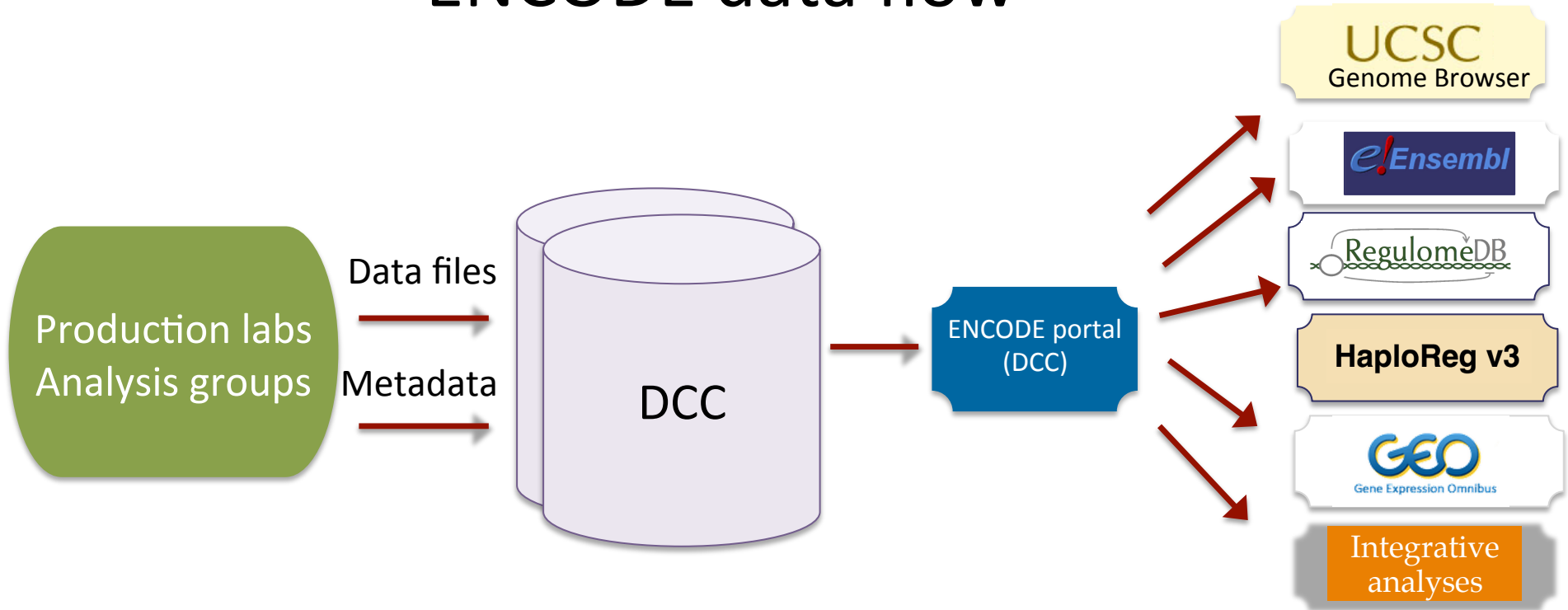


ENCODE data flow



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 Consortium standards REST API

ENCODE Portal

<https://www.encodeproject.org>

- Central source for ENCODE data: experimental and analysis data
- Hub for project information: data standards, publications
- High-quality metadata: data provenance & transparency

ENCODE Data - Methods - About ENCODE - Help - Search ENCODE Sign in

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 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)

Quick Start

To find and download ENCODE Consortium data:

- Click the Data toolbar above and browse data
 - By assay
 - By biosample
 - By genomic annotations
- Enter search terms like "skin", "ChIP-seq", or "CTCF"

Additional help about the ENCODE Portal:

- [Getting Started](#)
- [Batch Download](#)

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

UPDATED: The agenda has been posted for the [First ENCODE Users Meeting](#) will be held at the [Bojler Center](#) in Potomac, MD from June 29 - July 1, 2015

June 23, 2015: Data release: 3 human and 91 mouse datasets. [\[read more\]](#)

May 28, 2015: Changelog released for metadata schema updates. [\[read more\]](#)

May 18, 2015: Data release: 12 human datasets. [\[read more\]](#)

May 13, 2015: 519 publications that use ENCODE data, published by authors not funded by ENCODE, added to the ENCODE Portal [\[read more\]](#)

April 13, 2015: Data release: 28 human datasets. [\[read more\]](#)

March 31, 2015: Data release: 4 human datasets. [\[read more\]](#)

March 11, 2015: Experiment pages have been updated to show a graphical display of the pipeline used to generate the processed files associated with that experiment. [\[read more\]](#)

March 9, 2015: Batch Download of files released. [\[read more\]](#)

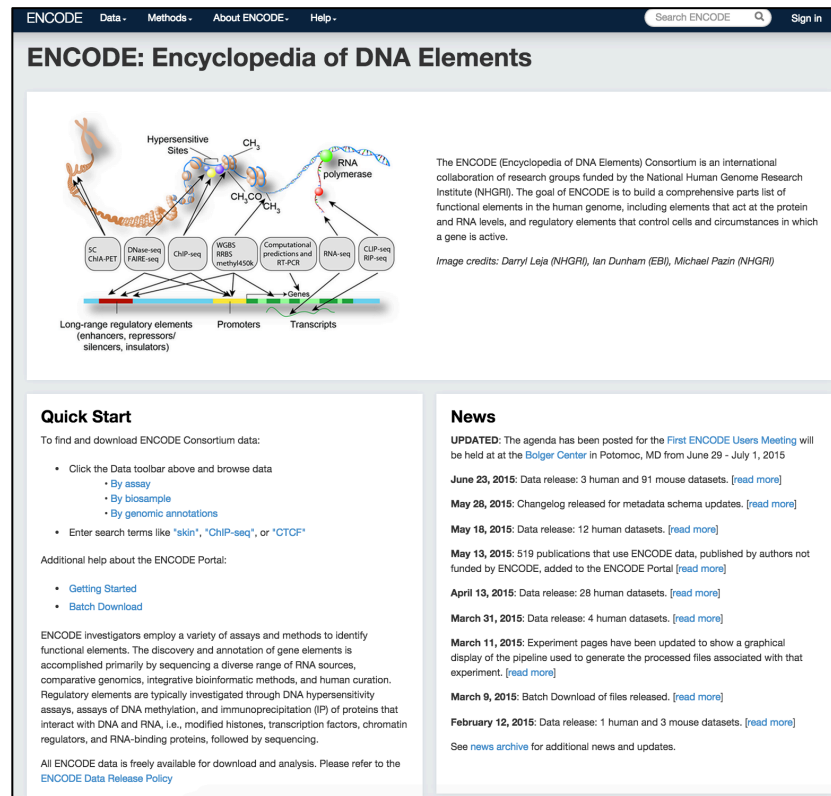
February 12, 2015: Data release: 1 human and 3 mouse datasets. [\[read more\]](#)

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

ENCODE Portal: Workshop goals

<https://www.encodeproject.org>

- View information about the project
- Find ENCODE data
- Visualize and download assays & data files
- Access ENCODE data and data files via the REST API



The screenshot displays the ENCODE Portal website. At the top, there is a navigation bar with links for 'Data', 'Methods', 'About ENCODE', and 'Help', along with a search bar and a 'Sign In' button. The main heading is 'ENCODE: Encyclopedia of DNA Elements'. Below this is a diagram illustrating the relationship between various assays and genomic elements. The diagram shows a DNA strand with 'Hypersensitive Sites' and 'CH₃' methyl groups. Assays shown include SC ChIP-seq, DNase-seq, FAIRE-seq, ChIP-seq, WGBS, RBS, VISTA, and RNA-seq. These assays are linked to 'Long-range regulatory elements (enhancers, repressors/silencers, insulators)', 'Promoters', and 'Transcripts'. A 'Gene' is also shown. To the right of the diagram is a text box describing the ENCODE Consortium and its goals, followed by image credits: 'Image credits: Darryl Leja (NHGRI), Ian Dunham (EBI), Michael Pazin (NHGRI)'. Below the diagram is a 'Quick Start' section with instructions on how to find and download data, and a 'News' section with a list of recent updates and data releases.

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 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)

Quick Start

To find and download ENCODE Consortium data:

- Click the Data toolbar above and browse data
 - By assay
 - By biosample
 - By genomic annotations
- Enter search terms like "skin", "ChIP-seq", or "CTCF"

Additional help about the ENCODE Portal:

- Getting Started
- Batch Download

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

UPDATED: The agenda has been posted for the [First ENCODE Users Meeting](#) will be held at the [Bolger Center](#) in Potomoc, MD from June 29 - July 1, 2015

June 23, 2015: Data release: 3 human and 91 mouse datasets. [\[read more\]](#)

May 28, 2015: Changelog released for metadata schema updates. [\[read more\]](#)

May 18, 2015: Data release: 12 human datasets. [\[read more\]](#)

May 13, 2015: 519 publications that use ENCODE data, published by authors not funded by ENCODE, added to the ENCODE Portal [\[read more\]](#)

April 13, 2015: Data release: 28 human datasets. [\[read more\]](#)

March 31, 2015: Data release: 4 human datasets. [\[read more\]](#)

