



Disclosures

Presenters have no relevant
conflicts to disclose



Introduction to ENCODE



2:30 - 2:45

Introduction & overview of the Consortium

Eurie Hong, Ph.D.

Project Manager, ENCODE Data Coordination Center
Stanford University

2:45 - 3:00

Using the ENCODE Portal (hands-on)

J. Seth Strattan, Ph.D.

Data Wrangler, ENCODE Data Coordination Center
Stanford University

3:00 - 3:15

Genome annotations

Mark Gerstein, Ph.D.

Albert L. Williams Professor of Biomedical Informatics
Yale University

3:15 - 3:30

How to interpret and read a track (hands-on)

Richard Sandstrom

Research Scientist, Stamatoyannopoulos lab
University of Washington

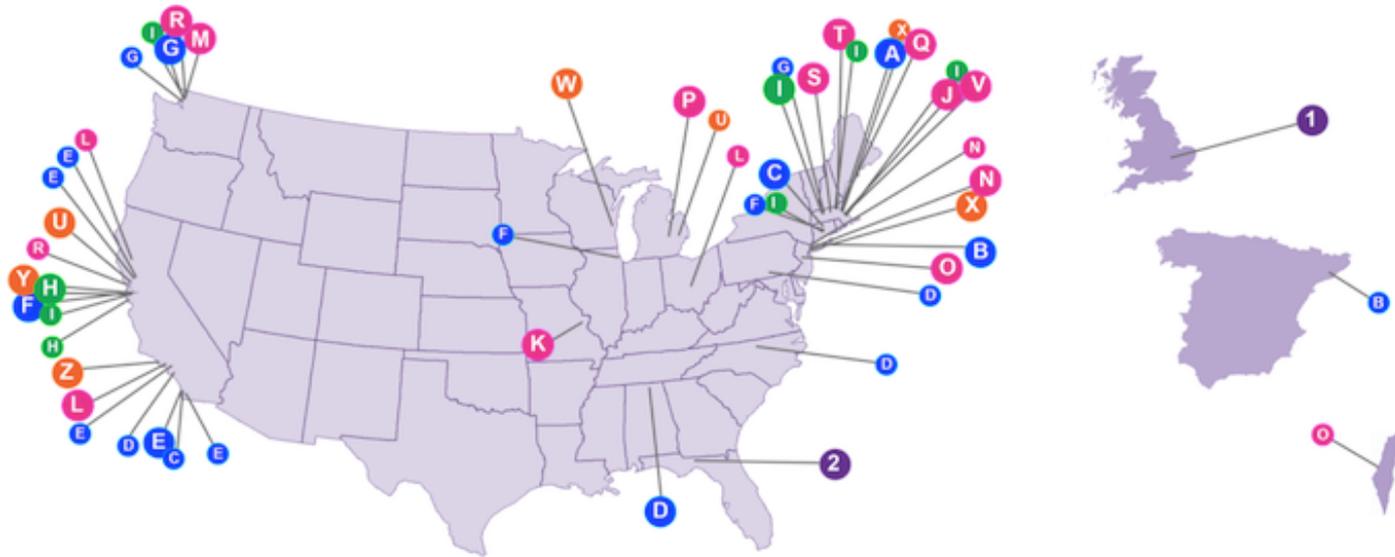
3:30 - 4pm

Variant annotation using ENCODE data (hands-on)

Luke Ward, Ph.D.

MIT

Who is the ENCODE Consortium?



Production Groups

- A** Broad Institute
- B** Cold Spring Harbor; Centre for Genomic Regulation (CRG);
- C** University of Connecticut Health Center; UCSD
- D** HudsonAlpha; Pennsylvania State; UC Irvine; Duke; Caltech
- E** UCSD; Salk Institute ; Joint Genome Institute; Lawrence Berkeley National Laboratory; UCSD
- F** Stanford; University of Chicago; Yale
- G** University of Washington; Fred Hutchinson Cancer Research Center; University of Massachusetts Medical School

Data Coordination Center

- H** Stanford; UCSC

Data Analysis Center

- I** University of Massachusetts Medical School; Yale; MIT; Stanford; Harvard; University of Washington

