



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



Display of ENCODE Data

General
Resources & FAQ *New*
Publications
Software Tools
Data Standards
Human
Downloads
Experiment Matrix
Search
Genome Browser (hg19) 
Integrative Analysis

About 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

chr1:206,850,000-207,050,000 200,001 bp. enter position, gene symbol or search terms go

chr1 (q32.1) p31.1 q12 q41 4344

Scale chr1: | 100 kb | hg19
206,900,000 | 206,950,000 | 207,000,000

Basic Gene Annotation Set from ENCODE/GENCODE Version 14
DYRK3 SNORD112 MAPKAPK2 Y_RNA1 IL10 IL19 IL20 IL20 IL20

Transcription
ln(x+1) 8
Transcription Levels Assayed by RNA-seq on 9 Cell Lines from ENCODE

Layered H3K4Me3
150
H3K4Me3 Mark (Often Found Near Promoters) on 7 cell lines from ENCODE

Layered H3K4Me1
50
H3K4Me1 Mark (Often Found Near Regulatory Elements) on 7 cell lines from ENCODE

Layered H3K27Ac
100
H3K27Ac Mark (Often Found Near Active Regulatory Elements) on 7 cell lines from ENCODE

Txn Factor ChIP
Transcription Factor ChIP-seq from ENCODE

DNase Clusters
Digital DNaseI Hypersensitivity Clusters in 125 cell types from ENCODE

GWAS Catalog
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
HSMM ChromHMM
HUVEC ChromHMM
NHEK ChromHMM
NHLF ChromHMM

<http://encodeproject.org>



Configuring The Display



UCSC Genome Browser

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

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

chr1 (q32.1) | p31.1 | q12 | q41 | 43 | 44

Scale 2 kb hg19

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

IL10

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

Layered H3K4Me3 HSK4Me3 Mark (Often Found Near Promoters) on 7 cell lines from ENCODE

Layered H3K4Me1 HSK4Me1 Mark (Often Found Near Regulatory Elements) on 7 cell lines from ENCODE

Layered H3K27Ac HSK27Ac Mark (Often Found Near Active Regulatory Elements) on 7 cell lines from ENCODE

DNase Clusters Digital DNaseI Hypersensitivity Clusters from ENCODE

Txn Factor ChIP Transcription Factor CHIP-seq from ENCODE

GWAS Catalog 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 HSMM 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