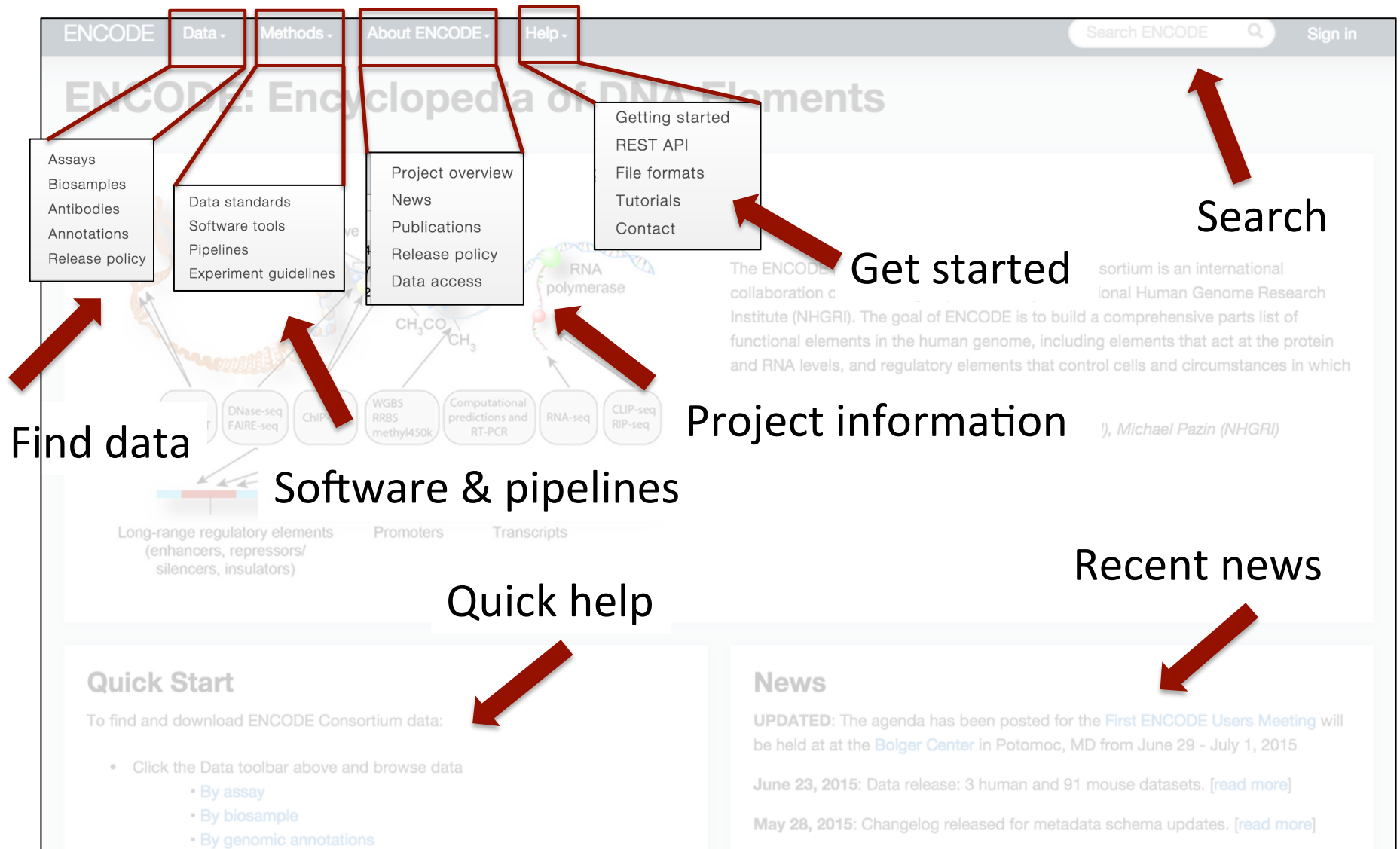
March 11, 2015: Experiment pages have been updated to show a graphical display of the pipeline used to generate the processed files associated with that experiment. [\[read more\]](#)

March 9, 2015: Batch Download of files released. [\[read more\]](#)

February 12, 2015: Data release: 1 human and 3 mouse datasets. [\[read more\]](#)

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

ENCODE Portal organization



ENCODE project documentation

The screenshot shows the ENCODE project website with several key sections highlighted by red boxes and arrows:

- Navigation Menu:** Includes 'Data -', 'Methods -', 'About ENCODE -', and 'Help -'. Red boxes highlight 'Methods -' and 'About ENCODE -'.
- Main Content Area:**
 - Methods:** A dropdown menu lists 'Data standards', 'Software tools', 'Pipelines', and 'Experiment guidelines'. A red arrow points to this menu.
 - About ENCODE:** A dropdown menu lists 'Project overview', 'News', 'Publications', 'Release policy', and 'Data access'. A red arrow points to this menu.
 - Help:** A dropdown menu lists 'Getting started', 'REST API', 'File formats', 'Tutorials', and 'Contact'. A red arrow points to this menu.
 - Publications:** A red arrow points to the 'Publications' link in the About ENCODE dropdown.
 - Experimental Standards:** A red arrow points to the 'Experimental Standards' link in the About ENCODE dropdown.
 - Tutorials:** A red arrow points to the 'Tutorials' link in the Help dropdown.
- Assay Icons:** A row of icons for various assays: 5C, ChIA-PET, DNase-seq, FAIRE-seq, ChIP-seq, WGBS, RRBS, methyl450k, Computational predictions and RT-PCR, RNA-seq, and CLIP-seq, RIP-seq. Red arrows point from these icons to the 'Experimental Standards' and 'Tutorials' sections.
- Long-range regulator:** A diagram showing a DNA segment with arrows indicating interactions, labeled 'Long-range regulator (enhancers, repressors, insulators)'.
- RNA polymerase:** A diagram showing a DNA segment with a red dot representing RNA polymerase, labeled 'RNA polymerase'.

Quick Start

To find and download ENCODE Consortium data:

- Click the Data toolbar above and browse data
 - By assay
 - By biosample
 - By genomic annotations

News

UPDATED: The agenda has been posted for the [First ENCODE Users Meeting](#) will be held at the [Bolger Center](#) in Potomoc, MD from June 29 - July 1, 2015

June 23, 2015: Data release: 3 human and 91 mouse datasets. [\[read more\]](#)

May 28, 2015: Changelog released for metadata schema updates. [\[read more\]](#)

Publications & related datasets

ENCODE Data - Methods - About ENCODE - Help - Search ENCODE Eurie Hong

Showing 9 of 9

A comparative encyclopedia of DNA elements in the mouse genome.
Mouse ENCODE Consortium, Yue F, Cheng Y, et al.
Nature. 2014 Nov 20;515(7527):355-64.
PMID:25409824

Comparative analysis of metazoan chromatin organization.
Ho JW, Jung YL, Liu T, et al.
Nature. 2014 Aug 28;512(7515):449-52.
PMID:25164756

Comparative analysis of the transcriptome across distant species.
Gerstein MB, Rozowsky J, Yan KK, et al.
Nature. 2014 Aug 28;512(7515):445-8.
PMID:25164755 PMID:PMG4155737

Comparative analysis of regulatory information and circuits across species.
Boyle AP, Araya CL, Brdlik C, et al.
Nature. 2014 Aug 28;512(7515):453-6.
PMID:25164757

Publication status
published 9

Category
technology development 131
pilot phase publication 79
software tool 27
genomic annotations 20
database 16
integrative analysis 12
key publication 9
model organism biology 1
tutorial 1

Available supplemental data
HMM regions 1
HOT regions 1
RNA annotations 1
enhancer annotations 1
peak calls 1

Publication year
2014 4
2012 2
2004 1
2007 1
2011 1

Journal
Nature 6
Genome biology 1
PLoS biology 1
Science 1
ENCODE 0

Published by
ENCODE 9
mouseENCODE 4
modENCODE 3

ENCODE Data - Methods - About ENCODE - Help - Search ENCODE Eurie Hong

Dataset ENCSR570AVM

Accession: ENCSR570AVM

Description: Data from the uniformly processed ChIP-seq experiments generated by the ENCODE Consortium between September 2007 through March 2012. The data represents peak calls (regions of enrichment) that were generated by the ENCODE Analysis Working Group (AWG) based on a uniform processing pipeline developed for the ENCODE Integrative Analysis effort and published in a set of coordinated papers in September 2012.

Dataset type: Composite

Lab: ENCODE Consortium Analysis Working Group

External resources: [UCSC-GB-hg19:wgEncodeAwgTfbsUniform](#)

References: [PMID:21526222](#) [PMID:22955616](#) [PMID:19029915](#)

Dataset documents

General protocol

Description excerpt: Methods describing the uniform processing pipeline for ChIP-seq experiments. The uniform...

[Uniform-processing-methods.pdf](#)

More

Files for dataset ENCSR570AVM

Accession	File type	Output type	Paired end	Associated replicates	Added by	Date added	File download
ENCF002CDN	bed_narrowPeak	UniformlyProcessedPeakCalls			Anshul Kundaje	2014-06-06	Download
ENCF002CDO	bed_narrowPeak	UniformlyProcessedPeakCalls			Anshul Kundaje	2014-06-06	Download
ENCF002CDP	bed_narrowPeak	UniformlyProcessedPeakCalls			Anshul Kundaje	2014-06-06	Download

Experiment standards: antibody characterizations

ENCODE Data - Methods - About ENCODE - Help - Search ENCODE Eurie Hong -

ENCAB000BFX

Antibody against *Homo sapiens* PCBP1

2

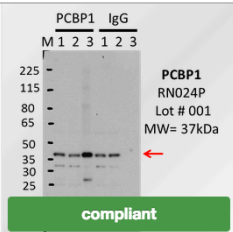
Homo sapiens K562 Eligible for new data

Source (vendor): [MBLI](#)
Product ID: [RN024P](#)
Lot ID: 001
Targets: [PCBP1 \(*Homo sapiens*\)](#)
Host: Rabbit
Clonality: Polyclonal
Purification: Affinity
Antigen description: KLH conjugated synthetic peptide, corresponding to internal region of human PCBP1

PCBP1 (*Homo sapiens*)

Method: immunoprecipitation

Caption excerpt: IP-Western blot analysis of K562 whole cell lysate using PCBP1 specific antibody. Lane 1 is 2.5% of 0.5mg input lysate, lane 2 is 2.5% of supernatant after immunoprecipitation and Lane 3 is 50% of...