Technology Development Groups

- J** MIT
- K** Washington University, St. Louis
- L** USC; Ohio State University; UC, Davis
- M** University of Washington
- N** Sloan-Kettering; Weill Cornell Medical College
- O** Princeton; Weizmann
- P** University of Michigan
- Q** Broad Institute
- R** University of Washington; UCSF
- S** Advanced RNA Technologies, LLC
- T** Harvard

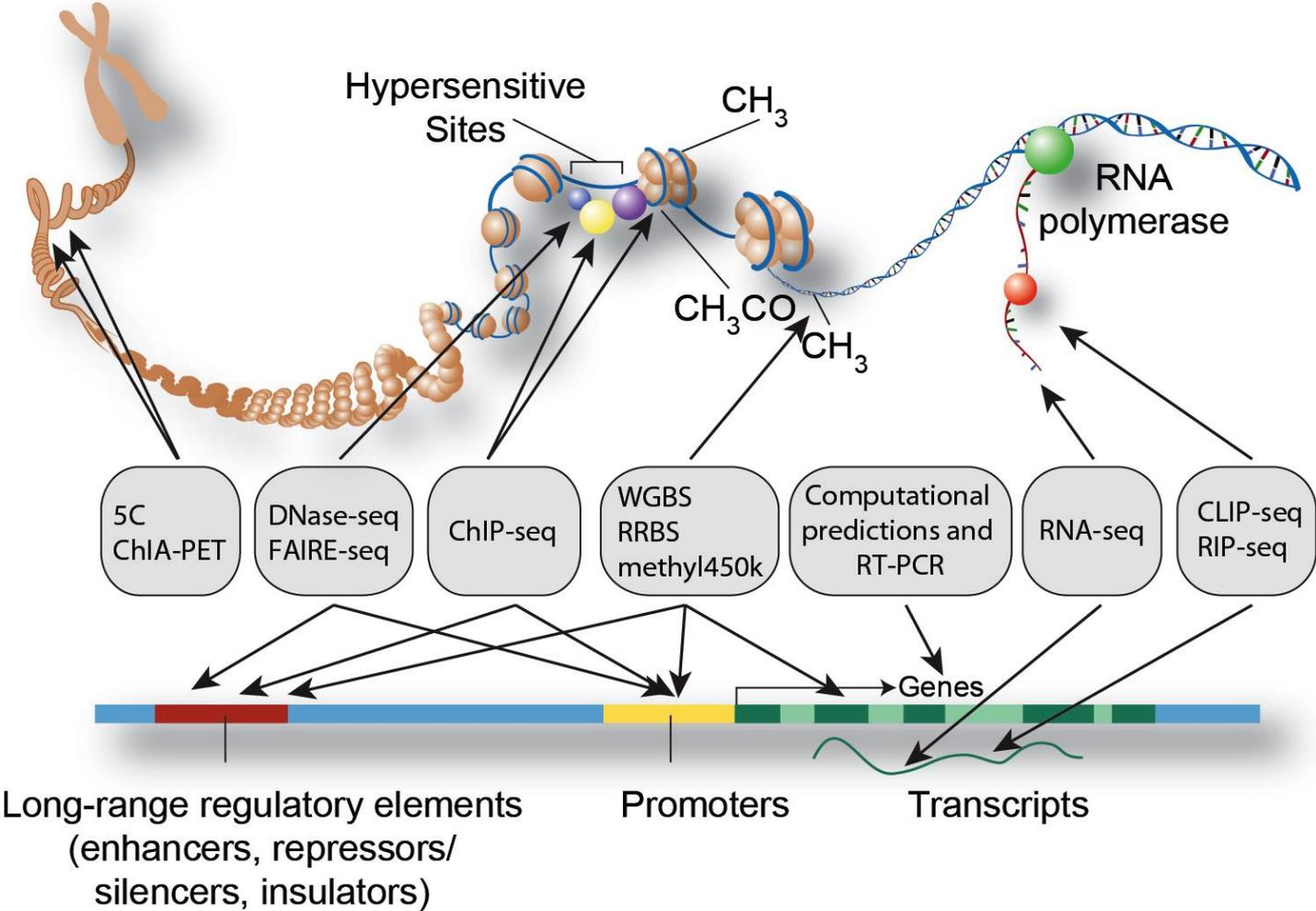
Computational Analysis Groups

- U** Berkeley; Wayne State University
- V** MIT
- W** University of Wisconsin
- X** Sloan-Kettering; Broad Institute
- Y** Stanford
- Z** UCLA