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



ENCODE Experiment Matrix



About ENCODE Data

Search for: tracks files

Cell Types

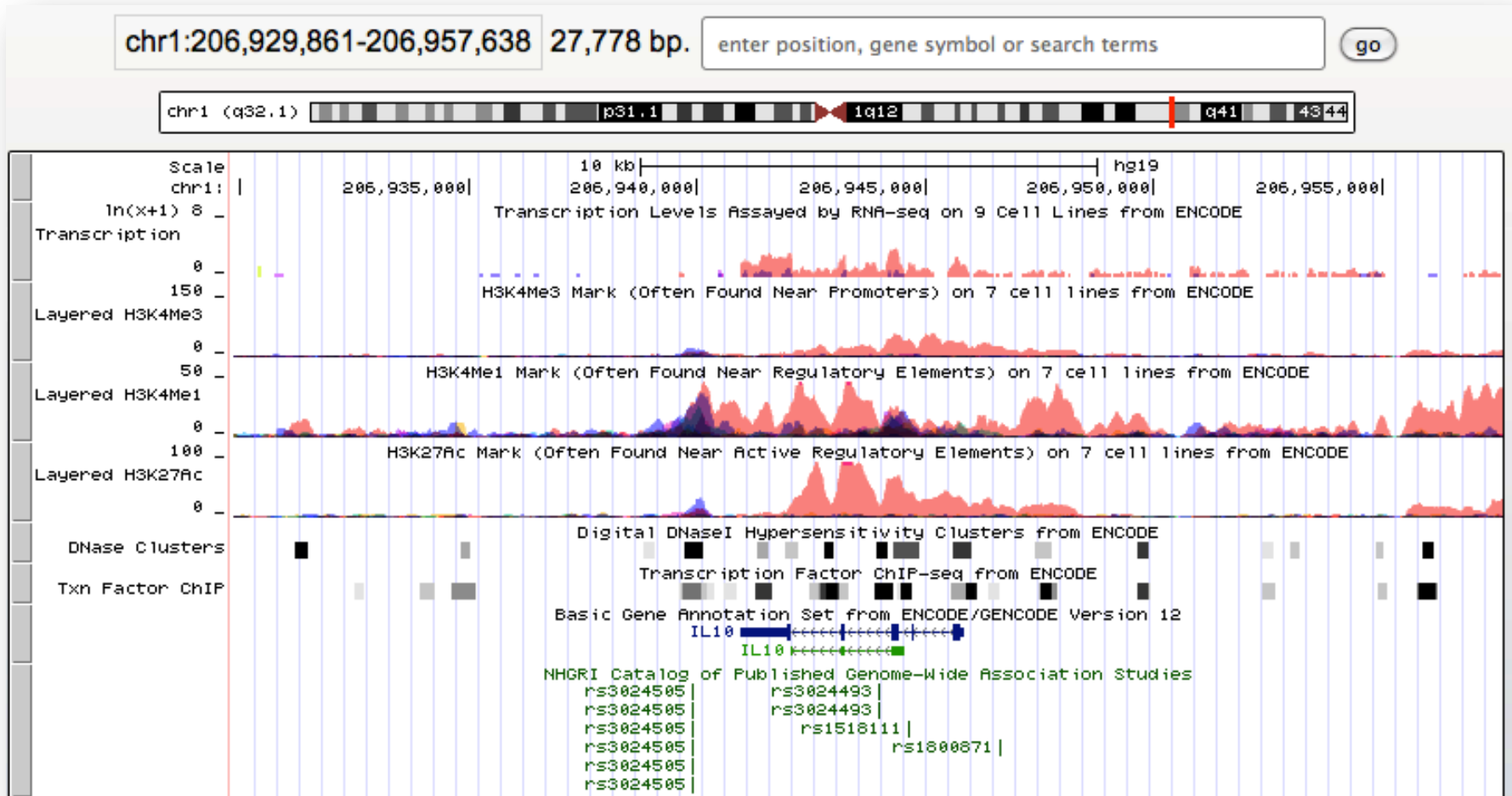
	DNA Methylation	Methyl Array	Methyl IRBS	Open Chromatin	DNase-DGF	DNase-seq	FAIRE-seq	RNA Binding Proteins	RIP Gene ST	RIP Tiling Array	RIP Validation	RIP-seq	RNA Profiling	CAGE	Exon Array	RNA-chip	RNA-PET	RNA-seq	Small RNA-seq	TFBS & Histones	ChIP-seq	Other	5C	ChIA-PET	Combined	DNA-PET	Genotype	
Tier 1																												
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																		3		4								
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>



ENCODE Browser

Viewing Locus of Interest





RegulomeDB Disease Database



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

RegulomeDB About Help

Linking Rheumatoid rs4810485

chr1:2,553,624 [rs3890745](#)
chr1:17,674,537 [rs2240335](#)
chr1:38,624,129 [rs12131057](#)
chr1:114,303,808 [rs6679677](#)
chr1:114,377,568 [rs2476601](#)
chr1:167,408,670 [rs840016](#)
chr2:61,136,129 [rs13031237](#)
chr2:61,164,331 [rs13017599](#)
chr2:65,595,586 [rs934734](#)
chr2:100,806,940 [rs1167692](#)
chr2:100,835,734 [rs1086503](#)
chr2:191,964,633 [rs7574865](#)
chr2:204,693,876 [rs231735](#)
chr2:204,738,919 [rs3087243](#)
chr3:56,966,246 [rs2062583](#)
chr3:58,556,841 [rs13315591](#)
chr4:25,417,244 [rs3816587](#)
chr4:26,108,197 [rs874040](#)
chr4:123,218,313 [rs1311972](#)
chr5:55,438,580 [rs6859219](#)
chr5:102,596,720 [rs26232](#)
chr6:29,789,171 [rs1610677](#)
chr6:31,379,931 [rs1063635](#)
chr6:32,218,989 [rs9296015](#)
chr6:32,282,854 [rs6910071](#)
chr6:32,429,643 [rs9268853](#)
chr6:32,574,171 [rs615672](#)
chr6:32,577,380 [rs660895](#)
chr6:32,602,269 [rs9272219](#)
chr6:32,662,851 [rs6457617](#)

