
Searching the ENCODE Portal

Asia-Pacific Bioinformatics Conference 2016
San Francisco, California

Aditi Narayanan, ENCODE DCC
Department of Genetics, Stanford University



Topics for this workshop

1. **“Getting started” and additional resources on the ENCODE portal**
2. **Faceted browsing and metadata access for assays**
3. **Keyword searches**
4. **Access to an individual object’s metadata and individual data files**
5. **Faceted browsing of Biosamples and Antibodies**
6. **Independent hands-on exercise**

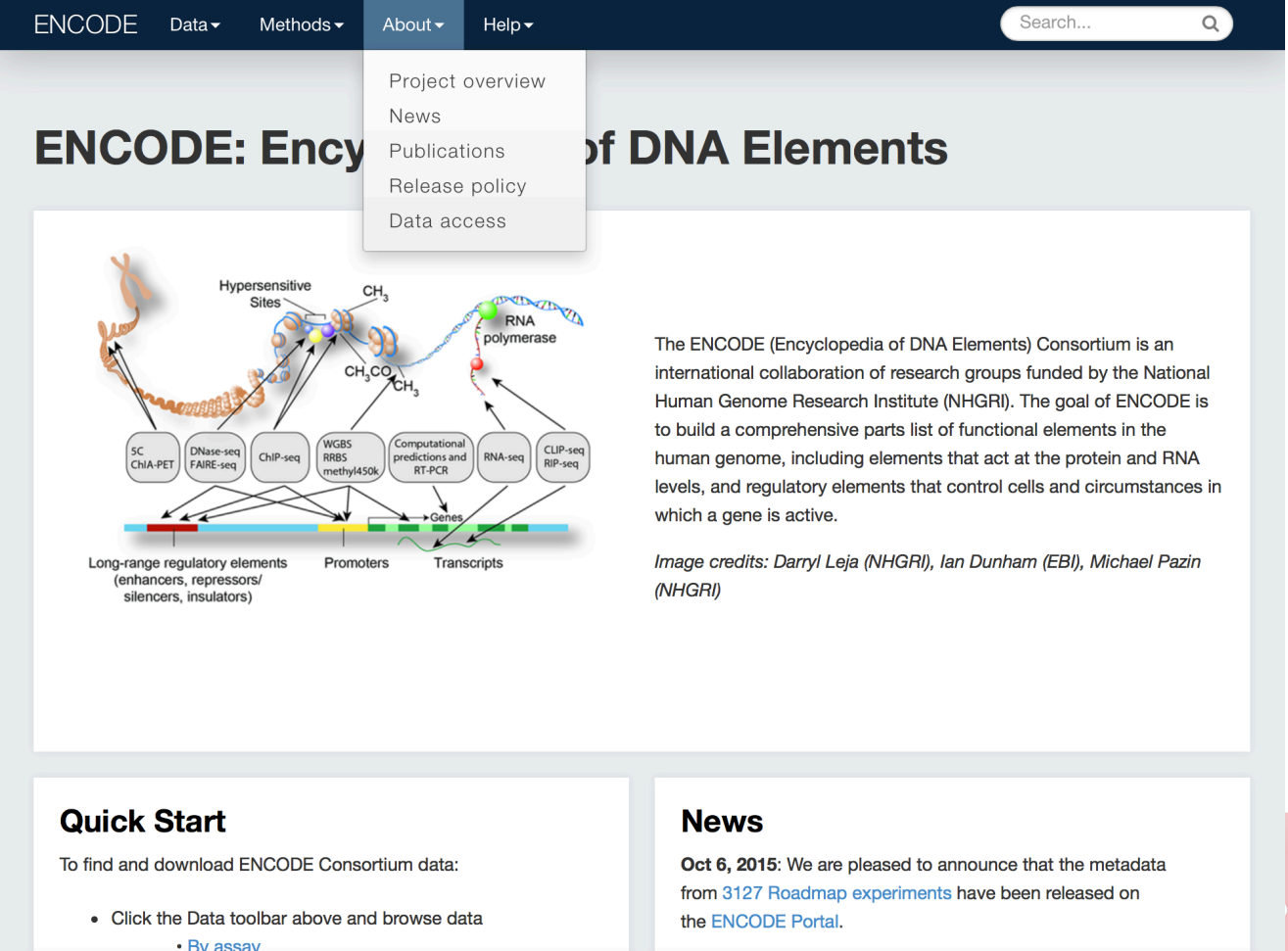


Getting Started

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

About tab points of interest

- *Publications*: How ENCODE data is used by members of the consortium and the general public
- *Data Access and Project Overview*: ENCODE consortium and partnerships with other projects

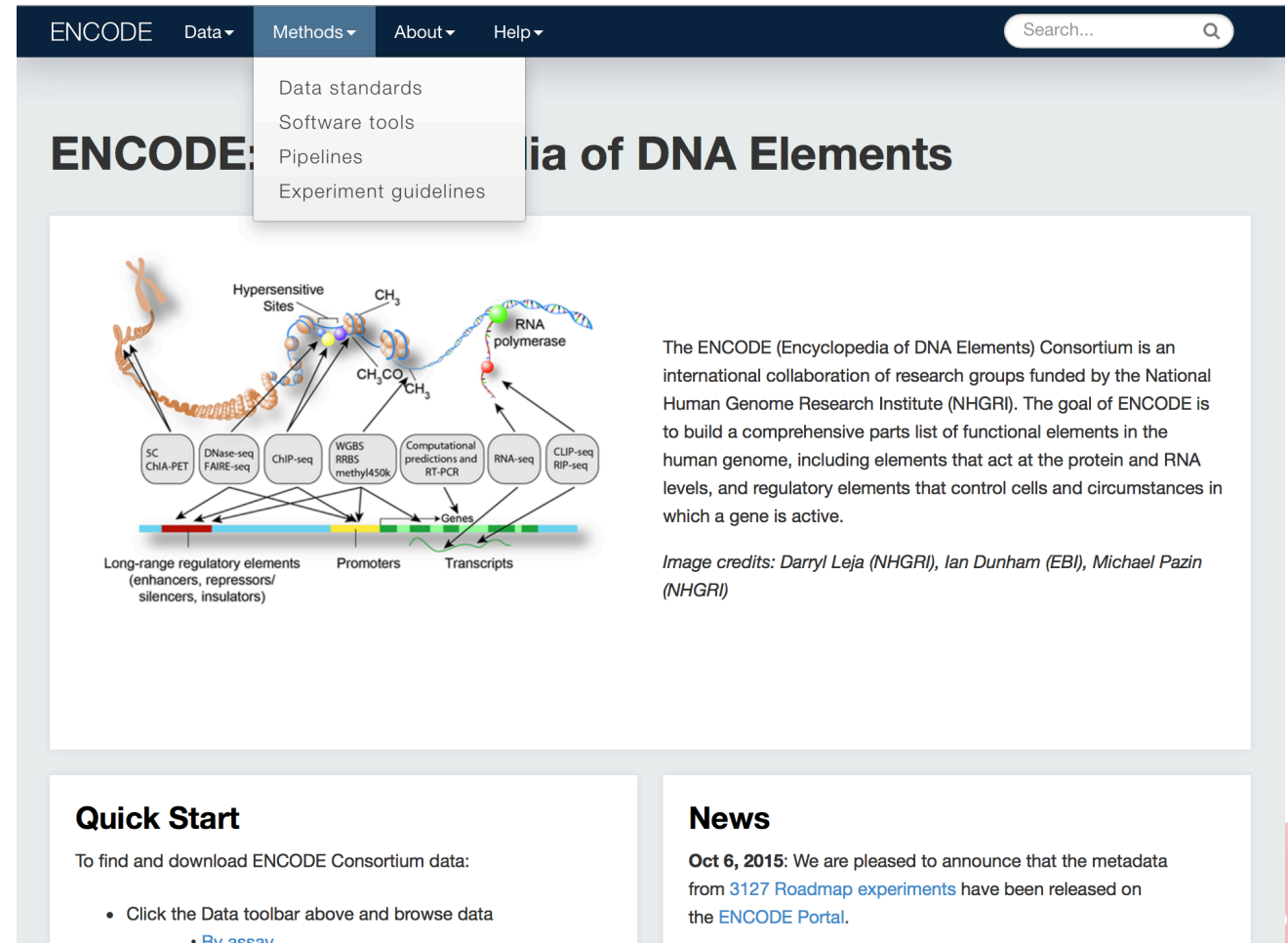


The screenshot shows the ENCODE website interface. At the top, there is a navigation bar with 'ENCODE', 'Data', 'Methods', 'About', and 'Help' menus, along with a search bar. The 'About' menu is open, showing options: 'Project overview', 'News', 'Publications', 'Release policy', and 'Data access'. Below the navigation bar, the main heading reads 'ENCODE: Encyclopedia of DNA Elements'. A central diagram illustrates the genome with various elements: 'Long-range regulatory elements (enhancers, repressors/silencers, insulators)', 'Promoters', and 'Transcripts'. Above the diagram, several assays are listed in boxes: 'SC ChIA-PET', 'DNase-seq FAIRE-seq', 'ChIP-seq', 'WGBS RRBS methyl450k', 'Computational predictions and RT-PCR', 'RNA-seq', and 'CLIP-seq RIP-seq'. The diagram also shows 'Hypersensitive Sites', 'CH₃CO', 'CH₃', and 'RNA polymerase'. To the right of the diagram, a text block explains the ENCODE Consortium's goal: '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.' Below this text, it says 'Image credits: Darryl Leja (NHGRI), Ian Dunham (EBI), Michael Pazin (NHGRI)'. At the bottom of the screenshot, there are two sections: 'Quick Start' with instructions on how to find and download data, and 'News' with a recent announcement from October 6, 2015, regarding the release of metadata from 3127 Roadmap experiments.