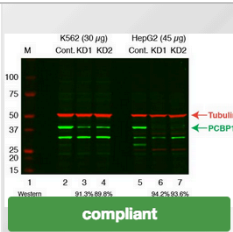
compliant

More

PCBP1 (*Homo sapiens*)

Method: knockdown or knockout

Caption excerpt: Western blot following shRNA against PCBP1 in K562 and HepG2 whole cell lysate using PCBP1 specific antibody. Lane 1 is a ladder, lane 2 is K562 non-targeting control knockdown, lane 2 and 3 are two...



compliant

More

Find ENCODE data

Browse

Search


The screenshot shows the ENCODE website interface. At the top, there is a dark blue navigation bar with the following items: ENCODE, Data (with a dropdown arrow), Methods (with a dropdown arrow), About ENCODE (with a dropdown arrow), Help (with a dropdown arrow), a search box labeled "Search ENCODE" with a magnifying glass icon, and a "Sign in" link. A red arrow points to the search box with the label "Search". Below the navigation bar, a dropdown menu is open under "Data", listing: Assays, Biosamples, Antibodies, Annotations, and Release policy. A red arrow points to this menu with the label "Browse". The main content area features the title "ENCyclopedia of DNA Elements" and a large diagram. The diagram illustrates the relationship between various genomic features and the data generated by different assays. At the top, it shows "Hypersensitive Sites" (orange), "CH₃" (blue), "CH₃CO" (green), and "RNA polymerase" (red). Below these are several assay boxes: "3C", "ChIA-PET", "DNase-seq", "FAIRE-seq", "ChIP-seq", "WGBS", "RRBS", "methyl450k", "Computational predictions and RT-PCR", "RNA-seq", and "CLIP-seq", "RIP-seq". Arrows point from these assays to a DNA strand at the bottom, which is divided into "Long-range regulatory elements (enhancers, repressors/silencers, insulators)", "Promoters", and "Transcripts".

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)

Search ENCODE data

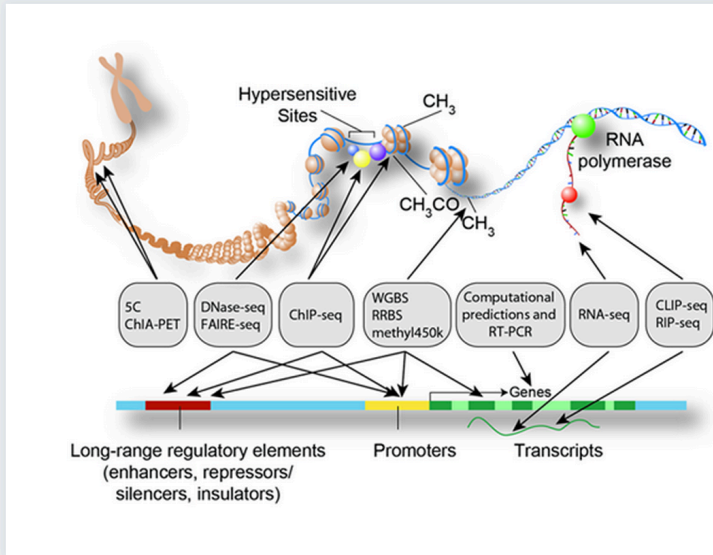
Enter "skin" into search box



ENCODE Data Methods About ENCODE Help

skin Sign in

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 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)

Search ENCODE data

The screenshot shows the ENCODE website search results for the term 'skin'. The page is titled 'ENCODE: Encyclopedia of DNA Elements'. The search bar at the top right contains the text 'skin'. A red arrow points to the search bar. Below the search bar, there is a navigation menu with 'Data', 'Methods', 'About ENCODE', and 'Help'. The main content area is divided into several sections:

- Data Type**: A table showing the number of experiments for each data type.

Data Type	Count
Experiments	170
Biosamples	108
Publications	48
Web page	2
- Assay**: A table showing the number of experiments for each assay type.

Assay	Count
ChIP-seq	53
RNA-seq	26
DNase-seq	25
RNA profiling by array assay	24
CAGE	9
- Experiment status**: A table showing the number of experiments for each status.

Experiment status	Count
released	169
revoked	1
- Genome assembly (visualization)**: A table showing the number of experiments for each genome assembly.

Genome assembly (visualization)	Count
hg19	147
mm9	2
- Organism**: A table showing the number of experiments for each organism.

Organism	Count
<i>Homo sapiens</i>	167
<i>Mus musculus</i>	2
- Target of assay**: A table showing the number of experiments for each target of assay.

Target of assay	Count
histone	31
histone modification	29
transcription factor	13
control	9
- Biosample type**: A table showing the number of experiments for each biosample type.

Biosample type	Count
primary cell	163
tissue	5
stem cell	2

The main content area also displays a list of 25 experiments, with the first few being:

- RAMPAGE of skin of body (*Homo sapiens*, fetal)**: Experiment ENCSR000AGU, released. Lab: Thomas Gingeras, CSHL. Project: ENCODE.
- RNA-seq of skin of body (*Homo sapiens*, fetal)**: Experiment ENCSR000AGA, released. Lab: Thomas Gingeras, CSHL. Project: ENCODE.
- RNA-seq of melanocyte of skin (*Homo sapiens*, adult)**: Experiment ENCSR000CUR, released. Lab: Thomas Gingeras, CSHL. Project: ENCODE.
- RNA-seq of skin of body (*Homo sapiens*, fetal)**: Experiment ENCSR000AFG, released. Lab: Thomas Gingeras, CSHL. Project: ENCODE.
- DNA methylation profiling by array assay of zone of skin (*Homo sapiens*, adult 83 year)**: Experiment ENCSR000BVZ, released. Lab: Richard Myers, HAIB. Project: ENCODE.
- CAGE of melanocyte of skin (*Homo sapiens*, child)**: Experiment ENCSR000CKY, released. Lab: Piero Carninci, RIKEN. Project: ENCODE.

Browse ENCODE data

The screenshot shows the ENCODE project website interface. At the top, there is a navigation bar with 'ENCODE', 'Data', 'Methods', 'About ENCODE', and 'Help'. A search bar and 'Sign in' link are also present. A dropdown menu is open under 'Data', listing 'Assays', 'Biosamples', 'Antibodies', 'Annotations', and 'Release policy'. A red arrow points from this menu to the 'Assays' section of the main content area.

The main content area features a diagram titled 'ENCODE Encyclopedia of DNA Elements'. The diagram illustrates various genomic features: 'Hypersensitive Sites', 'CH₃' and 'CH₃CO' modifications, 'SC ChIA-PET', 'DNase-seq FAIRE-seq', 'ChIP-seq', 'WGBS RRBS methyl450k', and 'Comp prediction'. Below the diagram, it identifies 'Long-range regulatory elements (enhancers, repressors/silencers, insulators)' and 'Promoters'.

Below the diagram, there are several summary tables:

Assay	
ChIP-seq	2466
RNA-seq	714
DNase-seq	270
shRNA knockdown followed by RNA-seq	245
RNA profiling by array assay	180
+ See more...	

Experiment status	
released	4670
revoked	14

Genome assembly (visualization)	
hg19	2799
mm9	558
mm10	164
dm3	108

Organism	
<i>Homo sapiens</i>	3531
<i>Mus musculus</i>	980
<i>Drosophila melanogaster</i>	108

Target of assay	
transcription factor	1199
histone	871
histone modification	844
control	451
RNA binding protein	312

Below these tables, the page displays 'Showing 25 of 4684 experiments'. There are buttons for 'Filter to 500 to visualize', 'Download', and 'View All'. The list of experiments includes:

- RNA-seq of cardiac ventricle fibroblast (*Homo sapiens*)** - Experiment ENCSR369RVN released. Lab: Thomas Gingeras, CSHL. Project: ENCODE.
- RNA-seq of hair follicular keratinocyte (*Homo sapiens*, adult 55 year)** - Experiment ENCSR680USE released. Lab: Thomas Gingeras, CSHL. Project: ENCODE.
- RNA-seq of airway epithelial cell (*Homo sapiens*)** - Experiment ENCSR822SUG released. Lab: Thomas Gingeras, CSHL. Project: ENCODE.
- RNA-seq of epithelial cell of proximal tubule (*Homo sapiens*)** - Experiment ENCSR118TVR released. Lab: Thomas Gingeras, CSHL. Project: ENCODE.
- RNA-seq of pericardium fibroblast (*Homo sapiens*)** - Experiment ENCSR362HMX released. Lab: Thomas Gingeras, CSHL. Project: ENCODE.
- RNA-seq of mesangial cell (*Homo sapiens*)** - Experiment ENCSR198TKA released. Lab: Thomas Gingeras, CSHL. Project: ENCODE.
- RNA-seq of bronchial fibroblast (*Homo sapiens*)** - Experiment ENCSR620NSN released. Lab: Thomas Gingeras, CSHL. Project: ENCODE.

Filter all ENCODE data using metadata

- Categories on the left are metadata describing assays and files
- Selecting an option in the category narrows results
- Choosing more than one option in a category acts as an “OR”
- Choosing more than one category acts as an “AND”
- Select “skin” under the “Organ” category



ENCODE Data Methods About ENCODE Help Search ENCODE Sign in

Showing 25 of 4778 experiments Filter to 500 to visualize Download View All

Assay

ChIP-seq	2530
RNA-seq	729
DNase-seq	270
shRNA knockdown followed by RNA-seq	245
RNA profiling by array assay	180

+ See more...

Experiment status

released	4764
revoked	14

Genome assembly (visualization)

hg19	2828
mm9	558
mm10	164
dm3	108

Organism

Homo sapiens	3531
Mus musculus	1071
Drosophila melanogaster	108

Target of assay

transcription factor	1215
histone	927
histone modification	900
control	461
RNA binding protein	314

+ See more...

Biosample type

immortalized cell line	2633
tissue	890
primary cell	789
stem cell	208
in vitro differentiated cells	134

+ See more...

