

Currently 125 cell/tissue types represented including....
iPS and ES cells, some differentiated forms
Fetal tissues (heart, brain, kidney, lung, others)
Adult primary cells and tissues (hematopoietic, brain regions, breast cell types, liver, kidney, colon, muscle, adipocytes, others)

Some samples will have:

Expanded panel of histone modifications (currently 20 additional)

Most samples will have:

DNA methylation data (RRBS, MRE-seq, MeDIP-seq, whole genome bisulfite seq)
ChIP-seq data (currently H3K27me3, H3K36me3, H3K4me1, H3K4me3, H3K9me3)
DNase I hypersensitivity data
Gene expression data (arrays or RNA-seq)

Can download:

.wig, .bed, some .bam, some SRA, working on peak calls

The screenshot shows the Roadmap Epigenomics Project website. At the top, there is a navigation bar with tabs for HOME, PARTICIPANTS, DATA, PROTOCOLS, QUALITY METRICS, TOOLS, and PUBLICATIONS. Below this is a search bar with a 'GO' button. The main content area is divided into several sections: OVERVIEW, PROJECT DATA, MAPPING CENTERS, PROTOCOLS & STANDARDS, PUBLICATIONS, and NEWS. The PROJECT DATA section features a large diagram illustrating various epigenomic features: Chromatin, RNA, DNA methylation, DNase I hypersensitive sites, and Histone Modifications. To the right of the diagram is a 'VIEW/DOWNLOAD QUICK LINKS' section with 'UCSC Browser Mirrors' and 'Data Repositories'. Below that is a 'NEWS' section with a date of 27 MAY and a headline 'Disease Risk Links to Gene Regulation'. At the bottom of the page, there is a brief description of the NIH Roadmap Epigenomics Mapping Consortium.

<http://roadmapepigenomics.org> – Find and view data, protocols, links to other sites associated with the program.

The screenshot shows the UCSC Genome Browser Gateway interface. At the top, there is a navigation bar with tabs for Home, Genomes, Blat, Tables, Gene Sorter, PCR, Session, FAQ, and Help. Below this is the title 'Human (Homo sapiens) Genome Browser Gateway'. The main content area contains a search form with fields for 'clade', 'genome', 'assembly', 'position or search term', 'gene', and 'image width'. The 'clade' is set to 'Mammal', 'genome' to 'Human', 'assembly' to 'Feb 2009 (GRCh37/hg19)', and 'position or search term' to 'chrX:70,752,933-70,795,740'. There is a 'submit' button. Below the search form, there is a link to 'Click here to reset the browser user interface settings to their defaults.' and a row of buttons: 'track search', 'add custom tracks', 'track hubs', 'configure tracks and display', and 'clear position'.

View Roadmap data at <http://genome.ucsc.edu> via TRACK HUB – click ‘track hubs’ to load Roadmap data and summary data tracks on UCSC.

Where is the data?

Find data, protocols, and analysis/viewing tools from the Roadmap Epigenomic Mapping Consortium at these sites:

- <http://roadmapepigenomics.org>
- <http://ncbi.nlm.nih.gov/epigenomics>
- <http://ncbi.nlm.nih.gov/geo/roadmap/epigenomics>
- <http://epigenomeatlas.org>
- <http://vizhub.wustl.edu>
- <http://genome.ucsc.edu> (via track hub)