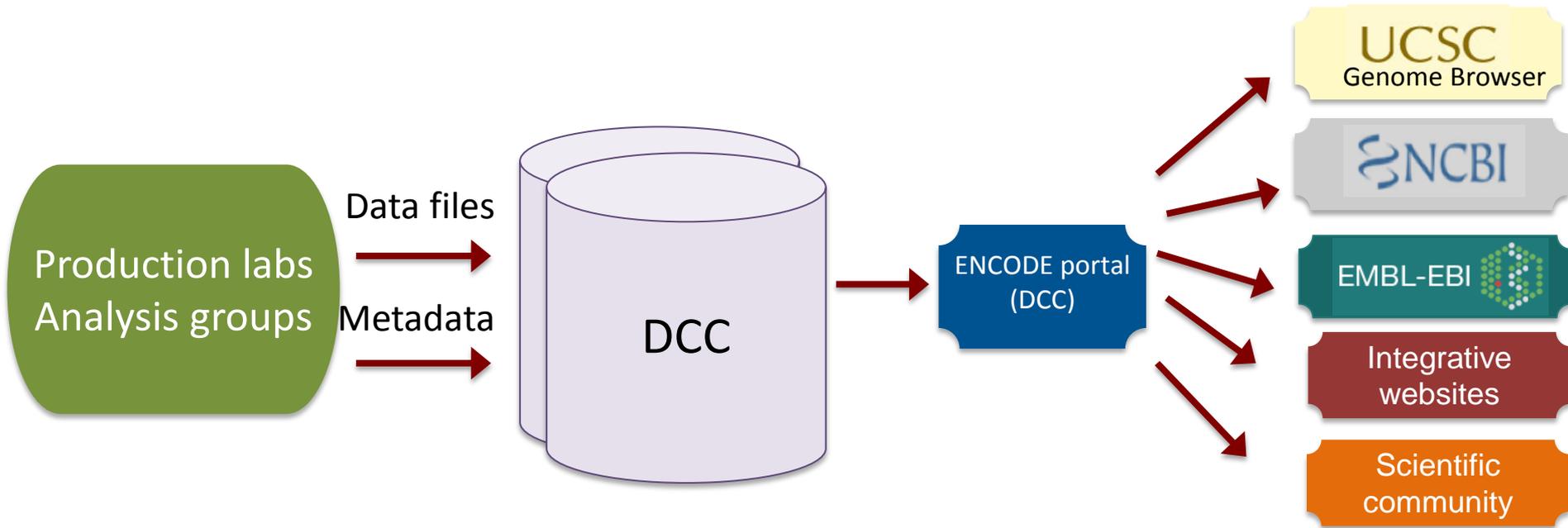
Affiliated Groups

- 1** Wellcome Trust Sanger Institute
- 2** Florida State University

What data does the consortium generate?



ENCODE data flow



Role:	Data generation	Data organization	Data access
Tasks:	Perform assays Perform analyses Validate data Submit data files Submit metadata	Define submission process Data processing & validation Data file storage Metadata curation	Web-based searches Data downloads

Principles of data organization

Transparency of methods

- How was the experiment performed?
- Are ENCODE data standards met?