Associations

Study: Genome-wide association study meta-analysis identifies seven new rheumatoid arthritis risk loci
First author: [unreadable]
Journal: Nat Genet
Date: 05/09/2012
Phenotype: Rheumatoid arthritis
Study population: [unreadable]
Replication population: [unreadable]
Association interval: [unreadable]
Odds ratio [95% CI]: [unreadable]
Gene: CD40
Risk allele: T
Minor allele frequency: [unreadable]
Study: Common
First author: [unreadable]
Journal: Nat Genet
Date: 09/14/2012
Phenotype: Rheumatoid arthritis
Study population: [unreadable]
Replication population: [unreadable]
Association interval: [unreadable]
Odds ratio [95% CI]: [unreadable]
Gene: CD40
Risk allele: G
Minor allele frequency: [unreadable]

Lead SNP

rs4810485
Position: chr20:44,747,946
Distance to nearest gene: [unreadable]
GENCODE version: [unreadable]
RegulomeDB score: 1f

Linkage Disequilibrium

Linkage disequilibrium (LD) plot showing correlations with other SNPs. No SNPs found for this LD threshold.

Data supporting chr20:44747946 (rs4810485)

Score: 1f
Likely to affect binding and linked to expression of a gene target

Human Feb. 2009 (GRCh37/hg19) chr20:44,747,746-44,748,146 (401 bp)
100 bases | hg19
44,747,800 | 44,747,850 | 44,747,900 | 44,747,950 | 44,748,000 | 44,748,050 | 44,748,100 |
RefSeq Genes
Publications: Sequences in scientific articles
H3K27Ac Mark (Often Found Near Active Regulatory Elements) on 7 cell lines from ENCODE
Digital DNaseI Hypersensitivity Clusters in 125 cell types from ENCODE
Transcription Factor ChIP-seq from ENCODE
Placental Mammal Basewise Conservation by PhyloP
Simple Nucleotide Polymorphisms (dbSNP 137) Found in >= 1% of Samples
Repeating Elements by RepeatMasker



RegulomeDB



Annotation of SNPs

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

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

rs4810485 ← 1

Submit ← 2

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

Protein Binding

Filter:

Method	Location	Bound Protein	? Cell Type	Additional Info	Reference
ChIP-seq	chr20:44747675..44747985	NFKB1	GM12878		ENCODE
ChIP-seq	chr20:44747677..44747987	NFKB1	GM12878	TNFa	ENCODE
ChIP-seq	chr20:44747695..44747951	MEF2A	GM12878		ENCODE
ChIP-seq	chr20:44747751..44747995	MEF2C	GM12878		ENCODE
ChIP-seq	chr20:44747763..44747979	SP1	GM12878		ENCODE

Chromatin structure

Filter:

Method	Location	? Cell Type	Additional Info	Reference
DNase-seq	chr20:44746436..44748294	Gm12892		ENCODE
DNase-seq	chr20:44746438..44748148	Urotsa		ENCODE
DNase-seq	chr20:44746482..44748340	Gm12878		ENCODE
DNase-seq	chr20:44746511..44748345	Gm19238		ENCODE
DNase-seq	chr20:44746548..44748220	Gm19239		ENCODE

Histone modifications

Filter:

Method	Location	Histone Mark	? Cell Type	Additional Info	Reference
ChIP-seq	chr20:44713831..44751698	H2az	H1hesc		ENCODE
ChIP-seq	chr20:44745395..44748640	H2az	Huvec		ENCODE
ChIP-seq	chr20:44745999..44747989	H2az	Monocd14ro1746		ENCODE
ChIP-seq	chr20:44745452..44749121	H3k04me1	Monocd14ro1746		ENCODE

Single nucleotides

Filter:

Method	Location	Affected Gene	? Cell Type	Additional Info	Reference
eQTL	chr20:44747946..44747947	NA	Lymphoblastoid	cis	20220756
eQTL	chr20:44747946..44747947	NA	Lymphoblastoid	cis	20220756
eQTL	chr20:44747946..44747947	CD40	Monocytes	cis	20502693