Recent ENCODE Publications

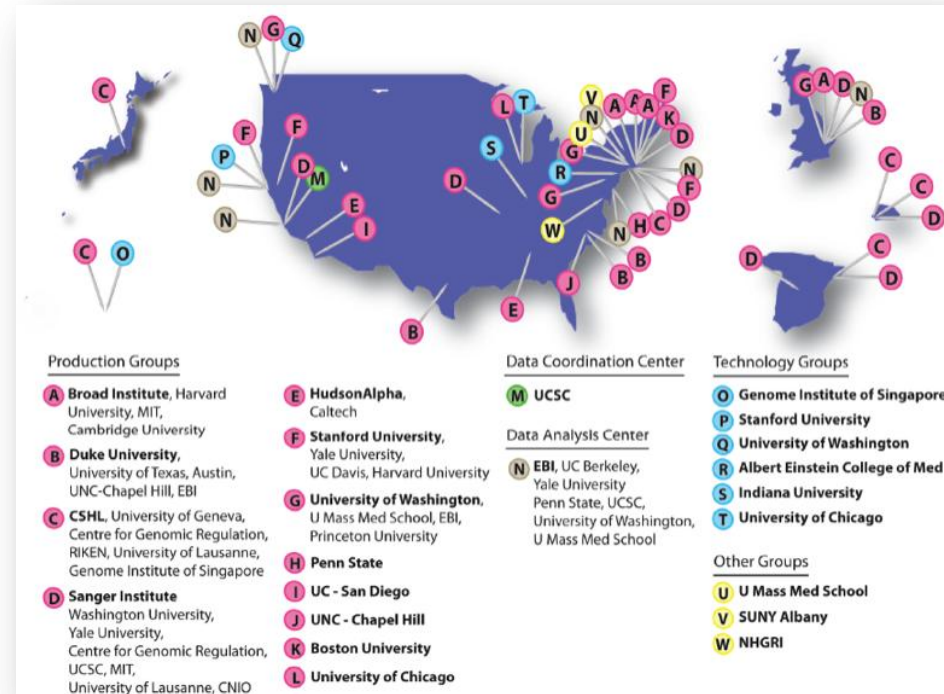
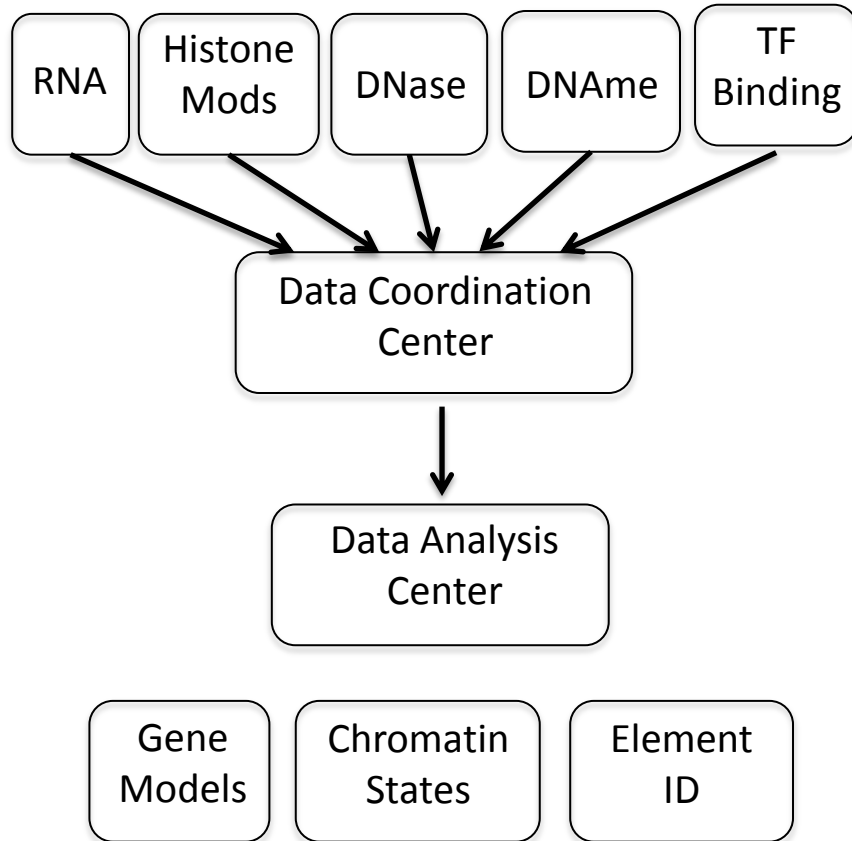


The image displays two screenshots of the Nature ENCODE explorer mobile application. The top screenshot shows the main interface with a 'Threads explained' section on the left and a 'Papers' section on the right. The bottom screenshot shows a detailed view of 'Thread 12 of 13' titled 'Impact of functional information on understanding variation'. This view includes a 'Read Thread' button and a diagram of a chromosome with highlighted regions. The interface is dark-themed and includes a 'Back' button and a 'Reset' button.