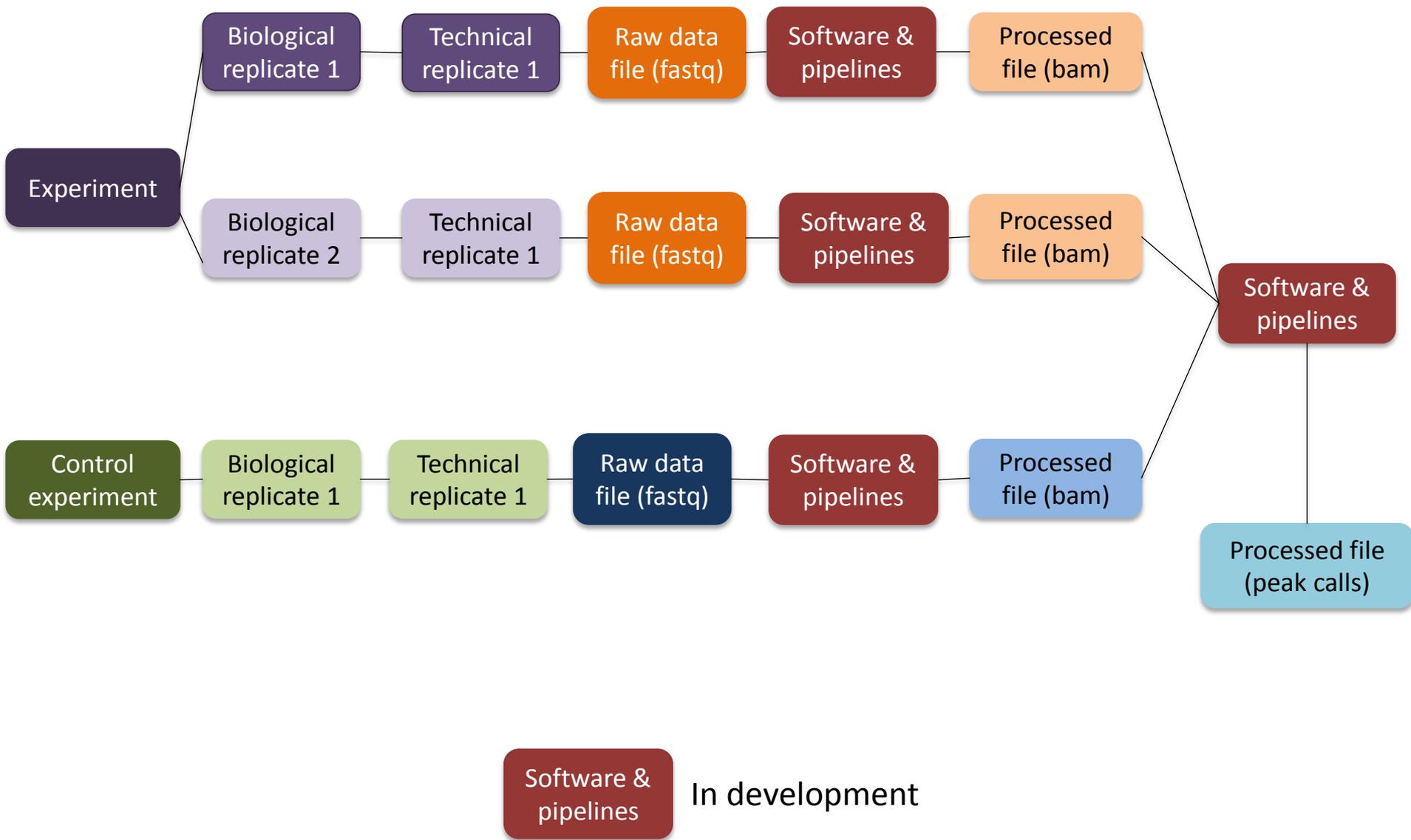
Reproducibility of results

- What files were used?
- What software & parameters were used to analyze the data?

Interoperability with other genomic projects

- Are standard file formats used?
- Can the metadata allow easy integration with other data?

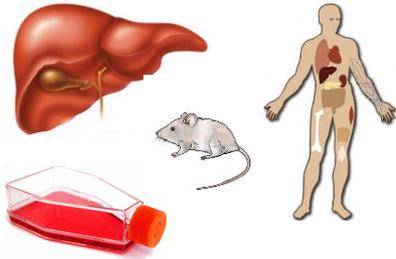
Metadata: capture the experimental design



Metadata: capture details of methods

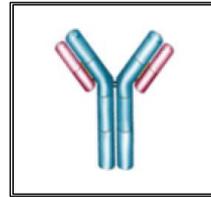
Experiment with replicates (*ENC SR000 DRY*)