Getting Started

Methods tab points of interest

- *Data Standards and Experiment Guidelines:* ENCODE's standards and requirements for submitting labs.
- *Software and Pipelines:* Data processing tools

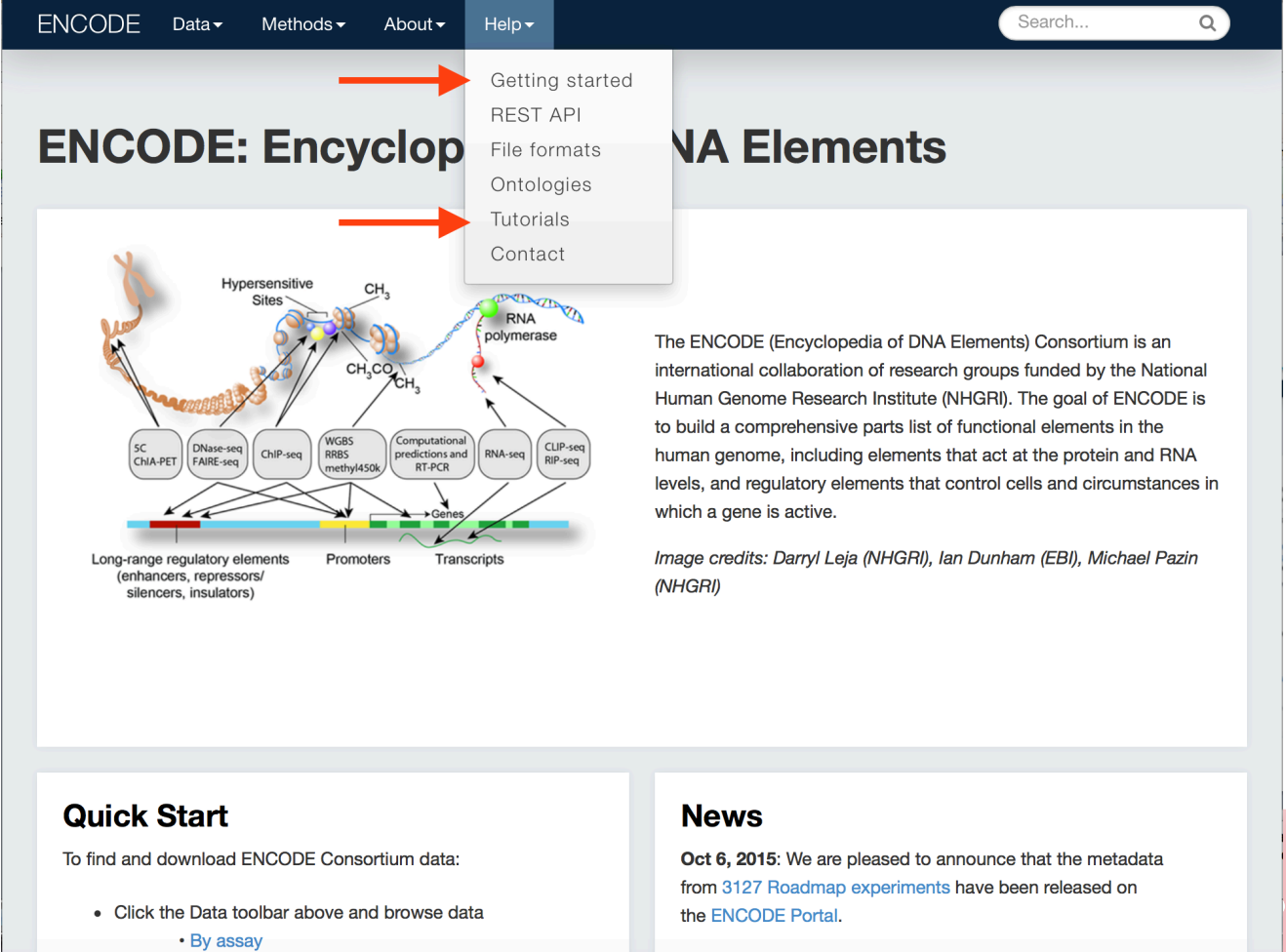


The screenshot shows the ENCODE website interface. The top navigation bar includes 'ENCODE', 'Data', 'Methods', 'About', and 'Help'. A search bar is located on the right. The 'Methods' dropdown menu is open, showing options: 'Data standards', 'Software tools', 'Pipelines', and 'Experiment guidelines'. The main content area features the title 'ENCODE: Encyclopedia of DNA Elements' and a diagram illustrating the relationship between various genomic features and experimental methods. The diagram shows a DNA strand with 'Long-range regulatory elements (enhancers, repressors/silencers, insulators)', 'Promoters', and 'Transcripts'. Above the DNA, 'Hypersensitive Sites' are marked with blue dots, and 'CH₃' and 'CH₃CO' methyl groups are shown. 'RNA polymerase' is depicted as a red and blue structure. Below the DNA, several experimental methods are listed in boxes: '5C ChIA-PET', 'DNase-seq FAIRE-seq', 'CHIP-seq', 'WGBS RRBS methyl450k', 'Computational predictions and RT-PCR', 'RNA-seq', and 'CLIP-seq RIP-seq'. Arrows indicate the mapping from these methods to the corresponding genomic features. To the right of the diagram, a text block describes the ENCODE Consortium's goal: '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.' Below this text, it says 'Image credits: Darryl Leja (NHGRI), Ian Dunham (EBI), Michael Pazin (NHGRI)'. At the bottom of the page, there are two sections: 'Quick Start' and 'News'. The 'Quick Start' section provides instructions on how to find and download ENCODE Consortium data, with a bullet point: 'Click the Data toolbar above and browse data' and a link 'By assav'. The 'News' section has a date 'Oct 6, 2015:' and text: 'We are pleased to announce that the metadata from 3127 Roadmap experiments have been released on the ENCODE Portal.'

Getting Started

Help tab points of interest

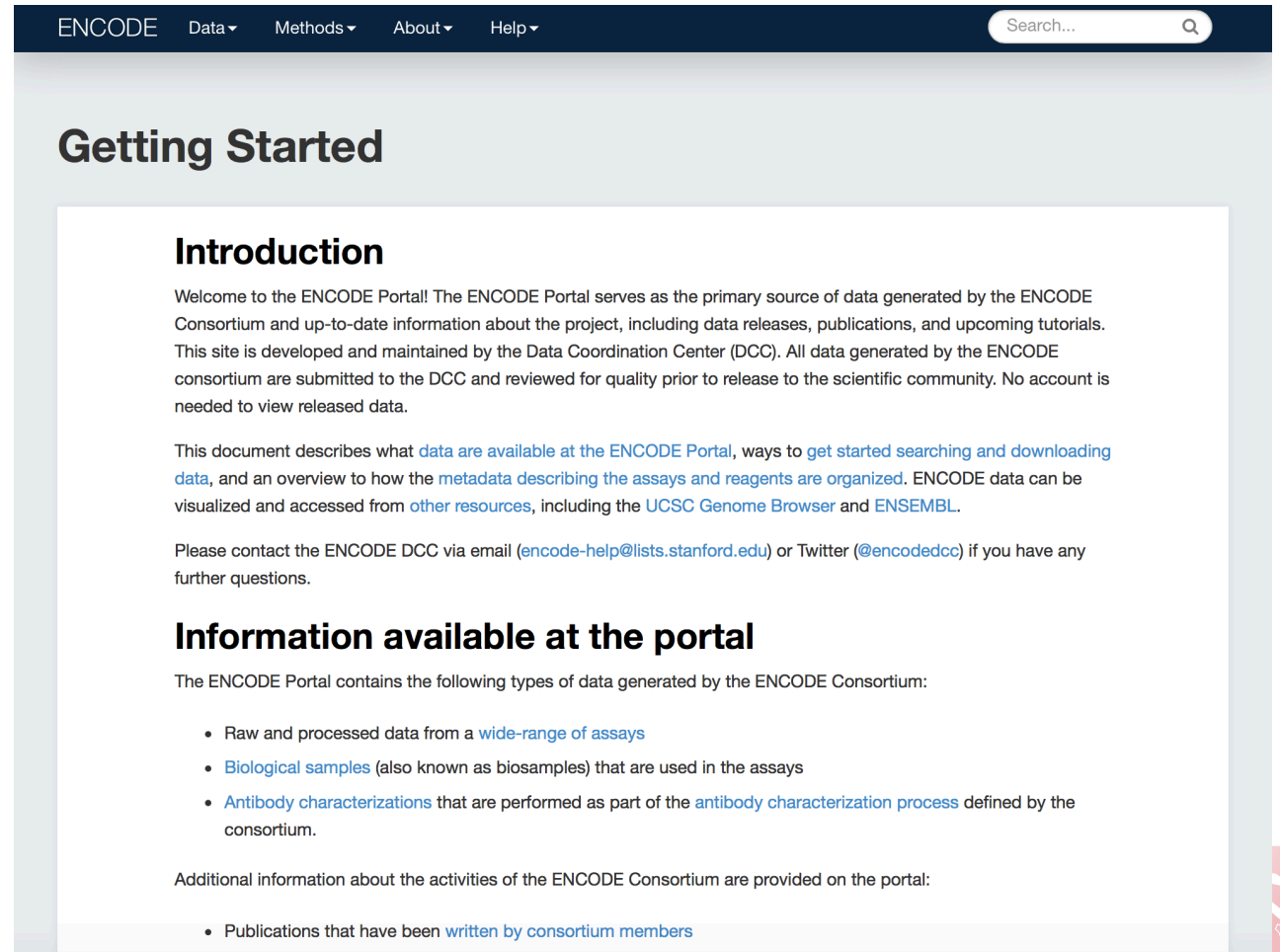
- *Getting Started*: Data organization and access
 - <https://www.encodeproject.org/help/getting-started>
- *Tutorials*:
 - <https://www.encodeproject.org/tutorials>
 - Links to slides from today's workshop
- *Contact*:
 - encode-help@lists.stanford.edu
 - [@encodedcc](#)



The screenshot shows the ENCODE website interface. At the top, there is a navigation bar with 'ENCODE' and dropdown menus for 'Data', 'Methods', 'About', and 'Help'. A search bar is located on the right. The 'Help' dropdown menu is open, showing options: 'Getting started', 'REST API', 'File formats', 'Ontologies', 'Tutorials', and 'Contact'. Two red arrows point to 'Getting started' and 'Tutorials'. Below the navigation bar, the main content area features a large diagram titled 'ENCODE: Encyclopedia of DNA Elements'. The diagram illustrates various genomic features: 'Long-range regulatory elements (enhancers, repressors/silencers, insulators)', 'Promoters', and 'Transcripts'. Above the diagram, 'Hypersensitive Sites' are shown with 'CH₃' and 'CH₃CO₂' modifications. 'RNA polymerase' is also depicted. Below the diagram, several assay types are listed in boxes: 'SC ChIA-PET', 'DNase-seq FAIRE-seq', 'ChIP-seq', 'WGBS RRBS methyl450k', 'Computational predictions and RT-PCR', 'RNA-seq', and 'CLIP-seq RIP-seq'. To the right of the diagram, a text block describes the ENCODE Consortium's goal: '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.' Below this text, it says 'Image credits: Darryl Leja (NHGRI), Ian Dunham (EBI), Michael Pazin (NHGRI)'. At the bottom of the page, there are sections for 'Quick Start' and 'News'. The 'Quick Start' section provides instructions on how to find and download ENCODE Consortium data, with a bullet point: 'Click the Data toolbar above and browse data' and a sub-bullet 'By assay'. The 'News' section has a date 'Oct 6, 2015' and text: 'We are pleased to announce that the metadata from 3127 Roadmap experiments have been released on the ENCODE Portal.'