Summary of SNP analysis

Show 10 entries

Coordinate (0-based)	dbSNP ID	? Regulome DB Score	Other Resources
chr20:44747946	rs4810485	1f ← 3	C E SEMBL dbSNP

Showing 1 to 1 of 1 entries

<http://www.genome.gov/2755193>



HaploReg v2

Annotation of SNPs



www.broadinstitute.org/mammals/haploreg/

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

HaploReg v2



HaploReg is a tool for exploring annotations of the noncoding genome at variants on haplotype blocks, such as candidate regulatory SNPs at disease-associated loci. Using LD information from the 1000 Genomes Project, linked SNPs and small indels can be visualized along with their predicted chromatin state, their sequence conservation across mammals, and their effect on regulatory motifs. HaploReg is designed for researchers developing mechanistic hypotheses of the impact of non-coding variants on clinical phenotypes and normal variation.

Update 2013.02.14: Version 2 now includes an expanded library of SNPs (based on dbSNP 137), motif instances (based on PWMs discovered from ENCODE experiments), enhancer annotations (adding 90 cell types from the Roadmap Epigenome Mapping Consortium), and eQTLs (from the GTex eQTL browser). In addition, LD calculations are provided based on the 1000 Genomes Phase 1 individuals, and r^2 and D' measurements are available down to an r^2 threshold of 0.2. Display improvements include improved cell metadata, gene metadata, and PWM display on the detail pages and the option for text output. Version 1 is available [here](#).

Query SNP: **rs4810485** and variants with $r^2 \geq 0.8$

chr	pos (hg19)	LD (r ²)	LD (D')	variant	Ref	Alt	AFR freq	AMR freq	ASN freq	EUR freq	SiPhy cons	Promoter histone marks	Enhancer histone marks	DNase	Proteins bound	eQTL tissues	Motifs changed	GENCODE genes	dbSNP func annot
20	44730245	0.98	0.99	rs6032660	G	A	0.98	0.73	0.59	0.75							Mtf1,Zfx	12kb 5' of NCOA5	
20	44732089	0.97	0.99	rs2024568	T	C	0.97	0.73	0.58	0.75							BDP1,GCNF,Nr2f2	13kb 5' of NCOA5	
20	44734310	0.98	0.99	rs6032662	C	T	0.98	0.73	0.59	0.75							Zfp410	13kb 5' of CD40	
20	44735263	0.95	0.99	rs6032663	T	G	0.98	0.72	0.58	0.74							RFX5	12kb 5' of CD40	
20	44735854	0.97	0.99	rs6065926	A	G	0.99	0.76	1.00	0.75			GM12878	HMVEC-LLy			HMG-IY,PU.1	11kb 5' of CD40	
20	44739419	0.98	0.99	rs6032664	A	T	0.98	0.73	0.59	0.75			GM12878				Spdef	7.5kb 5' of CD40	
20	44740196	0.95	0.99	rs6074022	C	T	0.97	0.73	0.58	0.74		H3M3	GM12878	7 cell types	6 bound proteins		CHD2,Nrf-2	6.7kb 5' of CD40	
20	44742064	0.98	0.99	rs1569723	C	A	0.98	0.73	0.59	0.75			HMEC	ProgFib			Irf	4.8kb 5' of CD40	
20	44746982	1	1	rs1883832	T	C	0.98	0.73	0.59	0.75				LNCaP,Chorion,GM19239	13 bound proteins		4 altered motifs	CD40	5'-UTR
20	4474794	0.88	1	rs4810485	T	G	0.94	0.73	0.59	0.75		8 cell types	NHLF	10 cell types	4 bound proteins		STAT	CD40	intronic
20	44749251	0.88	1	rs4239702	T	C	0.85	0.70	0.60	0.72			GM12878	Huvec	6 cell types		Myf,Sox,Zfp105	CD40	intronic



Publications



General
Resources & FAQ *Now*
Publications
Software Tools
Data Standards
Human
Download
Experiment Matrix
Search
Genome Browser (hg19)
Integrative Analysis

About ENCODE Data

The [Encyclopedia of DNA Elements](#) (ENCODE) Consortium is an international collaboration of scientists from the National Human Genome Research Institute ([NHGRI](#)). The goal of ENCODE is to build a comprehensive parts list of functional elements in the human genome, including protein and RNA levels, and regulatory elements that control cells and circumstances.