Organ

brain	274
skin of body	170
blood vessel	116
lung	102
liver	101

+ See more...

RNA Bind-n-Seq Experiment ENCSR118LLV released
Target: No protein target control
Lab: Chris Burge, MIT
Project: ENCODE

RNA Bind-n-Seq Experiment ENCSR883DFK released
Target: Input library control
Lab: Chris Burge, MIT
Project: ENCODE

ChIP-seq of neural tube (Mus musculus, embryonic 13.5 day) Experiment ENCSR087PLZ released
Target: H3K9ac
Lab: Bing Ren, UCSD
Project: ENCODE

ChIP-seq of heart (Mus musculus, embryonic 13.5 day) Experiment ENCSR069XHI released
Target: H3K9ac
Lab: Bing Ren, UCSD
Project: ENCODE

ChIP-seq of neural tube (Mus musculus, embryonic 13.5 day) Experiment ENCSR107SLP released
Target: H3K36me3
Lab: Bing Ren, UCSD
Project: ENCODE

ChIP-seq of heart (Mus musculus, embryonic 13.5 day) Experiment ENCSR239TSZ released
Target: H3K36me3
Lab: Bing Ren, UCSD
Project: ENCODE

ChIP-seq of neural tube (Mus musculus, embryonic 13.5 day) Experiment ENCSR580MEU released
Target: H3K27me3
Lab: Bing Ren, UCSD
Project: ENCODE

ChIP-seq of heart (Mus musculus, embryonic 13.5 day) Experiment ENCSR380BBL released
Target: H3K27me3
Lab: Bing Ren, UCSD
Project: ENCODE

ChIP-seq of neural tube (Mus musculus, embryonic 13.5 day) Experiment ENCSR973AYQ released
Target: H3K4me2
Lab: Bing Ren, UCSD
Project: ENCODE

Combine search & filtering

The screenshot illustrates the ENCODE project website's search and filtering interface. The search bar at the top contains the term 'skin'. The left sidebar provides various filters, with 'Data Type' and 'Life stage' being highlighted. The main content area shows a list of experiments, with the first view displaying 25 of 170 experiments and the second view displaying 10 of 10 experiments. Red arrows point to the search bar, the 'Data Type' filter, and the search bar in the second view.

Data Type

Experiments	170
Biosamples	108
Publications	48
Web page	

Assay

ChIP-seq	53
RNA-seq	26
DNase-seq	25

Showing 25 of 170 experiments

Assay

ChIP-seq	20
RNA-seq	10
DNase-seq	9
RNA profiling by array assay	6
DNA methylation profiling by array assay	5

Experiment status

released	10
----------	----

Genome assembly (visualization)

hg19	6
------	---

Organism

<i>Homo sapiens</i>	10
---------------------	----

Biosample type

primary cell	10
--------------	----

Organ

skin of body	10
lymphatic vessel	2

Life stage

unknown	11
adult	10
child	4
fetal	2

Showing 10 of 10 experiments

- RNA-seq of melanocyte of skin (*Homo sapiens*, adult)**
Lab: Thomas Gingeras, CSHL
Project: ENCODE
- RNA-seq of melanocyte of skin (*Homo sapiens*, adult)**
Lab: Thomas Gingeras, CSHL
Project: ENCODE
- RNA-seq of hair follicle dermal papilla cell (*Homo sapiens*, adult)**
Lab: Thomas Gingeras, CSHL
Project: ENCODE
- RNA-seq of fibroblast of arm (*Homo sapiens*, adult 53 year)**
Lab: Thomas Gingeras, CSHL
Project: ENCODE
- RNA-seq of fibroblast of arm (*Homo sapiens*, adult 53 year)**
Lab: Thomas Gingeras, CSHL
Project: ENCODE
- RNA-seq of fibroblast of dermis (*Homo sapiens*, adult)**
Lab: Thomas Gingeras, CSHL
Project: ENCODE

<https://www.encodeproject.org>

Visualize & download data

The image shows a screenshot of the ENCODE project website. At the top, there is a navigation bar with 'ENCODE', 'Data', 'Methods', 'About ENCODE', and 'Help'. A search bar contains the word 'skin' and a 'Sign in' button is on the right. Below the navigation bar, there is a section for 'Showing 10 of 10 experiments'. On the left, there is an 'Assay' filter menu with 'RNA-seq' selected. The main content area displays a list of experiments, with the first one being 'RNA-seq of melanocyte of skin (Homo sapiens, adult)'. A red arrow points from this experiment to a detailed view of the same experiment. In the detailed view, there is a 'Visualize' button and a 'Download' button. A red arrow points from the 'Download' button to a 'Using batch download' dialog box. The dialog box contains instructions on how to use the 'Download' button and provides a cURL command: `xargs -n 1 curl -o -L < files.txt`. Below the dialog box, there is a table of experiments with columns for 'Experiment' and 'Status'. The table shows several experiments, including 'RNA-seq of melanocyte of skin (Homo sapiens, adult)'.

ENCODE Data Methods About ENCODE Help skin Sign in

Assay
ChIP-seq 20
RNA-seq 10
DNase-seq 9
RNA profiling by array assay 6
DNA methylation profiling by array assay5
+ See more...

Showing 10 of 10 experiments Visualize Download

RNA-seq of melanocyte of skin (*Homo sapiens, adult*)
Lab: Thomas Gingeras, CSHL
Project: ENCODE

Experiment ENCSR000CUR released

RNA-seq of melanocyte of skin (*Homo sapiens, adult*)
Lab: Thomas Gingeras, CSHL

Using batch download

Click the "Download" button below to download a "files.txt" file that contains a list of URLs to a file containing all the experimental metadata and links to download the file. The first line of the file will always be the URL to download the metadata file. Further description of the contents of the metadata file are described in the [Batch Download help doc](#).

The "files.txt" file can be copied to any server.
The following command using cURL can be used to download all the files in the list:

```
xargs -n 1 curl -o -L < files.txt
```

Close Download

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
chr21:33031597-33041570 9,974 bp. enter position, gene symbol or search terms go [More on-site workshops available!](#)

chr21 (q22.11) 31p13 31p12 31p11.3 31q12 31q11.2 31q11.1 31q11.3 31q11.4 31q11.5 31q11.6 31q11.7 31q11.8 31q11.9 31q12.1 31q12.2 31q12.3 31q12.4 31q12.5 31q12.6 31q12.7 31q12.8 31q12.9

Scale chr21: 33,633,000 33,634,000 33,635,000 33,636,000 33,637,000 33,638,000 33,639,000 33,640,000 33,641,000 hg19

ENCF088HK RNA-seq of fibroblast of dermis - ENCSR000CUH
ENCF088HT
ENCF088HR
ENCF088HS
ENCF088HT
ENCF088HE
ENCF088HH
ENCF088HV
ENCF088HW
ENCF088HX
ENCF088HY
ENCF088HZ
ENCF088IA
ENCF088IB
ENCF088IC
ENCF088ID
ENCF088IE
ENCF088IF
ENCF088IG
ENCF088IH
ENCF088II
ENCF088IJ
ENCF088IK
ENCF088IL
ENCF088IM
ENCF088IN
ENCF088IO
ENCF088IP
ENCF088IQ
ENCF088IR
ENCF088IS
ENCF088IT
ENCF088IU

RNA-seq of fibroblast of dermis - ENCSR000CUV
RNA-seq of hair follicle dermal papilla cell - ENCSR000CUB
RNA-seq of melanocyte of skin - ENCSR000CUC
RNA-seq of melanocyte of skin - ENCSR000CVH
RNA-seq of melanocyte of skin - ENCSR000CUI

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 >

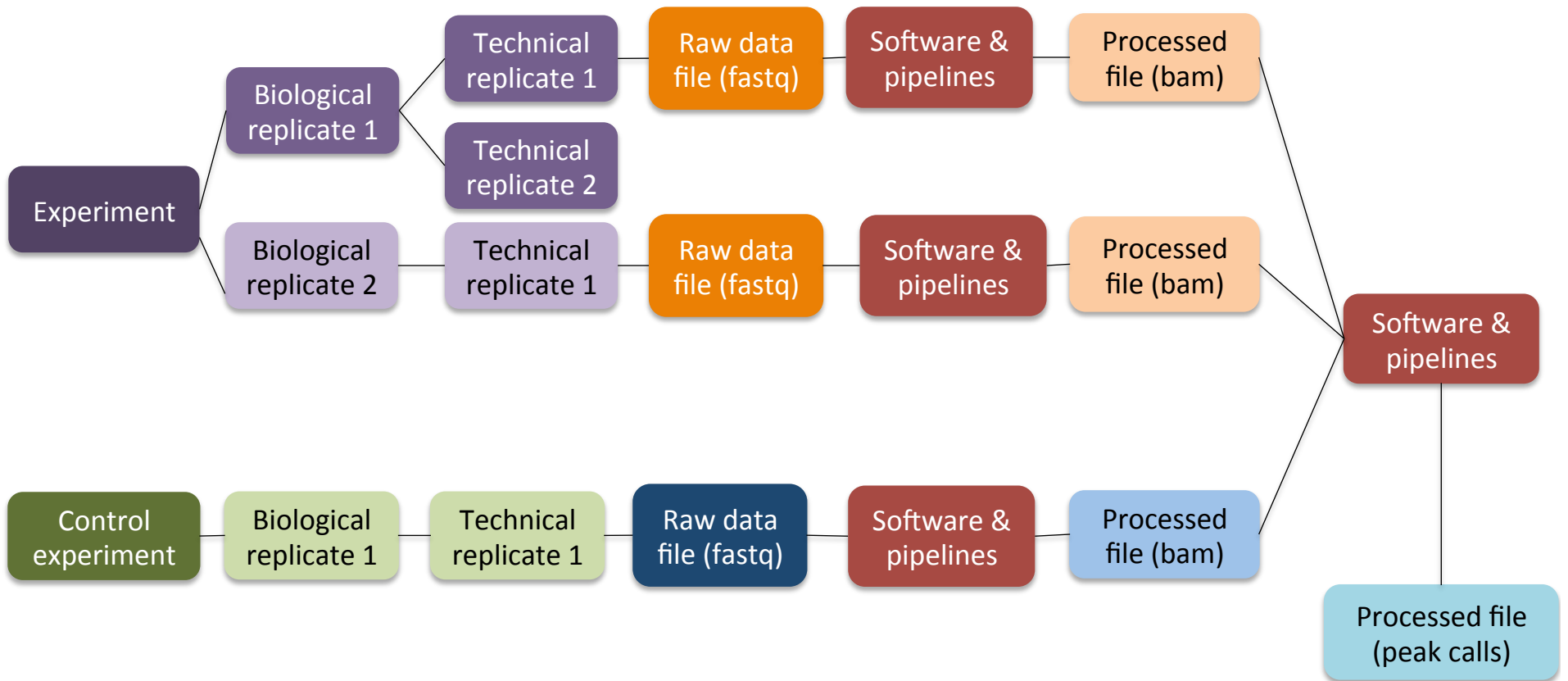
track search default tracks default order hide all add custom tracks track hubs configure reverse resize refresh