Getting Started

- <https://www.encodeproject.org/help/getting-started>
 - A good place to start
 - How to browse and facet
 - How to access one file at a time
 - Bulk download of files
 - ENCODE data model and data organization



The screenshot shows the ENCODE Portal's 'Getting Started' page. The navigation bar includes 'ENCODE', 'Data', 'Methods', 'About', and 'Help', along with a search bar. The main heading is 'Getting Started'. The 'Introduction' section welcomes users and explains the portal's role as the primary source of data. It also provides links for 'get started searching and downloading data' and 'metadata describing the assays and reagents are organized'. The 'Information available at the portal' section lists data types: raw and processed data, biological samples, and antibody characterizations. A final section mentions publications written by consortium members.

ENCODE Data Methods About Help Search...

Getting Started

Introduction

Welcome to the ENCODE Portal! The ENCODE Portal serves as the primary source of data generated by the ENCODE Consortium and up-to-date information about the project, including data releases, publications, and upcoming tutorials. This site is developed and maintained by the Data Coordination Center (DCC). All data generated by the ENCODE consortium are submitted to the DCC and reviewed for quality prior to release to the scientific community. No account is needed to view released data.

This document describes what [data are available at the ENCODE Portal](#), ways to [get started searching and downloading data](#), and an overview to how the [metadata describing the assays and reagents are organized](#). ENCODE data can be visualized and accessed from [other resources](#), including the [UCSC Genome Browser](#) and [ENSEMBL](#).

Please contact the ENCODE DCC via email (encode-help@lists.stanford.edu) or Twitter ([@encodedcc](https://twitter.com/encodedcc)) if you have any further questions.

Information available at the portal

The ENCODE Portal contains the following types of data generated by the ENCODE Consortium:

- Raw and processed data from a [wide-range of assays](#)
- [Biological samples](#) (also known as biosamples) that are used in the assays
- [Antibody characterizations](#) that are performed as part of the [antibody characterization process](#) defined by the consortium.

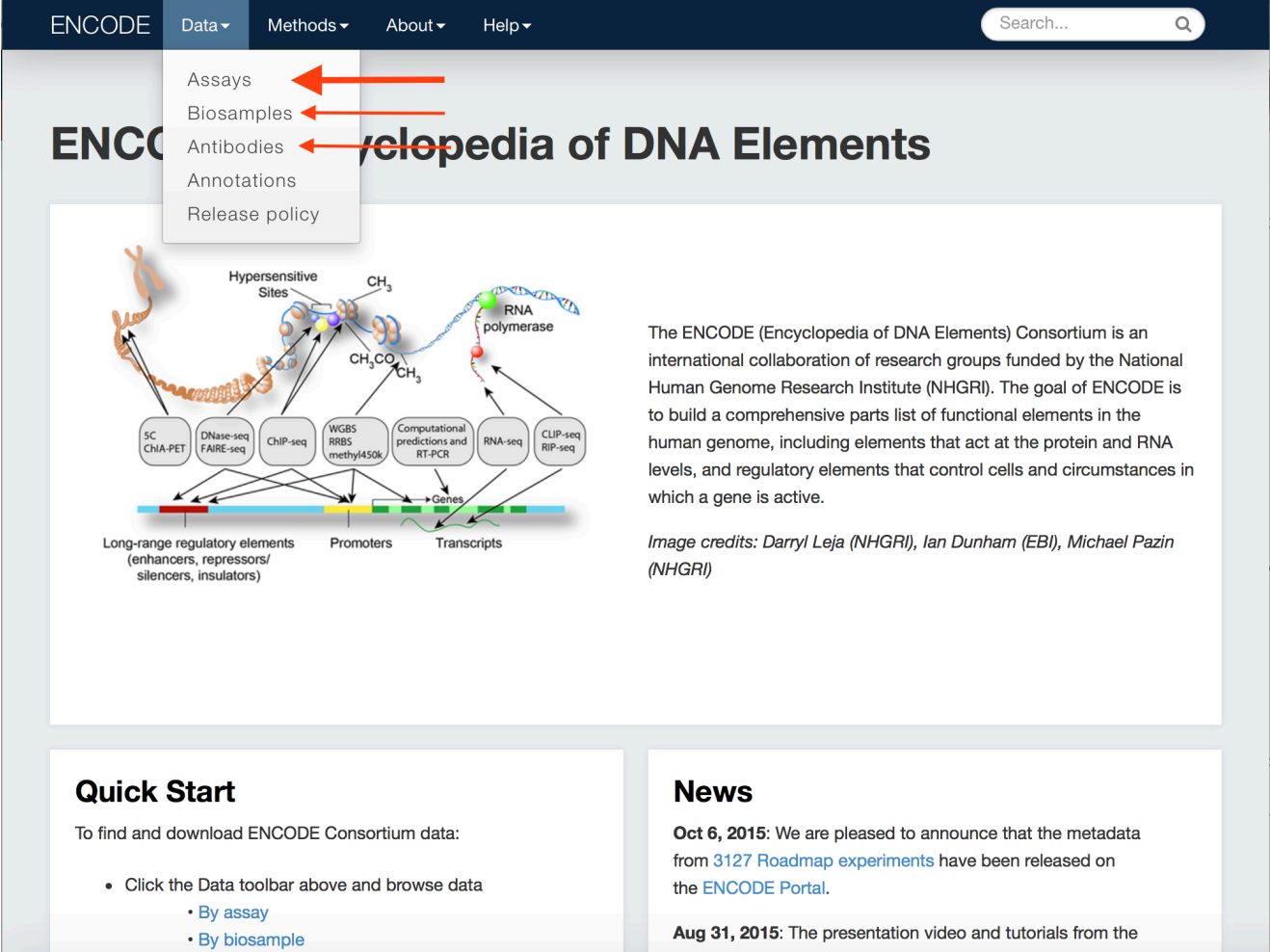
Additional information about the activities of the ENCODE Consortium are provided on the portal:

- Publications that have been [written by consortium members](#)



Getting Started

- **Data** tab points of interest
 - *Assays*: access to data, will be most important
 - *Biosamples and Antibodies*: Used in assays
 - *Annotations*: Broader biological conclusions



The screenshot shows the ENCODE website interface. The 'Data' dropdown menu is open, with red arrows pointing to 'Assays', 'Biosamples', and 'Antibodies'. The main content area features a diagram of DNA elements with various annotations and a 'Quick Start' section.

ENCODE Data Methods About Help Search...

ENCODE Encyclopedia of DNA Elements

Assays
Biosamples
Antibodies
Annotations
Release policy

Hypersensitive Sites CH₃ CH₃CO CH₃ RNA polymerase

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

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

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](#)

News

Oct 6, 2015: We are pleased to announce that the metadata from [3127 Roadmap experiments](#) have been released on the [ENCODE Portal](#).

Aug 31, 2015: The presentation video and tutorials from the [ENCODE 2015: Research Applications and User Meeting](#) have

Faceted Browsing: Experiments

Faceted Browsing of Assays

- Click on Data ... Assays
- <https://www.encodeproject.org/search/?type=Experiment>
- Facets are “filters”
- Limit to items of interest
- Facets from the same category can be stacked to combine results (logical OR)
- Facets from different categories further restrict results (logical AND)

The screenshot displays the ENCODE project search interface. The top navigation bar includes 'ENCODE', 'Data', 'Methods', 'About', and 'Help', along with a search bar. The main content area is divided into two columns. The left column shows faceted search results for 'Assay', 'Project', 'Experiment status', 'Genome assembly (visualization)', 'Organism', 'Biosample type', and 'Organ'. The right column shows a list of 25 results for RNA-seq experiments, including details like 'RNA-seq of stomach (Mus musculus, embryonic 16.5 day)', 'Lab: Barbara Wold, Caltech', and 'Project: ENCODE'. Buttons for 'Visualize', 'Download', and 'View All' are visible at the top right of the results section.

Facet Category	Facet Value	Count
Assay	ChIP-seq	383
	RNA-seq	96
	single cell isolation followed by RNA-seq	56
	microRNA-seq	20
	microRNA profiling by array assay	18
Project	ENCODE	96
Experiment status	released	96
Genome assembly (visualization)	hg19	395
	mm9	99
	mm10	96
Organism	Mus musculus	96
Biosample type	tissue	80
	primary cell	12
	immortalized cell line	4
Organ	brain	31
	liver	9
	heart	8
	kidney	4
	lung	4
	+ See more	

Assay	Project	Experiment status	Genome assembly (visualization)	Organism	Biosample type	Organ
RNA-seq of stomach (<i>Mus musculus</i> , embryonic 16.5 day)	ENCODE	released	mm10	Mus musculus	tissue	stomach
RNA-seq of kidney (<i>Mus musculus</i> , embryonic 16.5 day)	ENCODE	released	mm10	Mus musculus	tissue	kidney
RNA-seq of intestine (<i>Mus musculus</i> , embryonic 16.5 day)	ENCODE	released	mm10	Mus musculus	tissue	intestine
RNA-seq of lung (<i>Mus musculus</i> , embryonic 16.5 day)	ENCODE	released	mm10	Mus musculus	tissue	lung
RNA-seq of liver (<i>Mus musculus</i> , embryonic 16.5 day)	ENCODE	released	mm10	Mus musculus	tissue	liver
RNA-seq of heart (<i>Mus musculus</i> , embryonic 16.5 day)	ENCODE	released	mm10	Mus musculus	tissue	heart
RNA-seq of hindbrain (<i>Mus musculus</i> , embryonic 16.5 day)	ENCODE	released	mm10	Mus musculus	tissue	brain

Keyword Searches

The screenshot shows the ENCODE website interface with a search bar containing the word "skin". The search results are displayed in a list format. On the left side, there are several filter categories with horizontal bar charts and counts:

- Assay**: ChIP-seq (113), RNA-seq (58), DNase-seq (57), transcription profiling by array assay (24), CAGE (9). A link "+ See more..." is present.
- Project**: ENCODE (174), Roadmap (151), modENCODE (1).
- Experiment status**: released (325), revoked (1).
- Genome assembly (visualization)**: hg19 (154), mm9 (2).
- Organism**: *Homo sapiens* (322), *Mus musculus* (2), *Drosophila melanogaster* (1).