Home - Downloads - Data Policy - Data Standards

ENCODE Publications

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



ENCODE Software Tools



General
Resources & FAQ *Now*
Publication
Software Tools
Data Standards
Human
Downloads
Experiment Matrix
Search
Genome Browser (hg19)
Integrative Analysis

About ENCODE Data

Home - Downloads - Data Policy - Data Standards

ENCODE Software Tools

Software Tools

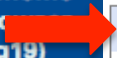
The goal of the ENCODE project is to generate a comprehensive catalog of all functional elements. To facilitate this task, members of the consortium have developed and refined software tools.

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

Mouse

Downloads

Experiment
Matrix

Search

Genome
Browser
(mm9)

Experi-
ment
List

Cell Types

Other

Registered
Variables

Antibodies

Education
and
Outreach

Release Lo

Data Policy

File Forma

Contributor

Pilot Project

Contacts

This [new page](#) provides links to ENCODE informational material and tools at the NHGRI, GEO, UCSC, pages at [encodeproject.org](#). It also includes a helpful FAQ section culled from ENCODE questions received.

25 January 2013 - Uniform DNaseI Hypersensitivity sites from ENCODE Analysis Working Group

ENCODE Consortium Data Release Policy Summary

Please observe the following guidelines when using ENCODE data:

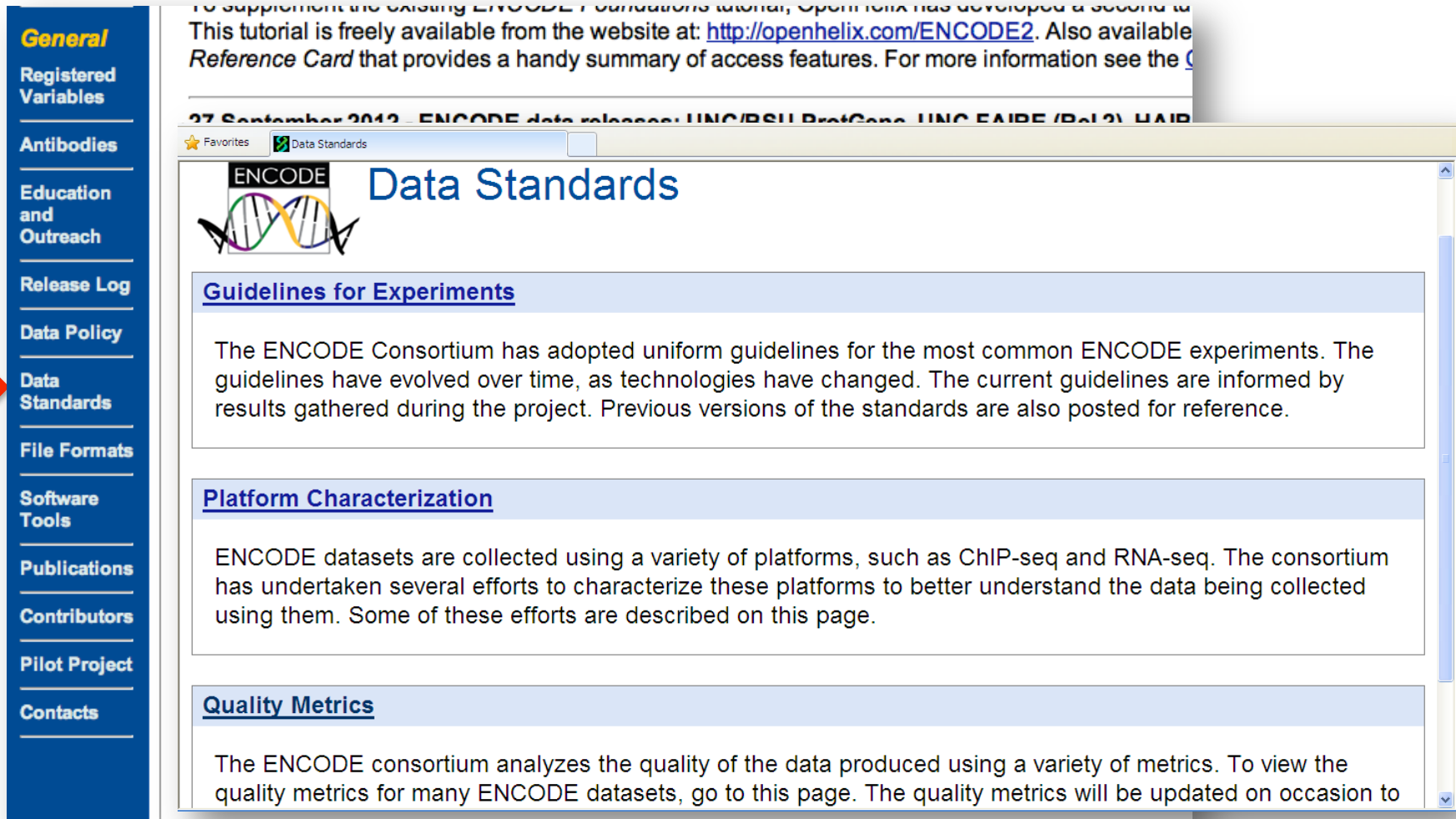
- Data users may freely download and analyze ENCODE data without restrictions. They may use ENCODE data in publications focused around individual genes; however, there is a narrow 9-month "moratorium" on the publication of global analysis of ENCODE data sets to allow the data producers the opportunity 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.