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

collapse all Hub (search) refresh

RNA-seq of fibroblast of dermis - ENCSR000CUH	RNA-seq of fibroblast of dermis - ENCSR000CUI	RNA-seq of hair follicle dermal papilla cell - ENCSR000CUB	RNA-seq of hair follicle dermal papilla cell - ENCSR000CVC	RNA-seq of melanocyte of skin - ENCSR000CUC	RNA-seq of melanocyte of skin - ENCSR000CVM
full ↑	full ↑	full ↑	hide ↑	full ↑	full ↑

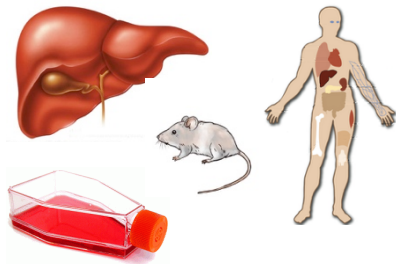
Metadata describing ENCODE assays



Identify reusable experimental variables

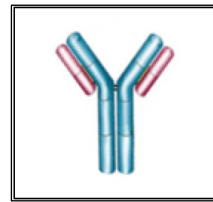
Experiment with replicates

Biosamples



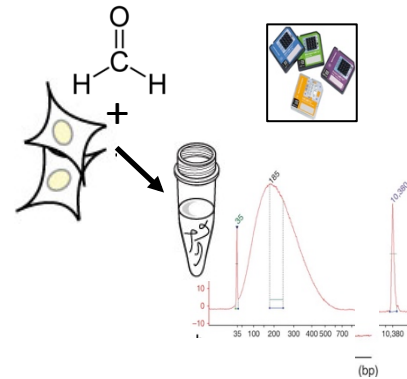
- Type (e.g. tissue, cell line)
- Ontology term name
- Source, product id, lot id
- Treatments
- Knockdown
- Fusion construct information
- Donor or strain information
- Dates (e.g. growth, harvest, procurement)
- Passage number
- Starting amount
- Lab assigned IDs

Antibodies



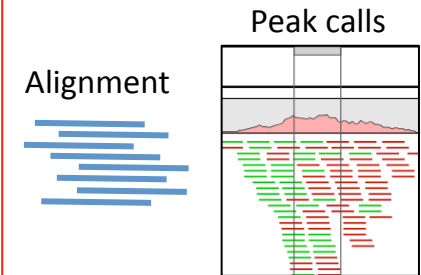
- Source, product id, lot id
- Isotype
- Antigen
- Host
- Purification method
- Validation status
- NHGRI approval status
- Target
- Species
- Dbxrefs

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



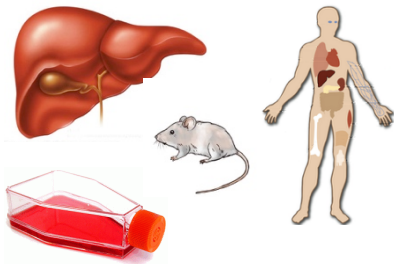
- Reference genome version
- Alignment software
- Software parameters
- Software version
- Quality metrics (e.g. NRF, FRiP)

(selected subset of all metadata)

Accession them

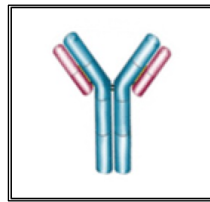
Experiment with replicates (*ENC SR000DRY*)

Biosamples



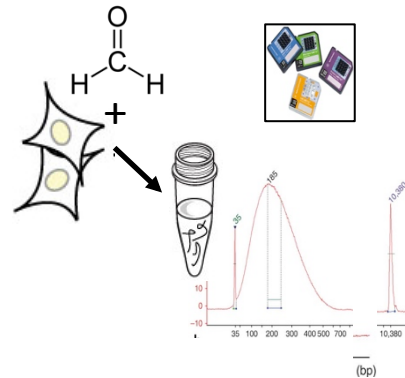
- Type (e.g. tissue, cell line)
- Ontology term name
- Source, product id, lot id
- Treatments
- Knockdown
- Fusion construct information
- Donor or strain information
- Dates (e.g. growth, harvest, procurement)
- Passage number
- Starting amount
- Lab assigned IDs

Antibodies



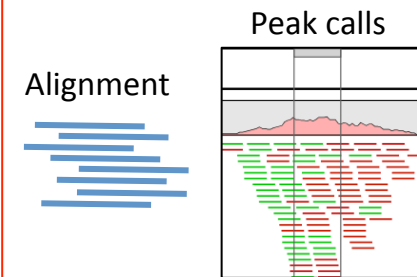
- Source, product id, lot id
- Isotype
- Antigen
- Host
- Purification method
- Validation status
- NHGRI approval status
- Target
- Species
- DBxrefs

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)

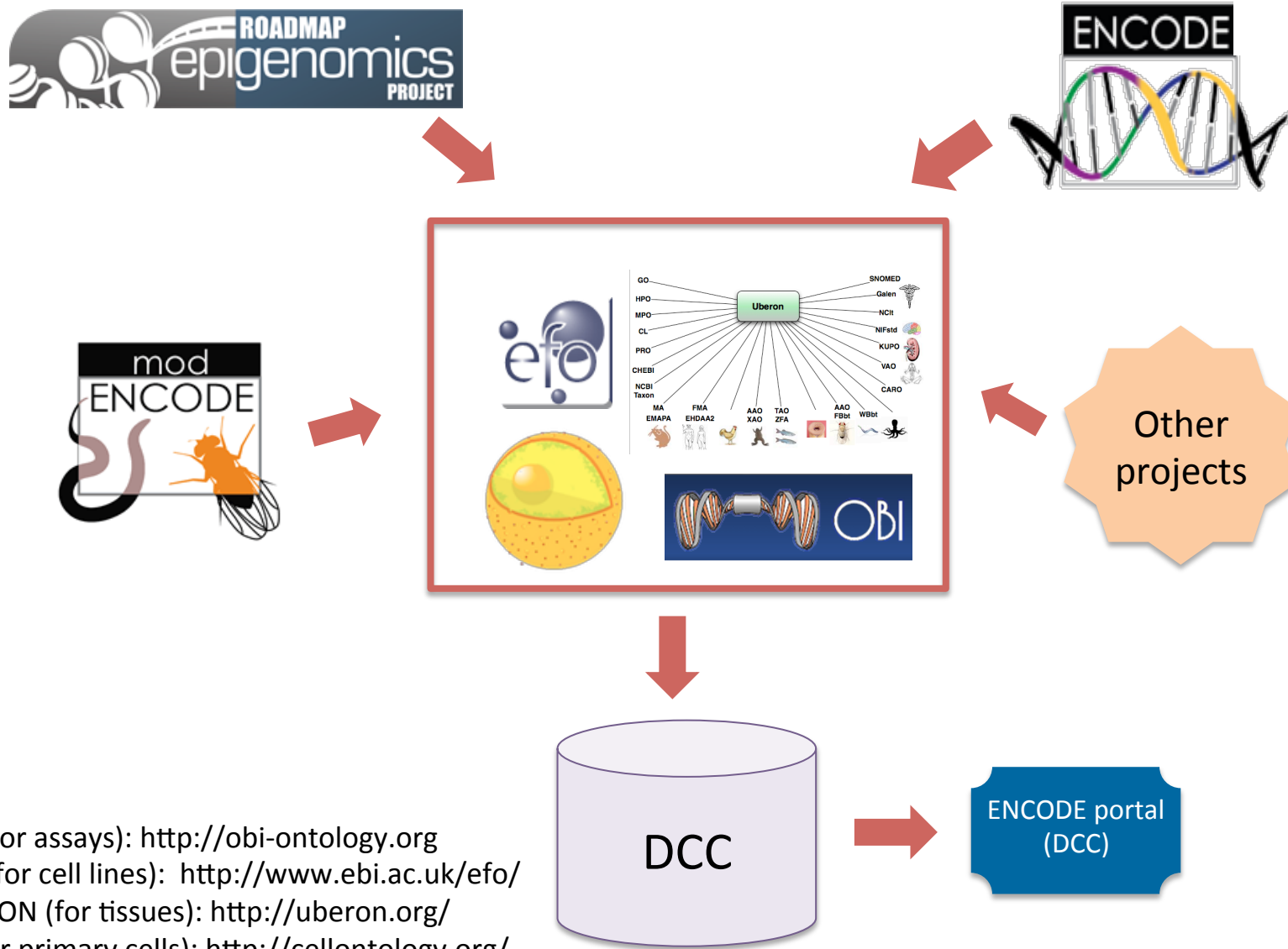
ENCBS095DKV (biosample)
ENCDO826IFN (donors)