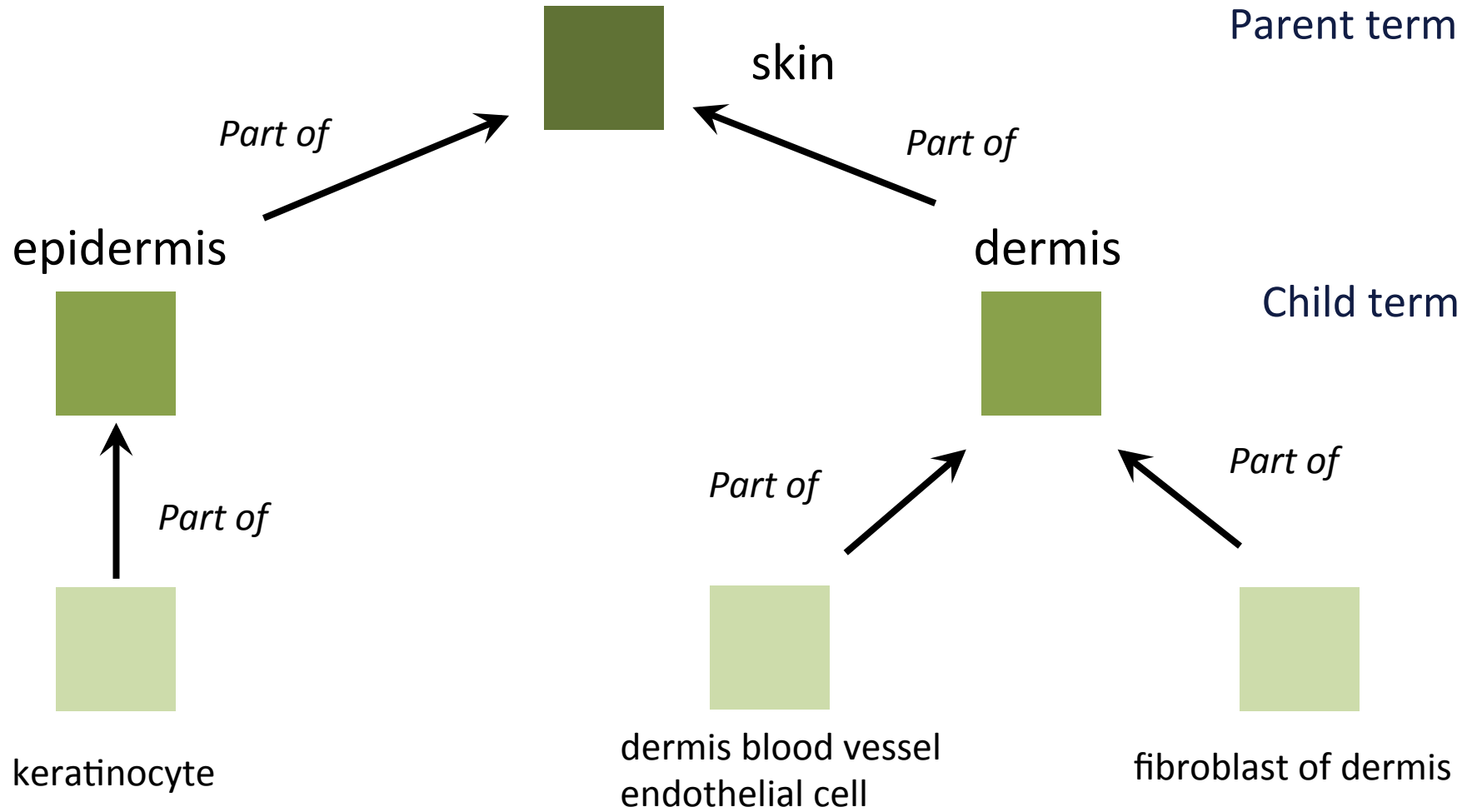
The main search results area shows "Showing 326 of 326 results" and three buttons: "Visualize", "Download", and "View 25". The results list includes:

- Repli-seq of keratinocyte (*Homo sapiens*)**: Experiment ENCSR000CXC, released. Lab: John Stamatoyannopoulos, UW. Project: ENCODE.
- DNase-seq of skin fibroblast (*Homo sapiens*, adult 31 year)**: Experiment ENCSR000EMO, released. Lab: John Stamatoyannopoulos, UW. Project: ENCODE.
- DNase-seq of dermis blood vessel endothelial cell (*Homo sapiens*)**: Experiment ENCSR000ENX, released. Lab: John Stamatoyannopoulos, UW. Project: ENCODE.
- transcription profiling by array assay of dermis microvascular lymphatic vessel endothelial cell (*Homo sapiens*)**: Experiment ENCSR000CZT, released. Lab: John Stamatoyannopoulos, UW. Project: ENCODE.

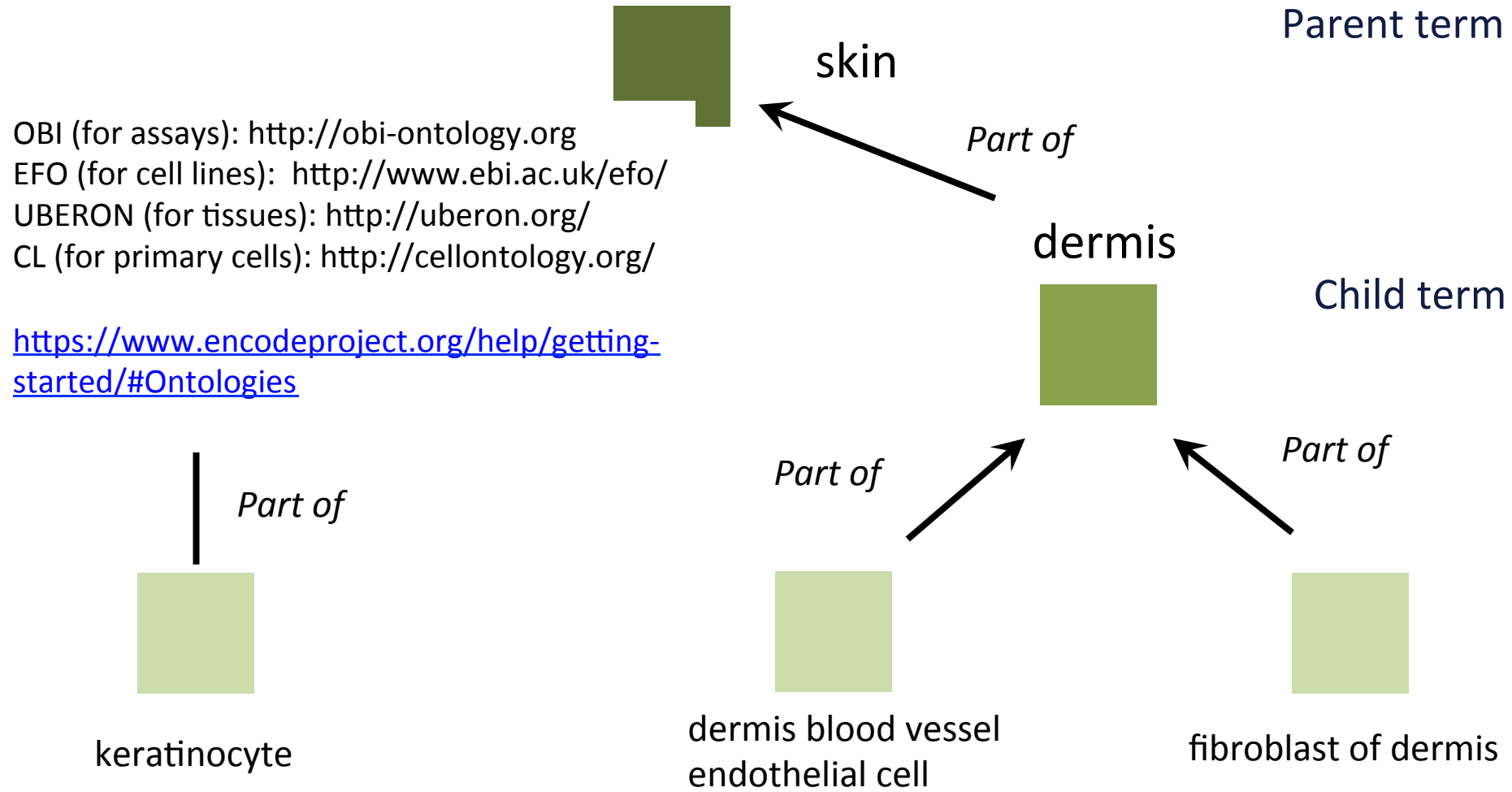
- Search for “Skin”
- Some of these search hits don’t contain the word skin...



Keyword Searches using Ontologies



Keyword Searches using Ontologies



Metadata to Data: Experiments

Use metadata to find data

- **Filter: RNAseq, ENCODE, mm10 assembly**
- Select an experiment:
<https://www.encodeproject.org/experiments/ENCSR466KZY/>
- Note metadata on protocols, replicates, etc.
- Graph: files are related by processing steps
- Download files from the graph or a list
 - Not all experiments have the graph yet

ENCODE Data Methods About Help Search...