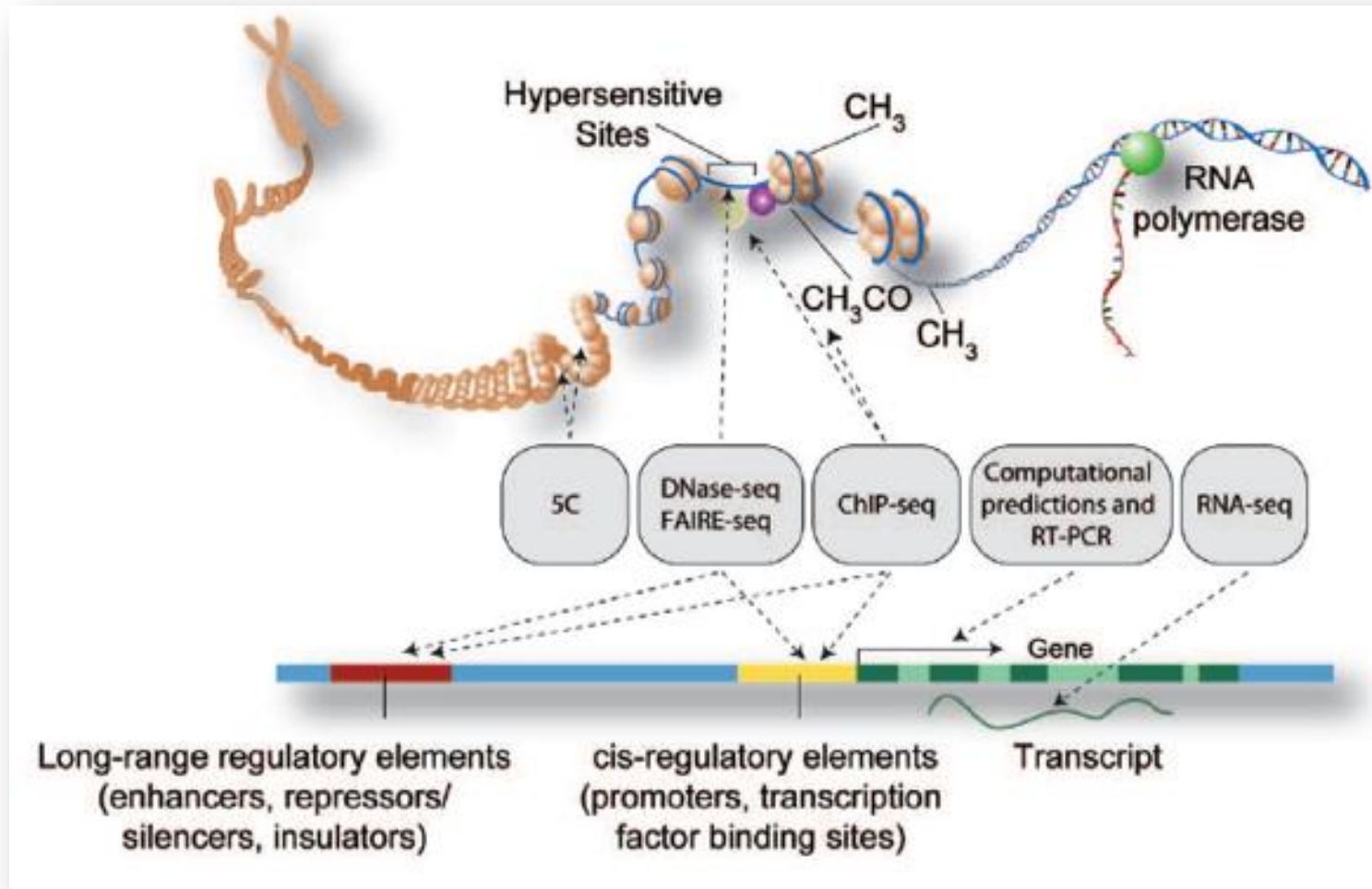
- More than 30 papers in Nature, Genome Research, Genome Biology, Science, Cell
- Publishing innovations
- ENCODE increased our understanding of non-coding DNA, and human disease