ENCAB964IAU

ENCLB239KAN

ENCFF254TDA

Metadata integration using ontologies



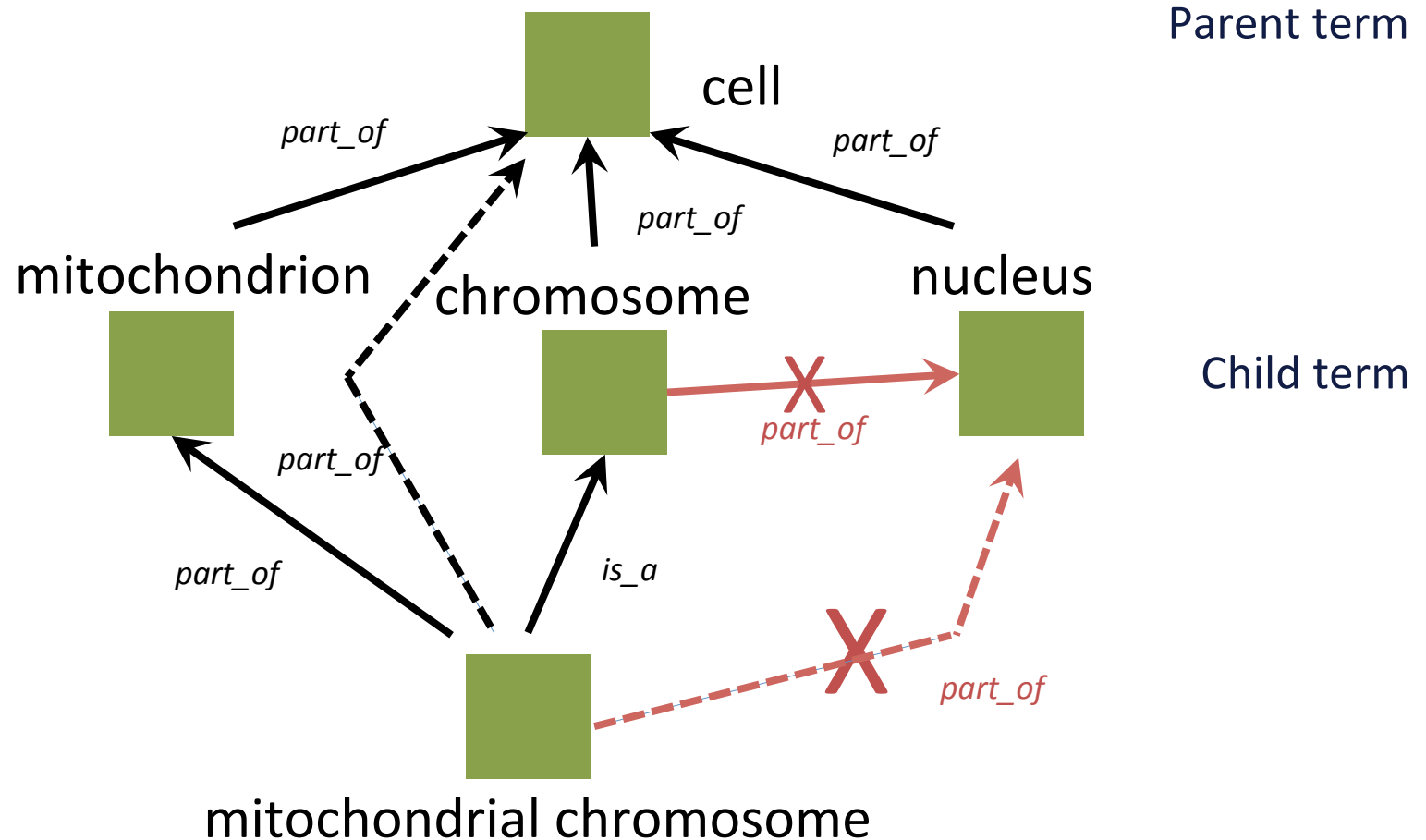
OBI (for assays): <http://obi-ontology.org>
EFO (for cell lines): <http://www.ebi.ac.uk/efo/>
UBERON (for tissues): <http://uberon.org/>
CL (for primary cells): <http://cellontology.org/>
ChEBI (for treatments): <https://www.ebi.ac.uk/chebi/>

An ontology is a set of words...

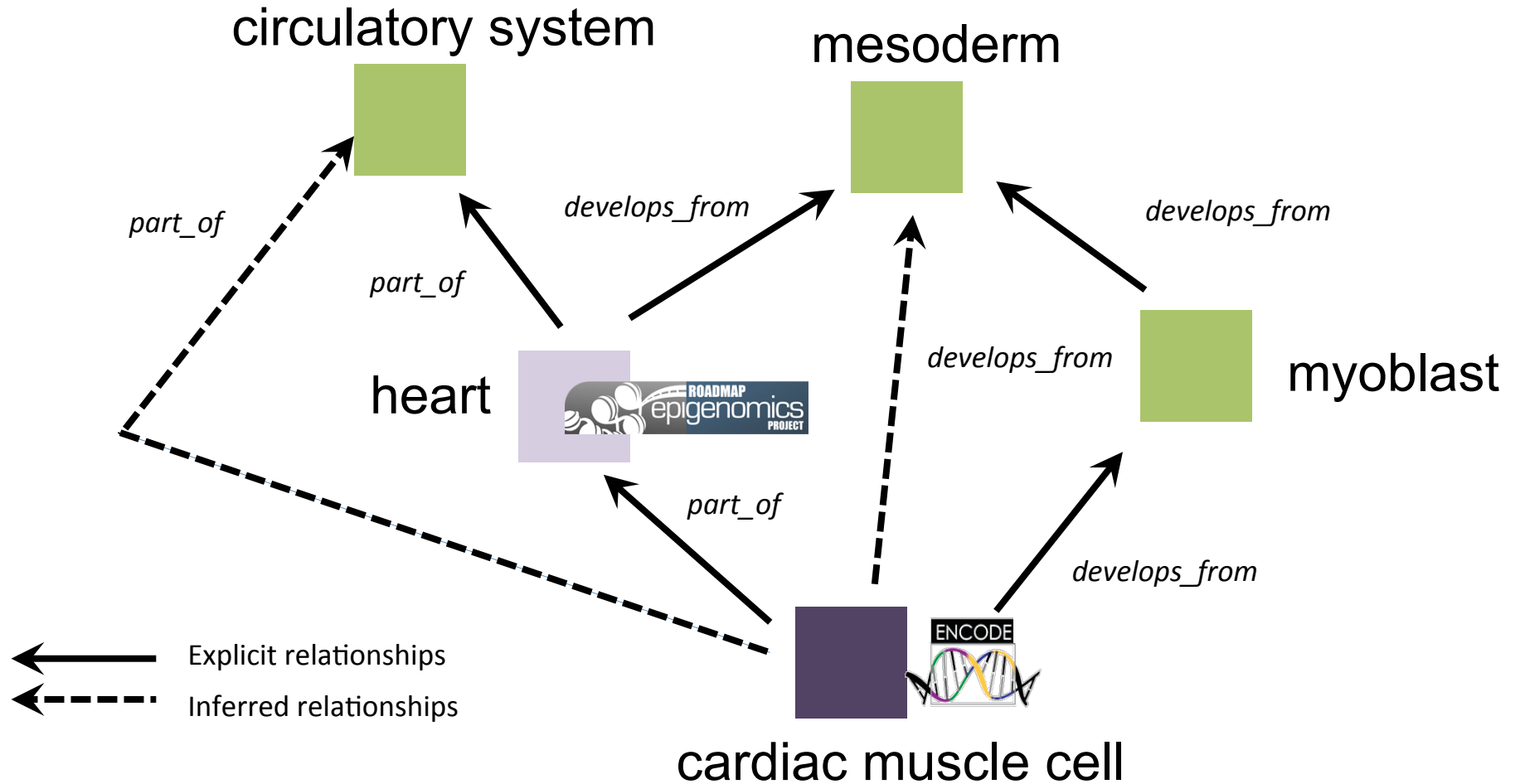
.. with different types of relationships to each other.