Showing 25 of 96 results Visualize Download View All

Assay

ChIP-seq	383
RNA-seq	96
single cell isolation followed by RNA-seq	56
microRNA-seq	20
microRNA profiling by array assay	18

Project

ENCODE	96
---------------	-----------

Experiment status

released	96
----------	----

Genome assembly (visualization)

hg19	395
mm9	99
mm10	96

Organism

Mus musculus	96
---------------------	-----------

Biosample type

tissue	80
primary cell	12
immortalized cell line	4

Organ

brain	31
liver	9
heart	8
kidney	4
lung	4

RNA-seq of stomach (*Mus musculus*, embryonic 16.5 day) Experiment
Lab: Barbara Wold, Caltech ENCSR466KZY released
Project: ENCODE

RNA-seq of kidney (*Mus musculus*, embryonic 16.5 day) Experiment
Lab: Barbara Wold, Caltech ENCSR537GNQ released
Project: ENCODE

RNA-seq of intestine (*Mus musculus*, embryonic 16.5 day) Experiment
Lab: Barbara Wold, Caltech ENCSR848GST released
Project: ENCODE

RNA-seq of lung (*Mus musculus*, embryonic 16.5 day) Experiment
Lab: Barbara Wold, Caltech ENCSR992WBR released
Project: ENCODE

RNA-seq of liver (*Mus musculus*, embryonic 16.5 day) Experiment
Lab: Barbara Wold, Caltech ENCSR826HIQ released
Project: ENCODE

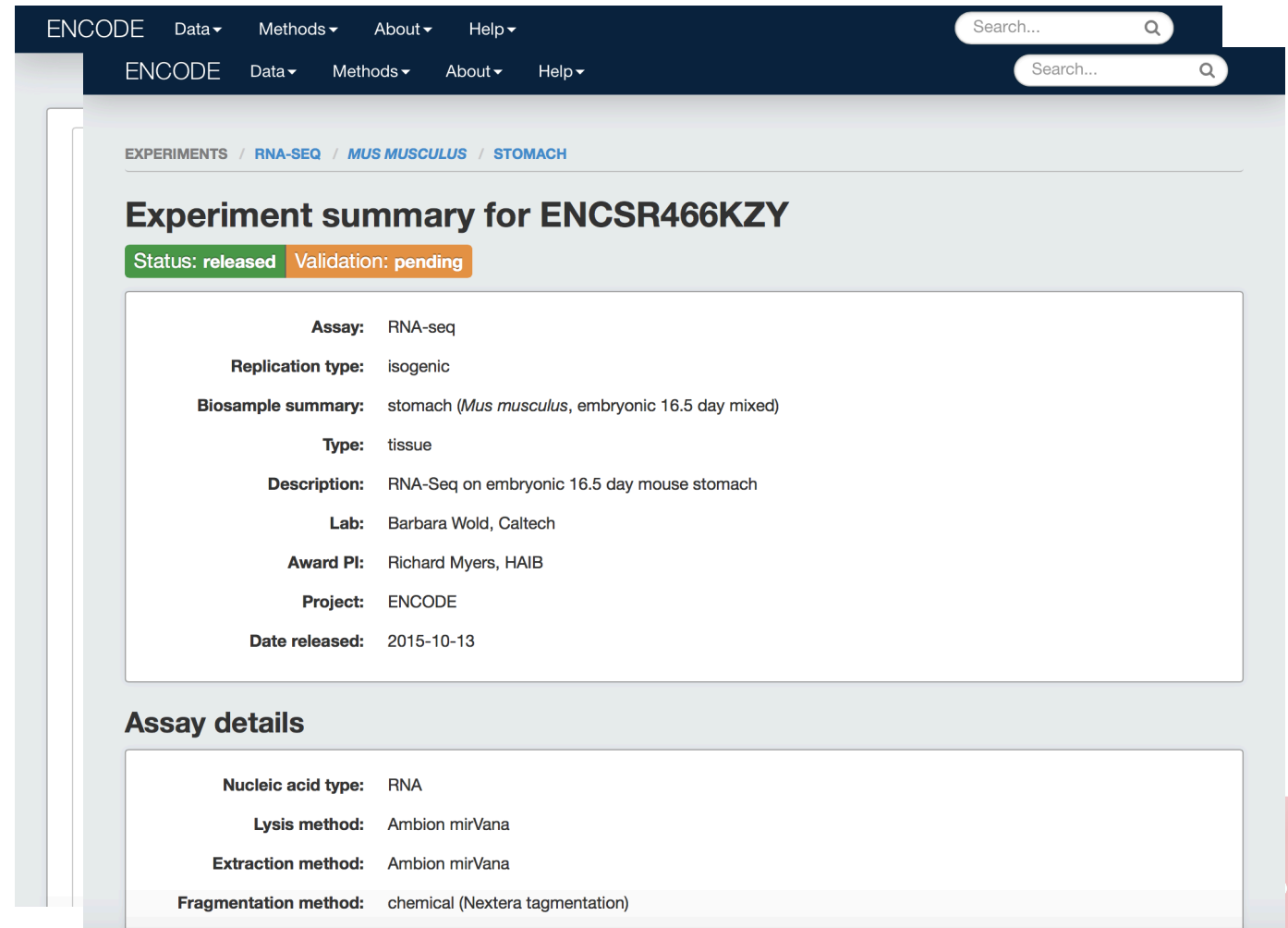
RNA-seq of heart (*Mus musculus*, embryonic 16.5 day) Experiment
Lab: Barbara Wold, Caltech ENCSR020DGG released
Project: ENCODE

RNA-seq of hindbrain (*Mus musculus*, embryonic 16.5 day) Experiment
Lab: Barbara Wold, Caltech ENCSR285WZV released
Project: ENCODE

Metadata to Data: Experiments

Use metadata to find data

- Filter: RNAseq, ENCODE, mm10 assembly
- **Select an experiment:**
<https://www.encodeproject.org/experiments/ENCSR466KZY/>
- **Note metadata on protocols, replicates, etc.**
- Graph: files are related by processing steps
- Download files from the graph or a list
 - Not all experiments have the graph yet



The screenshot displays the ENCODE project website interface. At the top, there is a navigation bar with 'ENCODE' and dropdown menus for 'Data', 'Methods', 'About', and 'Help'. A search bar is located on the right. Below the navigation bar, the breadcrumb trail reads 'EXPERIMENTS / RNA-SEQ / MUS MUSCULUS / STOMACH'. The main heading is 'Experiment summary for ENCSR466KZY'. Below the heading, there are two status indicators: 'Status: released' (in a green box) and 'Validation: pending' (in an orange box). The main content area contains a table of metadata:

Assay:	RNA-seq
Replication type:	isogenic
Biosample summary:	stomach (<i>Mus musculus</i> , embryonic 16.5 day mixed)
Type:	tissue
Description:	RNA-Seq on embryonic 16.5 day mouse stomach
Lab:	Barbara Wold, Caltech
Award PI:	Richard Myers, HAIB
Project:	ENCODE
Date released:	2015-10-13

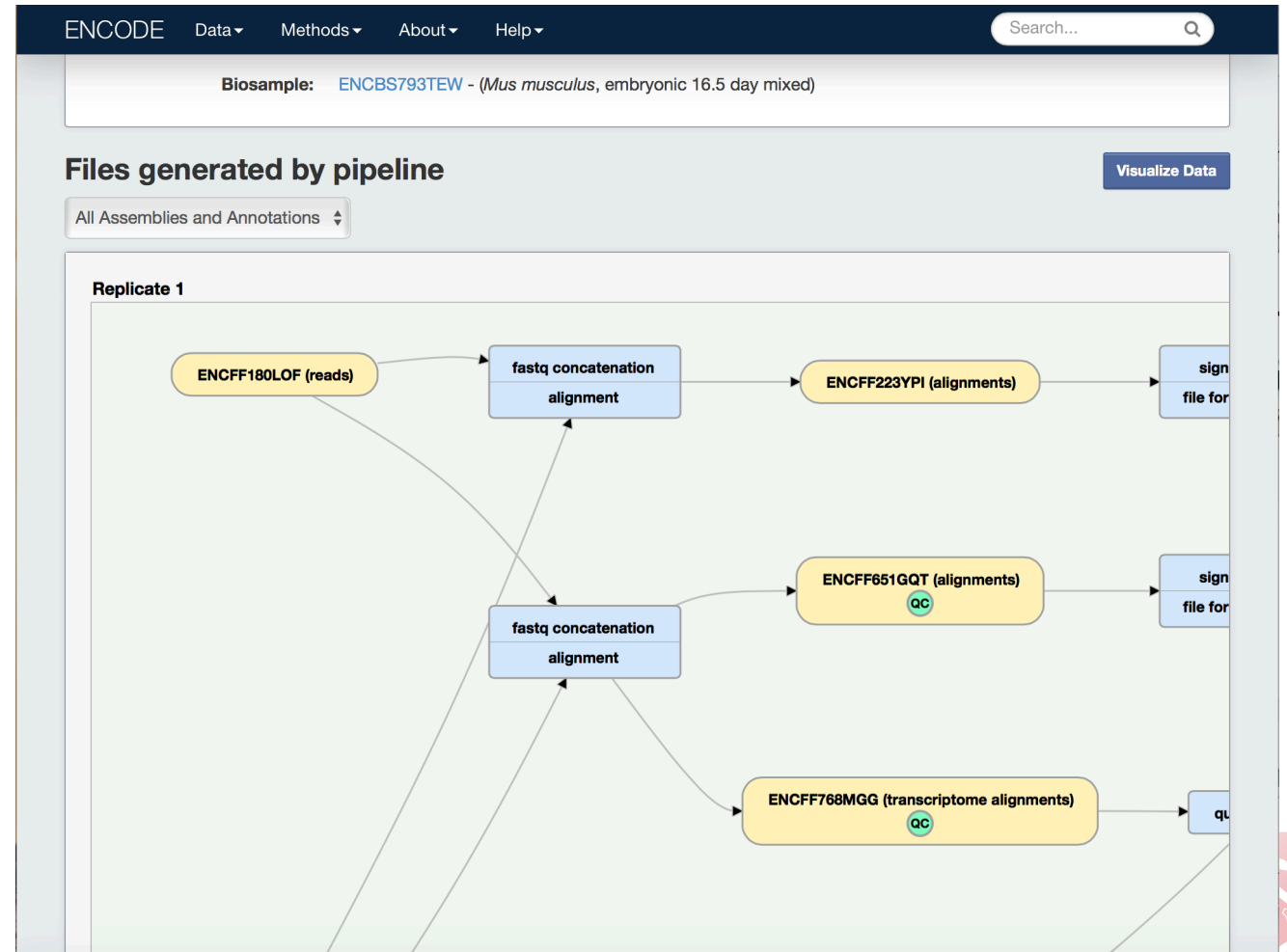
Below the metadata table, there is a section titled 'Assay details' with the following information:

Nucleic acid type:	RNA
Lysis method:	Ambion mirVana
Extraction method:	Ambion mirVana
Fragmentation method:	chemical (Nextera tagmentation)

Metadata to Data: Experiments

Use metadata to find data

- Filter: RNAseq, ENCODE, mm10 assembly
- Select an experiment:
<https://www.encodeproject.org/experiments/ENCSR466KZY/>
- Note metadata on protocols, replicates, etc.
- **Graph: files are related by processing steps**
- **Download files from the graph or a list**
 - Not all experiments have the graph yet



Metadata to Data: Experiments

Use metadata to find data

- Filter: RNAseq, ENCODE, mm10 assembly
- Select an experiment:

<https://www.encodeproject.org/experiments/ENCSR466KZY/>

- Note metadata on protocols, replicates, etc.
- **Graph: files are related by processing steps**
- **Download files from the graph or a list**
 - Not all experiments have the graph yet

The screenshot displays the ENCODE project interface. At the top, there is a navigation bar with 'ENCODE', 'Data', 'Methods', 'About', and 'Help' menus, along with a search bar. Below the navigation bar, a workflow graph is shown with three main components: 'ENCFF494GQK (reads)', 'fastq concatenation alignment', and 'ENCFF084TIX (transcriptome alignments)'. A 'sign file for' button is visible on the right side of the graph. Below the graph, there is a 'Download Graph' button. The main content area displays the following metadata for experiment ENCSR466KZY:

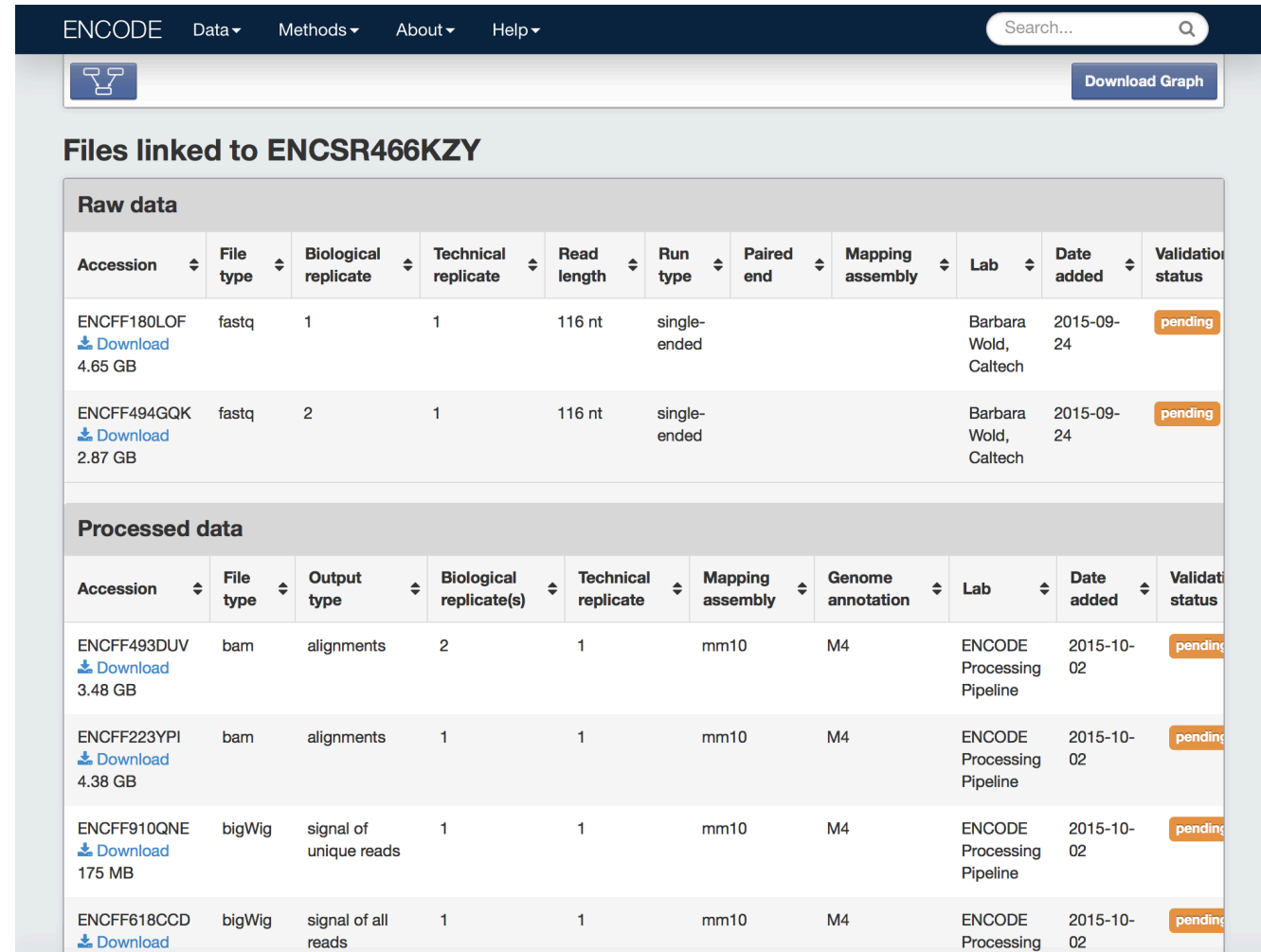
- Format:** bam
- Output:** transcriptome alignments
- Biological Replicate(s):** [2]
- Technical Replicate:** 1
- Mapping assembly:** mm10
- Genome annotation:** M4
- Lab:** ENCODE Processing Pipeline
- Date added:** 2015-10-02
- Software:** concat-fastqs 1.0.2, star 2.4.0k, samtools 0.1.19-96b5f2294a
- File download:** [Download](#)

At the bottom of the screenshot, the text 'Files linked to ENCSR466KZY' is visible.

Metadata to Data: Experiments

Use metadata to find data

- Filter: RNAseq, ENCODE, mm10 assembly
- Select an experiment:
<https://www.encodeproject.org/experiments/ENCSR466KZY/>
- Note metadata on protocols, replicates, etc.
- Graph: files are related by processing steps
- Download files from the graph or a list
 - **Not all experiments have the graph yet**



ENCODE Data Methods About Help Search... Download Graph

Files linked to ENCSR466KZY

Raw data

Accession	File type	Biological replicate	Technical replicate	Read length	Run type	Paired end	Mapping assembly	Lab	Date added	Validation status
ENCF180LOF Download 4.65 GB	fastq	1	1	116 nt	single-ended			Barbara Wold, Caltech	2015-09-24	pending
ENCF494GQK Download 2.87 GB	fastq	2	1	116 nt	single-ended			Barbara Wold, Caltech	2015-09-24	pending

Processed data

Accession	File type	Output type	Biological replicate(s)	Technical replicate	Mapping assembly	Genome annotation	Lab	Date added	Validation status
ENCF493DUV Download 3.48 GB	bam	alignments	2	1	mm10	M4	ENCODE Processing Pipeline	2015-10-02	pending
ENCF223YPI Download 4.38 GB	bam	alignments	1	1	mm10	M4	ENCODE Processing Pipeline	2015-10-02	pending
ENCF910QNE Download 175 MB	bigWig	signal of unique reads	1	1	mm10	M4	ENCODE Processing Pipeline	2015-10-02	pending
ENCF618CCD Download	bigWig	signal of all reads	1	1	mm10	M4	ENCODE Processing Pipeline	2015-10-02	pending

Metadata to Data: Roadmap Epigenomics

Finding data from different projects

- Filter: ChIP-seq, Roadmap, primary cell
- <https://www.encodeproject.org/experiments/ENCSR371FHD/>
- GEO page for Roadmap data:
<http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSM908967>

The screenshot shows the ENCODE Data Browser interface. The top navigation bar includes 'ENCODE', 'Data', 'Methods', 'About', and 'Help', along with a search bar. The left sidebar contains several filter categories:

- Assay:** ChIP-seq (558), RNA-seq (88), DNase-seq (86), MeDIP-seq (38), whole-genome shotgun bisulfite sequencing (38).
- Project:** Roadmap (558), ENCODE (295).
- Experiment status:** released (558).
- Organism:** *Homo sapiens* (558).
- Target of assay:** histone (471), histone modification (469), control (87).
- Biosample type:** tissue (682), primary cell (558), stem cell (322), in vitro differentiated cells (204), immortalized cell line (75).
- Organ:** extraembryonic structure (59).

The main content area displays 'Showing 25 of 558 results' with a 'View All' button. The results list several ChIP-seq experiments for trophoblast cells in *Homo sapiens*, embryonic stage, all performed by Bing Ren at UCSD as part of the Roadmap project. The first few results are:

- ChIP-seq of trophoblast cell (*Homo sapiens*, embryonic)**
Target: H4K91ac
Lab: Bing Ren, UCSD
Project: Roadmap
Experiment: ENCSR371FHD released
- ChIP-seq of trophoblast cell (*Homo sapiens*, embryonic)**
Target: H4K91ac
Lab: Bing Ren, UCSD
Project: Roadmap
Experiment: ENCSR916PNM released
- ChIP-seq of trophoblast cell (*Homo sapiens*, embryonic)**
Target: H4K8ac
Lab: Bing Ren, UCSD
Project: Roadmap
Experiment: ENCSR888MKB released
- ChIP-seq of trophoblast cell (*Homo sapiens*, embryonic)**
Target: H4K5ac
Lab: Bing Ren, UCSD
Project: Roadmap
Experiment: ENCSR182TPM released
- ChIP-seq of trophoblast cell (*Homo sapiens*, embryonic)**
Target: H3K9me3
Lab: Bing Ren, UCSD
Project: Roadmap
Experiment: ENCSR774MLM released
- ChIP-seq of trophoblast cell (*Homo sapiens*, embryonic)**
Target: H3K9me3
Lab: Bing Ren, UCSD
Project: Roadmap
Experiment: ENCSR346GRK released

Metadata to Data: Roadmap Epigenomics

Finding data from different projects

- Filter: ChIP-seq, Roadmap, primary cell
- <https://www.encodeproject.org/experiments/ENCSR371FHD/>
- GEO page for Roadmap data:
<http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSM908967>

The screenshot shows the ENCODE project website interface. At the top, there is a navigation bar with 'ENCODE' and menu items for 'Data', 'Methods', 'About', and 'Help'. A search bar is located on the right. Below the navigation bar, the breadcrumb trail reads 'EXPERIMENTS / CHIP-SEQ / HOMO SAPIENS / TROPHOBLAST CELL'. The main heading is 'Experiment summary for ENCSR371FHD', with a green status tag indicating 'Status: released'. The experiment details are listed in a table-like format:

Assay:	ChIP-seq
Replication type:	unreplicated
Biosample summary:	trophoblast cell (<i>Homo sapiens</i> , embryonic male)
Type:	primary cell
Target:	H4K91ac
Controls:	ENCSR890MJO
Lab:	Bing Ren, UCSD
Award PI:	Bing Ren, UCSD
Project:	Roadmap
External resources:	GEO:GSM908967
Aliases:	roadmap-epigenomics:Reference Epigenome: ChIP-Seq Input from BMP4 Mesendoderm Cells Thu Jan 12 15:50:22 -0600 2012 54330
Date released:	2015-10-06

Below the experiment details, the 'Assay details' section is partially visible, showing 'Nucleic acid type: DNA'.