One new track and three track updates were released on the human hg19 browser. [Read more.](#)

25 September 2012 - Mouse ENCODE data releases: UW DNaseI DGF, UW DNaseI HS (Rel 2), LIC



ENCODE Data Standards



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 Foundations 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 ProtGeno, UNC EAIBE (Bel2), HAIR

ENCODE Data Standards

Guidelines for Experiments

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 posted for reference.

Platform Characterization

ENCODE datasets are collected using a variety of platforms, such as ChIP-seq and RNA-seq. The consortium has undertaken several efforts to characterize these platforms to better understand the data being collected using them. Some of these efforts are described on this page.

Quality Metrics

The ENCODE consortium analyzes the quality of the data produced using a variety of metrics. To view the quality metrics for many ENCODE datasets, go to this page. The quality metrics will be updated on occasion to



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 search bar and navigation tabs for HOME, PARTICIPANTS, DATA, PROTOCOLS, QUALITY METRICS, TOOLS, and PUBLICATIONS. Below the navigation, there are several panels: OVERVIEW, PROJECT DATA, MAPPING CENTERS, PROTOCOLS & STANDARDS, PUBLICATIONS, and NEWS. A large diagram illustrates various epigenetic data types including DNA methylation, DNase I hypersensitive sites, Histone Modifications, Chromatin, RNA, and Genes. A sidebar on the right contains 'VIEW/DOWNLOAD QUICK LINKS' with links to mirrors and data repositories, and a 'NEWS' section with recent updates.

The screenshot shows the UCSC Genome Browser Gateway interface. It includes a navigation bar with links like Home, Genomes, Blat, Tables, Gene Sorter, PCR, Session, FAQ, and Help. Below the navigation, there is a search form with fields for clade (Mammal), genome (Human), assembly (Feb. 2009 (GRCh37/hg19)), position or search term (chrX:70,752,933-70,795,740), and image width (800). There are buttons for 'submit', '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)

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



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 Production



Brad Bernstein (Manolis Kellis, Tony Kouzarides, Eric Lander, John Rinn)

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

Rick Myers (Barbara Wold, Ross Hardison, Flo Pauli Behn, Ali Mortazavi, Tim Reddy, Greg Cooper, Devin Absher)

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

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

Brent Graveley (Gene Yeo, Chris Burge, Xiang-Dong Fu)

Bing Ren (Joe Ecker, Len Pennacchio, Axel Visel, Wei Wang)

Mike Cherry (Jim Kent, Eurie Hong, David Haussler, **Kate Rosenbloom**)

Zhiping Weng (Mark Gerstein, Manolis Kellis, Roderic Guigo, Shirley Liu, Bill Noble, Rafael Irizarry)

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

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

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

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

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

Questions? pazinm@mail.nih.gov