All relationships must be true

because inferences can be made based on these relationships

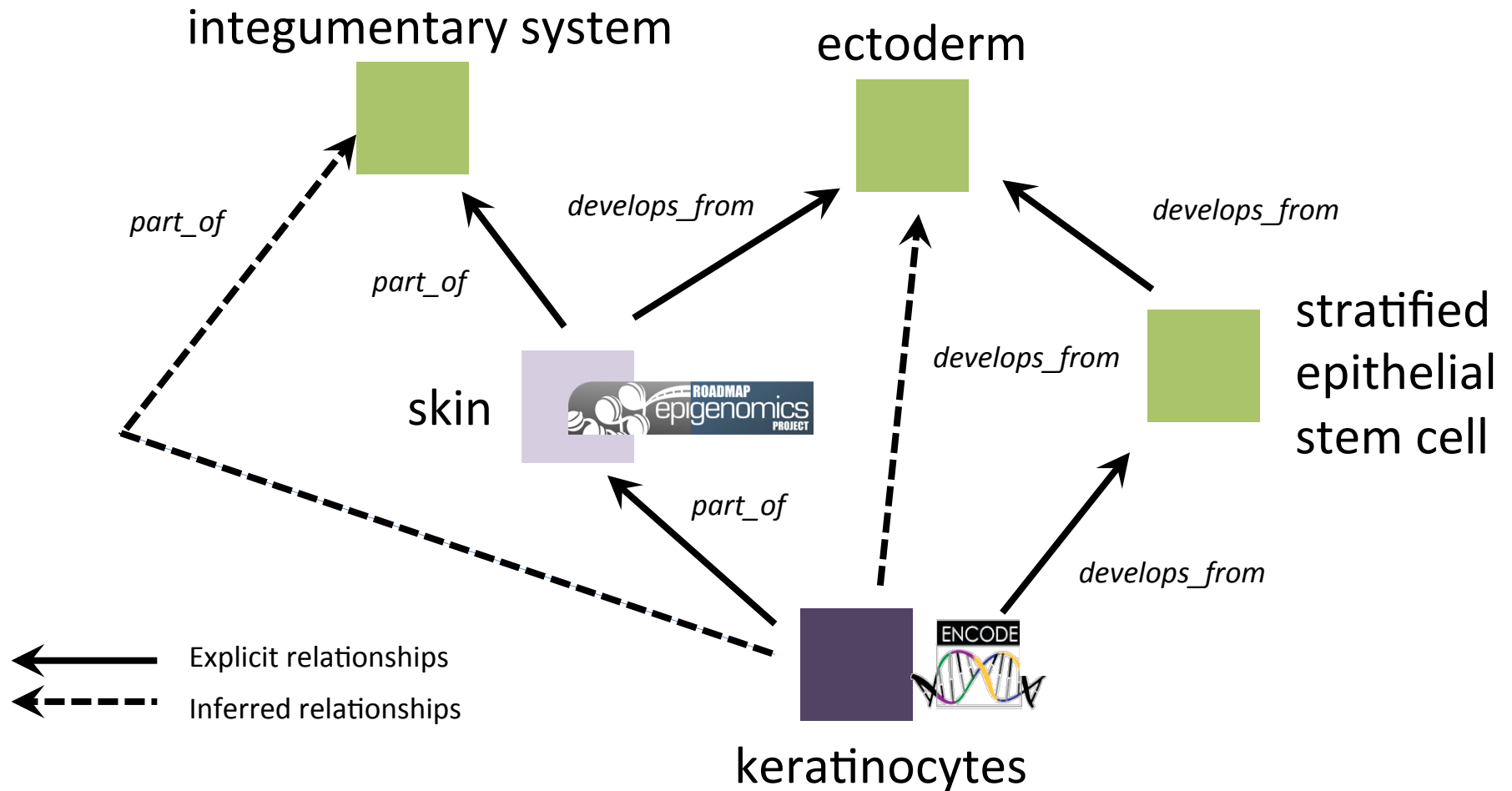


Impact of using ontologies: Common ontologies = instant interoperability



<http://uberon.org/>
<http://cellontology.org/>

Impact of using ontologies: Common ontologies = instant interoperability



<http://uberon.org/>
<http://cellontology.org/>

View experimental metadata




Experiment summary for ENCSR823VEE
Status: released Validation: pending

Assay: RNA-seq
Accession: ENCSR823VEE
Biosample summary: embryonic facial prominence (*Mus musculus*, embryonic 14.5 day)
Type: tissue
Description: RNA-seq on embryonic 14.5 day mouse embryonic facial prominence
Lab: Barbara Wold, Caltech
Project: ENCODE
Date released: 2014-12-17

Assay details

Nucleic acid type: RNA
Lysis method: Ambion mirVana
Extraction method: Ambion mirVana
Fragmentation method: chemical (Nextera tagmentation)
Size range: >200
Size selection method: SPRI beads
Platform: HiSeq 2500
Spike-ins datasets: ENCSR156CIL

Documents

Extraction protocol	Extraction protocol	General protocol
Description: Lysis Protocol	Description: Genomic DNA Removal Protocol	Description: A copy of the Smart-seq protocol
		
cms_055423.pdf	cms_055740.pdf	SmartSeqProtocol.pdf
More	More	More

Biological replicate - 1

Technical replicate: 1
Library: ENCLB254AMJ
Library starting quantity: 10.0 ng
Biosample: ENCBS776GWF - embryonic facial prominence

Biological replicate - 2

Technical replicate: 1
Library: ENCLB015WCJ
Library starting quantity: 10.0 ng
Biosample: ENCBS903HZW - embryonic facial prominence

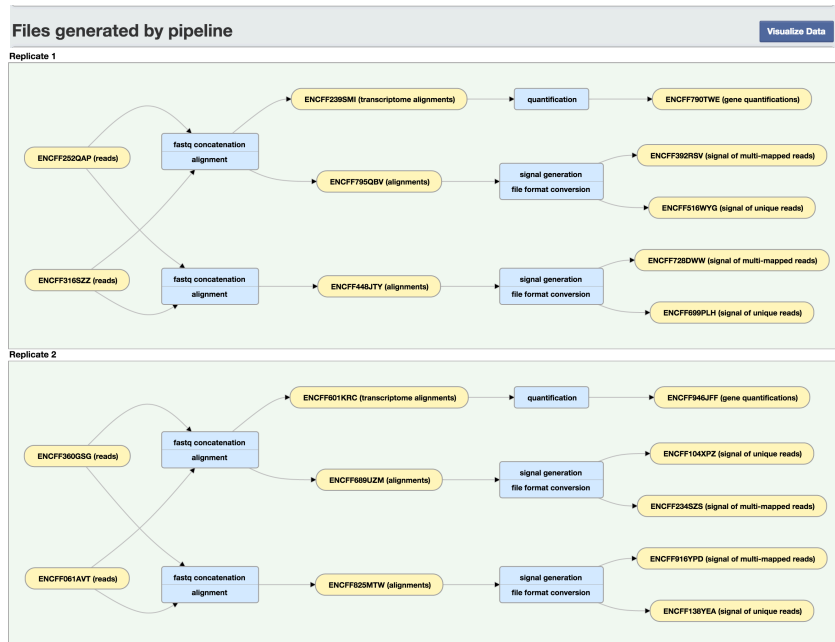
The set up of the assay, including controls (if needed)

Details of how the assay was done

Protocols

Links to more sample information

Data provenance & reproducibility



- File related based on analysis steps
- Interactive graph: click on file nodes and software steps
- Running pipelines will be covered in more detail tomorrow

- Visualize data for a single assay
- Download individual files

Files linked to ENCSR823VEE

Raw data										
Accession	File type	Biological replicate	Technical replicate	Read length	Run type	Paired end	Mapping assembly	Lab	Date added	Validation status
ENCF061AVT Download 2.28 GB	fastq	2	1	100 nt	single-ended			Barbara Wold, Caltech	2014-12-12	pending
ENCF252QAP Download 1.81 GB	fastq	1	1	100 nt	single-ended			Barbara Wold, Caltech	2014-12-12	pending
ENCF316SZZ Download 1.82 GB	fastq	1	1	100 nt	single-ended			Barbara Wold, Caltech	2014-12-12	pending
ENCF360GSG Download 2.27 GB	fastq	2	1	100 nt	single-ended			Barbara Wold, Caltech	2014-12-12	pending

Processed data										
Accession	File type	Output type	Biological replicate	Technical replicate	Mapping assembly	Genome annotation	Lab	Date added	Validation status	
ENCF392RSV Download 162 MB	bigWig	signal of multi-mapped reads	1	1	mm10	M4	ENCODE Processing Pipeline	2015-01-06	pending	
ENCF946JFF Download 8.24 MB	tsv	gene quantifications	2	1	mm10	M4	ENCODE Processing Pipeline	2015-01-06	pending	
ENCF790TWE Download 8.23 MB	tsv	gene quantifications	1	1	mm10	M4	ENCODE Processing Pipeline	2015-01-06	pending	
ENCF448JTY Download 3.42 GB	bam	alignments	1	1	mm10	M4	ENCODE Processing Pipeline	2015-01-06	pending	
ENCF689UZW Download 5.42 GB	bam	alignments	2	1	mm10	M4	ENCODE Processing Pipeline	2015-01-06	pending	
ENCF104XPZ Download 155 MB	bigWig	signal of unique reads	2	1	mm10	M4	ENCODE Processing Pipeline	2015-01-06	pending	
ENCF795QBV Download	bam	alignments	1	1	mm10	M4	ENCODE Processing Pipeline	2015-01-06	pending	