Metadata to Data: Roadmap Epigenomics

Finding data from different projects

- Filter: ChIP-seq, Roadmap, primary cell
- <https://www.encodeproject.org/experiments/ENCSR371FHD/>
- GEO page for Roadmap processed data: <http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSM908967>

Submission date	Apr 03, 2012
Last update date	Jan 29, 2015
Contact name	UCSD AND SALK
Organization name	University of California, San Diego
Street address	Health Sciences Drive
City	La Jolla
State/province	CA
ZIP/Postal code	92092
Country	USA

Platform ID [GPL11154](#)
Series (1) [GSE16256](#) UCSD Human Reference Epigenome Mapping Project

Relations
SRA [SRX135238](#)
BioSample [SAMN00778288](#)

[See the data on Genome Data Viewer](#)

Supplementary file	Size	Download	File type/resource
SRX/SRX135/SRX135238		(ftp)	SRA Experiment
GSM908967_UCSD.H1_BMP4_Derived_Trophoblast_Cultured_Cells.H4K91ac.AY206.bed.gz	238.6 Mb	(ftp)(http)	BED
GSM908967_UCSD.H1_BMP4_Derived_Trophoblast_Cultured_Cells.H4K91ac.AY206.wig.gz	27.6 Mb	(ftp)(http)	WIG

Raw data provided as supplementary file
Processed data provided as supplementary file

[NLM](#) | [NIH](#) | [GEO Help](#) | [Disclaimer](#) | [Accessibility](#)

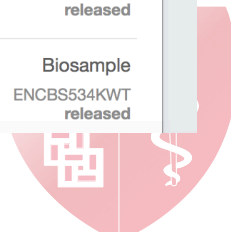


Faceted Browsing: Biosamples

Faceted Browsing of Biosamples

- Click on Data ... Biosamples
- <https://www.encodeproject.org/search/?type=Biosample>
- Filter: *Homo sapiens*, Tissue, Brain, ENCODE
- https://www.encodeproject.org/search/?type=Biosample&organism.scientific_name=Homo
[omo](https://www.encodeproject.org/search/?type=Biosample&organism.scientific_name=Homo)
[+sapiens&biosample_type=tissue&organ_slim](https://www.encodeproject.org/search/?type=Biosample&organism.scientific_name=Homo)
[s=brain&award.project=ENCODE](https://www.encodeproject.org/search/?type=Biosample&organism.scientific_name=Homo)

The screenshot displays the ENCODE project search interface. The top navigation bar includes 'ENCODE', 'Data', 'Methods', 'About', and 'Help', along with a search bar. The main content area is divided into two columns. The left column shows faceted search results for various categories: Organism (Mus musculus: 164, Homo sapiens: 39), Biosample status (released: 39), Biosample type (tissue: 39, primary cell: 38), Organ (brain: 39, heart: 12, kidney: 12, lung: 12, liver: 10), Sex (male: 27, female: 11, unknown: 1), Life stage (adult: 22, fetal: 16, child: 1), and Source (BioChain: 13, Thomas Montine: 10, NICHD: 6). The right column shows a list of 25 results out of 39, with a 'View All' button. The results are displayed in a table format with columns for sample name, type, and source. The first few results are: 'middle frontal gyrus (Homo sapiens, adult 78 year)' (Type: tissue, Source: Thomas Montine), 'middle frontal gyrus (Homo sapiens, adult 78 year)' (Type: tissue, Source: Thomas Montine), 'midbrain (Homo sapiens, adult 84 year)' (Type: tissue, Source: Thomas Montine), 'midbrain (Homo sapiens, adult 78 year)' (Type: tissue, Source: Thomas Montine), 'medulla oblongata (Homo sapiens, adult 78 year)' (Type: tissue, Source: Thomas Montine), 'medulla oblongata (Homo sapiens, adult 84 year)' (Type: tissue, Source: Thomas Montine), and 'brain (Homo sapiens, adult 66 year)' (Type: tissue, Source: Agilent).



Object Pages: Biosamples

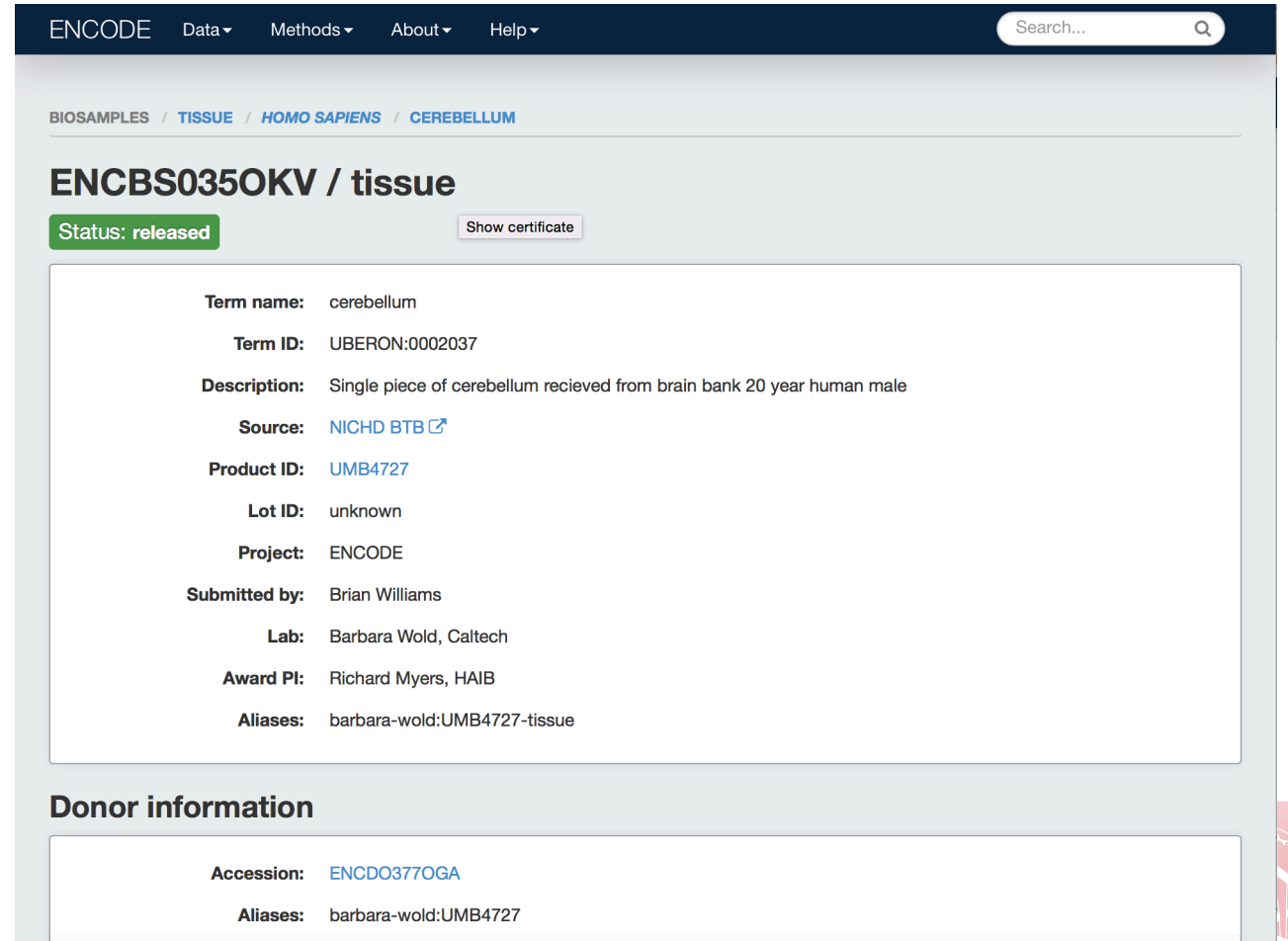
Biosample metadata

- Click “View all”
- Select

<https://www.encodeproject.org/biosamples/ENCBS035OKV/>

- Note donor metadata
- Click through to donor

<https://www.encodeproject.org/human-donors/ENCDO377OGA/>



The screenshot shows the ENCODE website interface for a biosample. The top navigation bar includes 'ENCODE', 'Data', 'Methods', 'About', and 'Help', along with a search bar. The breadcrumb trail is 'BIOSAMPLES / TISSUE / HOMO SAPIENS / CEREBELLUM'. The main title is 'ENCBS035OKV / tissue'. Below the title, there is a green 'Status: released' badge and a 'Show certificate' button. The main content area is a table with the following details:

Term name:	cerebellum
Term ID:	UBERON:0002037
Description:	Single piece of cerebellum recieved from brain bank 20 year human male
Source:	NICHD BTB
Product ID:	UMB4727
Lot ID:	unknown
Project:	ENCODE
Submitted by:	Brian Williams
Lab:	Barbara Wold, Caltech
Award PI:	Richard Myers, HAIB
Aliases:	barbara-wold:UMB4727-tissue