ENCODE Structure

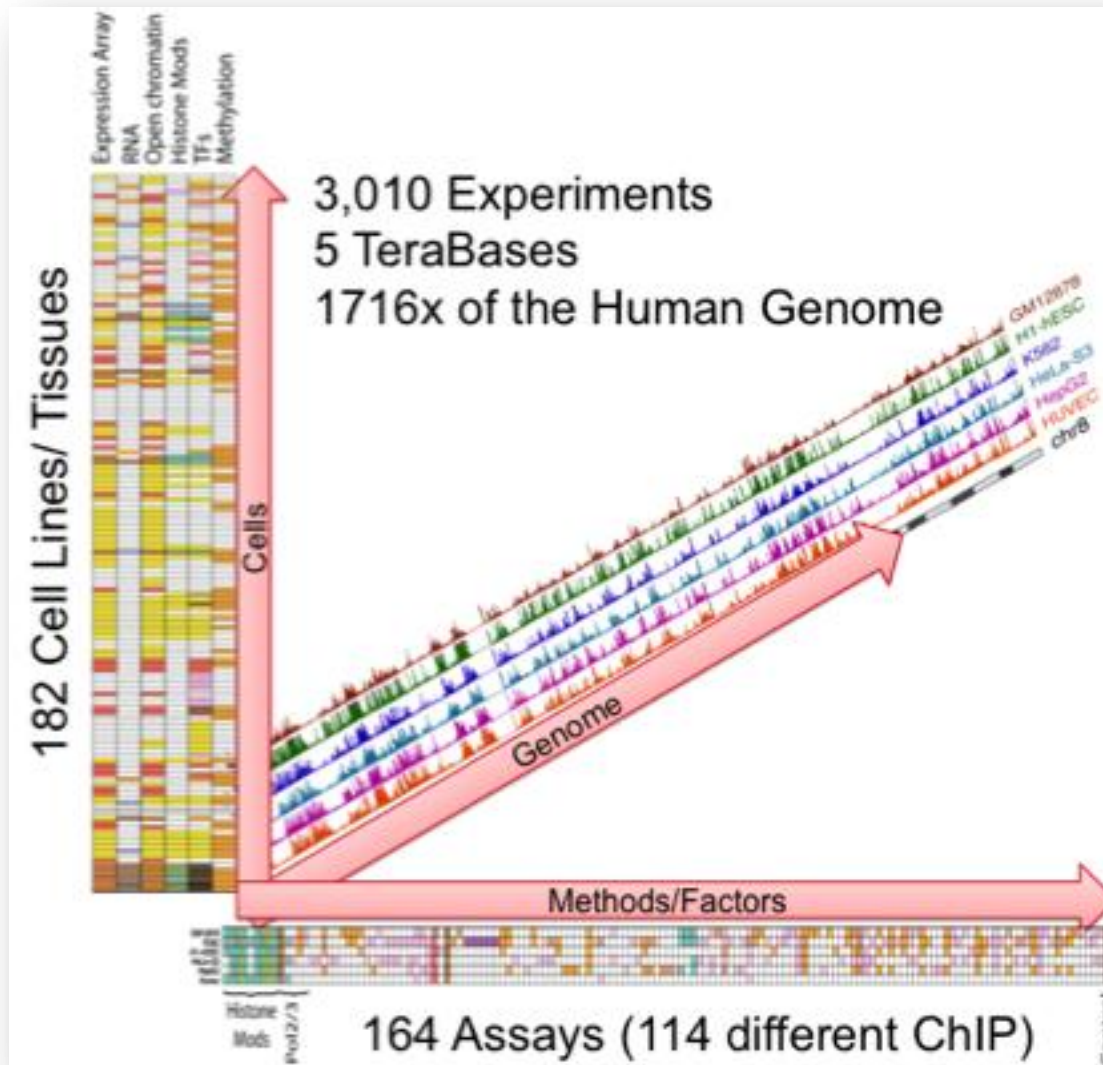


What Data Are Available?





ENCODE Dimensions



From Ewan Birney



Raw Genomic Coverage of Elements



Biochemical Mark	Genomic Coverage
Any ENCODE mark	80 %
Any ENCODE RNA	62 %
Any Histone modification	56 %
DHS or TF ChIP	19.4 %
Open chromatin	15.2 %
TF ChIP	8.1 %
DHS footprint	5.7 %
Purifying selection	~3-8 %
Exons (GENCODE)	2.94 %
Protein-coding regions (GENCODE)	1.22 %