<https://www.encodeproject.org>

Programmatic access to the ENCODE Portal

<https://www.encodeproject.org/help/rest-api/>

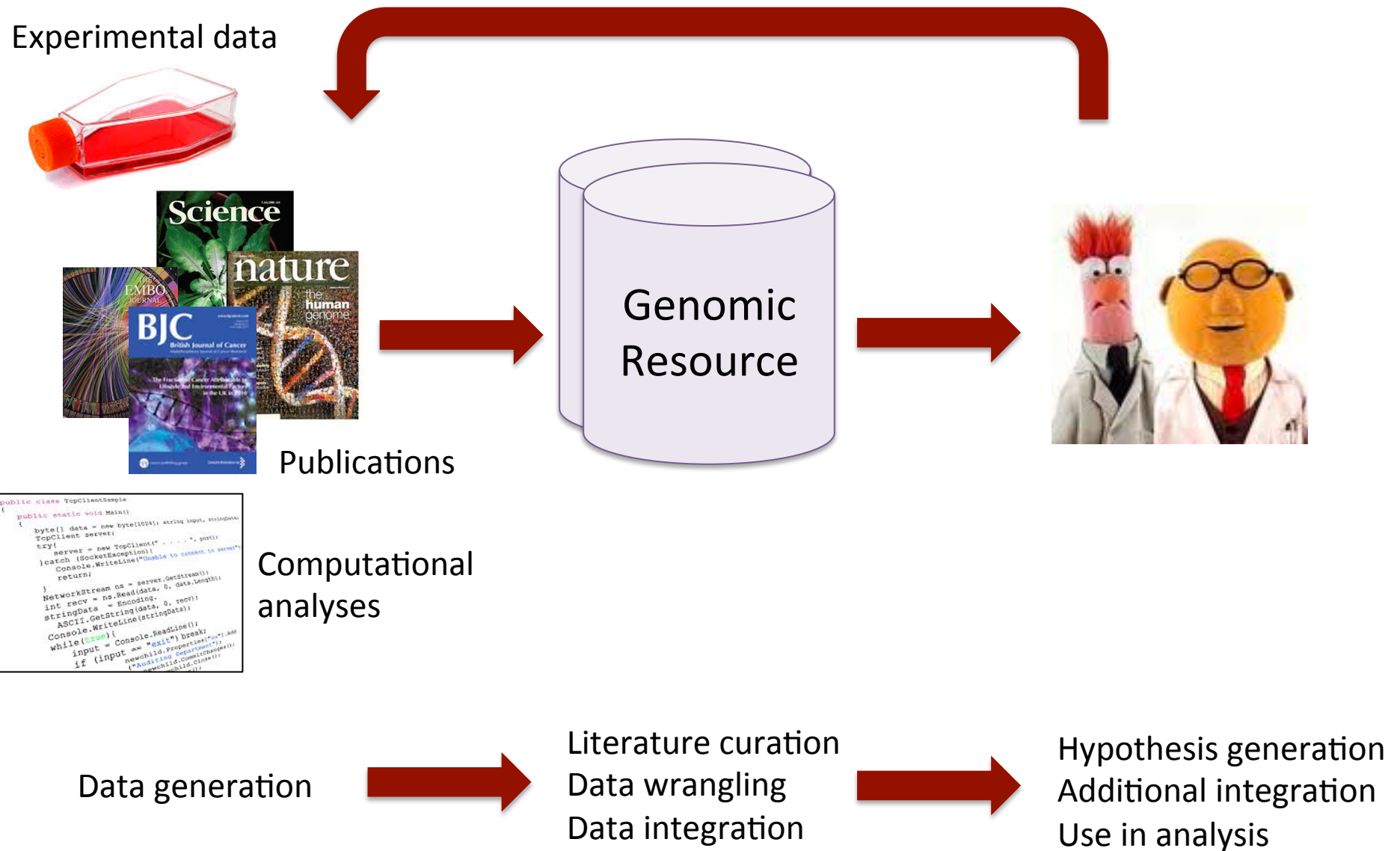
- All Portal content is accessible via URL's; just add ?format=json
- The database record is returned in JSON format
- JSON can be parsed in your language of choice

```
GET_object.py *
1  #!/usr/bin/env python
2
3  import requests
4
5  URL = 'https://www.encodeproject.org/experiments/ENCSR236EGS/?format=json'
6
7  response = requests.get(URL)
8
9  experiment = response.json()
10
11 print experiment['accession']
12 print experiment['description']
13
```

Retrieve search results via REST API

```
GET_search.py
1 #!/usr/bin/env python
2
3 import requests
4
5 URL = ('https://www.encodeproject.org/search/?'
6        'type=experiment&'
7        'assay_term_name=ChIP-seq&'
8        'replicates.library.biosample.donor.organism.scientific_name=Homo sapiens&'
9        'target.investigated_as=transcription factor&'
10       'replicates.library.biosample.biosample_type=in vitro differentiated cells&'
11       'format=json')
12
13 response = requests.get(URL)
14
15 search_result = response.json()['@graph']
16
17 #extract and print the target for each experiment
18 print '\n'.join([experiment['target']['label'] for experiment in search_result])
19
```

Role of a genomic resource



Tomorrow: ENCODE uniform processing pipelines

The image shows two screenshots of the ENCODE website. The left screenshot displays the 'Methods' dropdown menu with options: 'Data standards', 'Software tools', 'Pipelines', and 'Experiment guidelines'. A red arrow points from the 'Pipelines' option to the right screenshot. The right screenshot shows the 'RNA-seq pipeline for long RNAs' page, which includes a 'Pipeline overview' section with a list of outputs and a 'Pipeline schematic' flowchart.

ENCODE Data ▾ Methods ▾ About ENCODE ▾ Help ▾ Search ENCODE 🔍 Sign in

ENCODE Data standards Software tools Pipelines Experiment guidelines

RNA-seq pipeline for long RNAs

The ENCODE RNA-seq pipeline for long RNAs can be used if your libraries are generated from mRNAs (poly-A(+), rRNA-depleted total RNA, or poly-A(-) RNA populations that are size-selected to be longer than approximately 200 bp. The pipeline takes as inputs both RNA-seq reads (from paired-end stranded or single end unstranded libraries) and a gene annotation file (by default GENCODE), and outputs several products:

- mapping of the reads to the genome creates an alignment file in bam file format
- mapping of the reads to the transcriptome creates a transcriptome alignment file in bam file format
- normalized RNA-seq signal for each strand (plus and minus) for unique reads and for unique and multimapping reads in bigwig file format
- gene quantifications as a tsv file
- transcript quantifications as a tsv file

The mapping of the reads is done using the STAR program and the quantification of genes and transcripts is done with the RSEM program. Although there is general agreement between the mappings and the gene quantifications produced by different RNAseq pipelines, quantifications of individual transcript isoforms, being much more complex, can differ substantially depending on the processing pipeline employed, and are of unknown accuracy. Therefore, mapping and gene quantifications can be used confidently, while transcript quantifications should be used with care.

For paired-end, strand-specific library preparation: [View pipeline](#) | [Download software](#) | [Run pipeline at DNAnexus](#)
For single-end, unstranded library preparations: [View pipeline](#) | [Download software](#) | [Run pipeline at DNAnexus](#)
Find data generated by this pipeline: [All](#) | [paired-end only](#) | [single-end only](#)

Pipeline schematic

```
graph LR; subgraph Path1; A1[genome indexing STAR] --> B1[fastq concatenation, alignment Concat-fastq, STAR, Samtools]; end; subgraph Path2; A2[genome indexing TopHat, Bowtie] --> B2[fastq concatenation, alignment Concat-fastq, TopHat, Bowtie, Tophat BAM Repair]; end; B1 --> C[quantification RSEM]; B2 --> D[signal generation, file format conversion STAR, bedGraphToBigWig];
```

Reminder: See the 'Prepare to run web-based pipelines' PDF to prepare for tomorrow's workshop.

Tomorrow: ENCODE genomic annotations

ENCODE [Data](#) [Methods](#) [About ENCODE](#) [Help](#) [Sign in](#)

ENCODE Encyclopedia of DNA Elements

- Assays
- Biosamples
- Antibodies
- Annotations
- Release policy

Quick Start

To find and download ENCODE Consortium data:

- Click the Data toolbar above and browse data
 - By assay
 - By biosample
 - By genomic annotations
- Enter search terms like "skin", "ChIP-seq", or "CTCF"

Additional help about the ENCODE Portal:

- Getting Started

Genomic annotations

Introduction

The ENCODE Project provides a set of candidate genomic regions that can serve as predictions for further investigation. This page provides links to download a set of genomic annotations as well as a list of publications that contain additional data.

Annotated genomic regions

- Gene expression matrix over ENCODE2 cell lines (~60 cell lines in total) in GENCODE 19 [[Download data](#) | [Download methods](#)]
- Transcription start site (TSS) lists [[View README](#)]
 - GENCODE v19 TSS [[Download](#)]
 - GENCODE v19 TSS stratified by strict Fantom5 CAGE clusters [[Download](#)]
 - GENCODE v19 TSS stratified by robust Fantom5 CAGE clusters [[Download](#)]
 - GENCODE v19 TSS stratified by permissive Fantom5 CAGE clusters [[Download](#)]
- Candidate enhancers based on DNase hypersensitivity and H3K27ac and annotated with TF-ChIP peaks as well as candidate promoters annotated with TF-ChIP peaks. [[Visualize data](#) | [Download methods](#)]
 - Distal DNase peaks [[Download](#)]
 - Proximal DNase peaks [[Download](#)]
 - H3K27ac annotations [[Download](#)]
 - Distal TF binding sites [[Download](#)]
 - Proximal TF binding sites [[Download](#)]

Additional annotations

Papers previously published by the ENCODE Consortium contain data files that include additional genomic annotations. [Search for all publications with ENCODE element data.](#)

Peaks

Peaks are enriched regions of the genome corresponding to either sites of transcription factor binding or DNase hypersensitivity identified during various functional genomic assays. In this section, we provide a list of peaks in various cell lines using both DNase-Seq and ChIP-Seq assays. [View publications.](#)

RNAs

RNA represents the direct readout of the genetic information encoded by genomes and a significant proportion of a cell's regulatory capabilities are focused on its synthesis, processing, transport, modification and translation. A catalogue of the RNA species made inside the cell and the amount of RNA from each of these loci across various cell lines is provided in this section. [View publications.](#)

Coming soon: Search ENCODE data by genomic region

Search ENCODE data by region

chr15:40986872-40986972

Assay

DNase-seq	206
ChIP-seq	81

Biosample term

GM12878	29
K562	29
HeLa-S3	10
MCF-7	9
skin fibroblast	9

[+ See more...](#)

Target

POLR2A	10
YY1	6
MYC	5
CHD2	4
MAX	4

[+ See more...](#)

Showing 10 of 287

[List view](#) [Browser view](#)

15:40,977,143..40,997,243

100bp 2kb 50kb 500kb

Genome 40,980,000 40,985,000 40,990,000 40,995,000

GENCODE **<RAD51-AS1**

>RAD51

Repeats

peaks

peaks

peaks

hotspots

peaks

hotspots

peaks

hotspots

peaks

ENCODE DCC



Eurie Hong



Mike Cherry (PI)



Jim Kent (co-PI)



Ben Hitz

Data wranglers



Esther Chan



Jean Davidson



Cricket Sloan



Seth Strattan

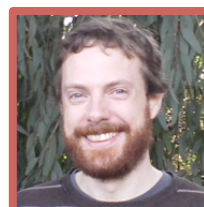
Software engineers



Tim Dreszer



Nikhil Podduturi



Laurence Rowe



Forrest Tanaka

Biocurator assistant



Marcus Ho



Aditi Narayanan

QA, systems, admin



Brian Lee



Stuart Miyasato



Matt Simison



Zhenhua Wang

Brian Lee, Stuart Miyasato, Matt Simison, Zhenhua Wang