Biosamples



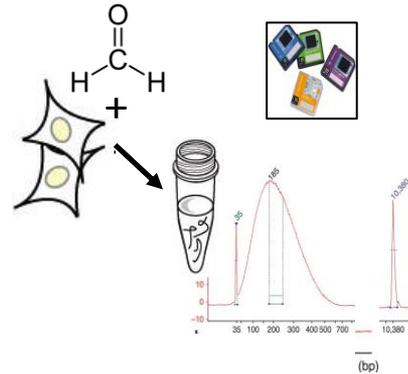
- Type (e.g. tissue, cell line)
- Ontology term name
- Source, product id, lot id
- Treatments
- Knockdown
- Fusion construct information
- Donor or strain information
- Dates (e.g. growth, harvest, procurement)
- Passage number
- Starting amount
- Lab assigned IDs

Antibodies



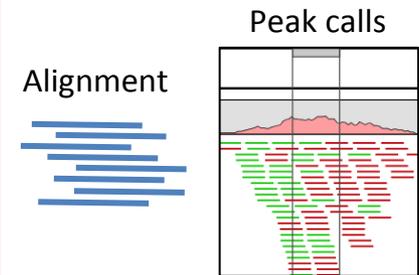
- Source, product id, lot id
- Isotype
- Antigen
- Host
- Purification method
- Validation status
- NHGRI approval status
- Target
- Species
- Dbxrefs

Libraries



- Library preparation protocol
- Strand specificity
- Size selection method
- Validation document
- Lysis method
- Sonication method
- Extraction method
- Nucleic acid type
- Nucleic acid size range

Files



- Reference genome version
- Alignment software
- Software parameters
- Software version
- Quality metrics (e.g. NRF, FRiP)

(selected subset of all metadata)

ENCBS095DKV (biosample)
ENCDO826IFN (donors)

ENCAB964IAU

ENCLB239KAN

ENCFF254TDA

Introduction to the ENCODE portal

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

ENCODE Data Methods About ENCODE Help Search ENCODE Sign in

ENCODE: Encyclopedia of DNA Elements

The diagram illustrates the ENCODE project's goal to identify functional elements in the human genome. It shows a DNA strand with various elements: Long-range regulatory elements (enhancers, repressors/silencers, insulators), Promoters, and Transcripts. Above the DNA, several assays are shown: SC ChIA-PET, DNase-seq, FAIRE-seq, ChIP-seq, WGBS, RRBS, methyl450k, Computational predictions and RT-PCR, RNA-seq, and CLIP-RIP-seq. These assays are linked to specific DNA features: Hypersensitive Sites, CH₃, CH₃CO, CH₃, and RNA polymerase.

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 parts list of functional elements in the human genome, including elements that act at the protein and RNA levels, and regulatory elements that control cells and circumstances in which a gene is active.

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

Data

To find and download ENCODE Consortium data:

- Click the Data toolbar above and browse data
 - [By assay](#)
 - [By biosample](#)
- Enter search terms like "skin", "ChIP-seq", or "CTCF"

ENCODE investigators employ a variety of assays and methods to identify functional elements. The discovery and annotation of gene elements is accomplished primarily by sequencing a diverse range of RNA sources, comparative genomics, integrative bioinformatic methods, and human curation. Regulatory elements are typically investigated through DNA hypersensitivity assays, assays of DNA methylation, and immunoprecipitation (IP) of proteins that interact with DNA and RNA, i.e., modified histones, transcription factors, chromatin regulators, and RNA-binding proteins, followed by sequencing.

All ENCODE data is freely available for download and analysis. Please refer to the [ENCODE Data Release Policy](#)

News

Sept 12, 2014: Data release: 23 human and 5 mouse datasets. [\[read more\]](#)

August 28, 2014: modENCODE and ENCODE [comparison papers](#) published. [\[read more\]](#)

August 19, 2014: New ENCODE portal released. The portal contains tools for browsing and searching data generated by the ENCODE consortium via assays, biological samples, and experimental reagents used. [\[read more\]](#)

July 17, 2014: Data Release: 760 experiments of ChIP-seq, RNA-seq, ChIA-Pet and 3 new assay types in human and mouse. [\[read more\]](#)

June 16, 2014: Visualize tracks on the UCSC Genome Browser via trackhubs-on-the-fly. [\[read more\]](#)

March 17, 2014: Antibody characterization standard updated to include antibodies against chromatin regulators and RNA binding-proteins. [\[read more\]](#)

See [news archive](#) for additional news and updates.