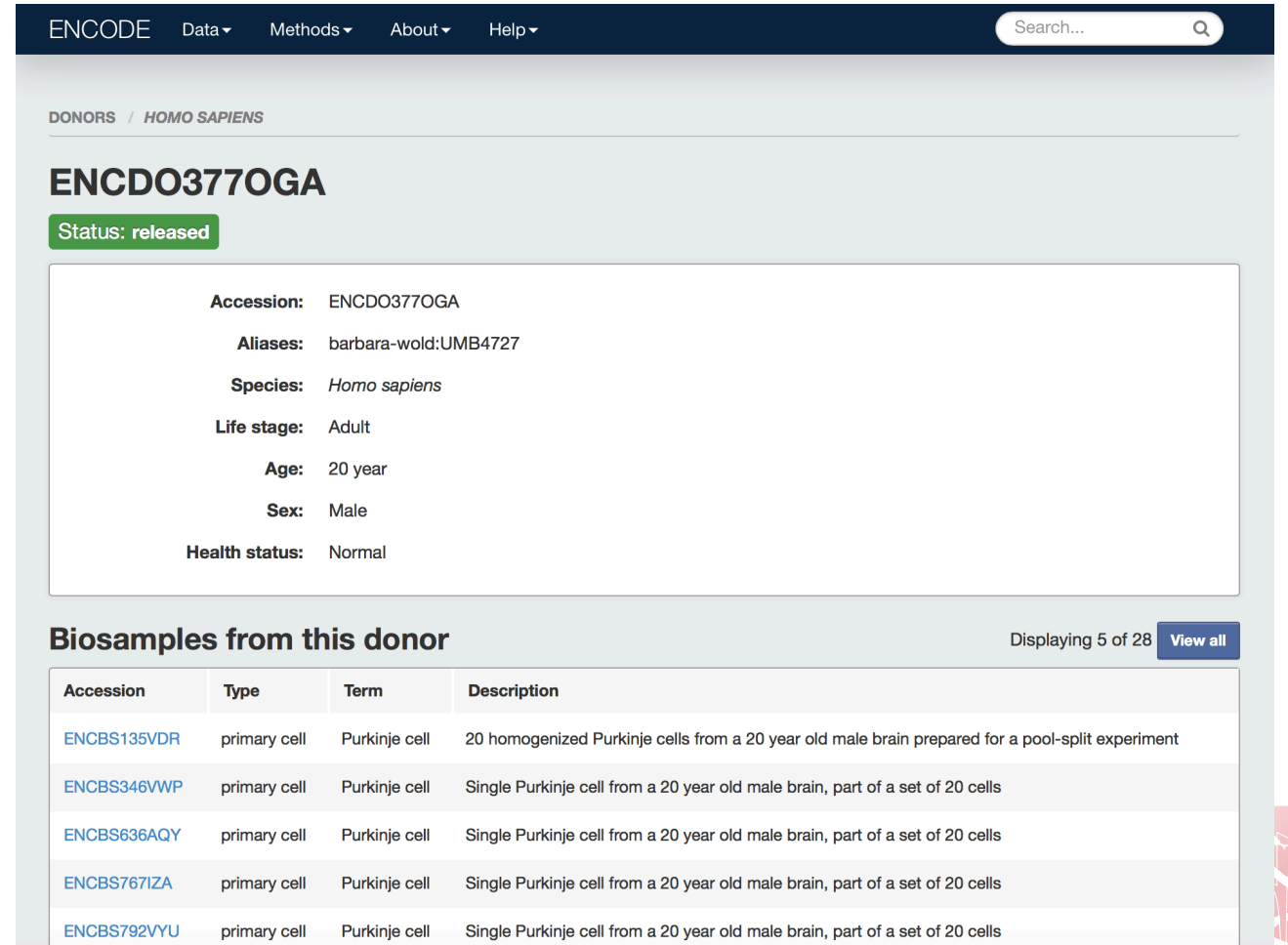
Below the table is a section titled 'Donor information' with the following details:

Accession:	ENCDO377OGA
Aliases:	barbara-wold:UMB4727

Object Pages: Biosamples

Biosample metadata

- Click “View all”
- Select <https://www.encodeproject.org/biosamples/ENCBS035OKV/>
- Note donor metadata
- Click through to donor <https://www.encodeproject.org/human-donors/ENCDO377OGA/>



ENCODE Data Methods About Help Search...

DONORS / HOMO SAPIENS

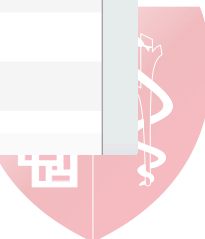
ENCDO377OGA

Status: released

Accession: ENCDO377OGA
Aliases: barbara-wold:UMB4727
Species: *Homo sapiens*
Life stage: Adult
Age: 20 year
Sex: Male
Health status: Normal

Biosamples from this donor Displaying 5 of 28 [View all](#)

Accession	Type	Term	Description
ENCBS135VDR	primary cell	Purkinje cell	20 homogenized Purkinje cells from a 20 year old male brain prepared for a pool-split experiment
ENCBS346VWP	primary cell	Purkinje cell	Single Purkinje cell from a 20 year old male brain, part of a set of 20 cells
ENCBS636AQY	primary cell	Purkinje cell	Single Purkinje cell from a 20 year old male brain, part of a set of 20 cells
ENCBS767IZA	primary cell	Purkinje cell	Single Purkinje cell from a 20 year old male brain, part of a set of 20 cells
ENCBS792VYU	primary cell	Purkinje cell	Single Purkinje cell from a 20 year old male brain, part of a set of 20 cells



Browsing Antibodies

Faceted Browsing of Antibodies

- Click on Data ... Antibodies
- <https://www.encodeproject.org/search/?type=AntibodyLot>
- Characterization standards
- <https://www.encodeproject.org/about/experiment-guidelines/>
- Example: filtered on “Eligible for new data”:
<https://www.encodeproject.org/antibodies/ENCAB934MDN/>

ENCODE Data Methods About Help Search...

Showing 25 of 3088 results [View All](#)

Eligibility status

- awaiting lab characterization 1679
- not pursued 1024
- not eligible for new data 260
- eligible for new data 121
- eligible for new data (via exemption) 3

+ See more...

Target Organism

- Homo sapiens 2704
- Drosophila melanogaster 207
- Caenorhabditis elegans 132
- Mus musculus 118
- Aequorea victoria 4

+ See more...

Target of antibody

- transcription factor 1831
- RNA binding protein 946
- histone 295
- histone modification 284
- control 13

+ See more...

Characterization method

- immunoprecipitation 1574
- immunoblot 620
- knockdown or knockout 213
- ChIP-seq comparison 153
- immunoprecipitation followed by mass spectrometry 109

+ See more...

SF3B1 (*Homo sapiens*) ● Antibody
Source: Abcam
Product ID / Lot ID: ab66774 / GR216813-1

H3K9me3 (*Homo sapiens*) ● Antibody
Source: Abcam
Product ID / Lot ID: ab8898 / GR223198-1

H3K27ac (*Homo sapiens*) ● Antibody
Source: Abcam
Product ID / Lot ID: ab4729 / GR211892-1

H3K4me2 (*Homo sapiens*) ● Antibody
Source: Abcam
Product ID / Lot ID: ab7766 / GR160184-1

H3K4me1 (*Homo sapiens*) ● Antibody
Source: Abcam
Product ID / Lot ID: ab8895 / GR193882-1

NR3C1 (*Homo sapiens*) ● Antibody
Source: Santa Cruz Biotech
Product ID / Lot ID: sc-1003 / B1915

DDX41 (*Homo sapiens*) ● Antibody
Source: GeneTex
Product ID / Lot ID: GTX111146 / 40044

Browsing Antibodies

Faceted Browsing of Antibodies

- Click on Data ... Antibodies
- <https://www.encodeproject.org/search/?type=AntibodyLot>
- Characterization standards
- <https://www.encodeproject.org/about/experiment-guidelines/>
- Example, filtered on “Eligible for new data”:
<https://www.encodeproject.org/antibodies/ENCAB934MDN/>

The screenshot displays the ENCODE project website interface. At the top, there is a navigation bar with 'ENCODE' and dropdown menus for 'Data', 'Methods', 'About', and 'Help'. A search bar is located on the right. Below the navigation bar, the breadcrumb trail reads 'ANTIBODIES / HOMO SAPIENS / IGF2BP3'. The main heading is 'ENCAB934MDN' with the subtitle 'Antibody against *Homo sapiens* IGF2BP3'. A status bar indicates 'Homo sapiens K562' and 'Eligible for new data' with a green dot. The main content area lists antibody details: Source (vendor): MBLI, Product ID: RN009P, Lot ID: 002, Targets: IGF2BP3 (*Homo sapiens*), Host: Rabbit, Clonality: Polyclonal, and Isotype: IgA. The Antigen description is 'KLH-conjugated synthetic peptide HQQKALQSGPPQSRRK (563-579 aa)'. At the bottom, two panels show experimental data for IGF2BP3 (*Homo sapiens*). The left panel, titled 'Method: immunoprecipitation', shows a Western blot for HepG2 cells with lanes 1, 2, 3, 4, and M. Molecular weight markers are indicated at 225 and 115 kDa. The right panel, titled 'Method: knockdown or knockout', shows a Western blot for HepG2 cells with lanes M, Cont. KD1 KD2, and Cont. KD1 KD2. Molecular weight markers are indicated at 150 and 100 kDa.

An exercise

Search for CTCF

1. Find the number of Experiments targeting CTCF
2. How many of them are performed using stem cells?
3. Filter: ChIP-seq, immortalized cell line, postnatal life stage
4. Only one of these experiments has treatment information available- which one?
5. What is the accession for the donor of the biosample for Replicate 1 of the human experiment?



An exercise

Search for CTCF...

1. Find the number of Experiments targeting CTCF

- ANSWER: 164

ENCODE Data Methods About Help

Data Type

Dataset	175
Experiment	164
Publication	98
AntibodyLot	14
FileSet	11

[+ See more...](#)

Showing 25 of 300 results [View All](#)

UCSC browser composite file set: Data from UCSC Genome Browser composite hg19-wgEncodeUwTfbs FileSet
ENCSR182KPN released
Lab: John Stamatoyannopoulos, UW
Project: ENCODE

UCSC browser composite file set: Data from UCSC Genome Browser composite mm9-wgEncodeLicrTfbs FileSet
ENCSR655MAT released
Lab: Bing Ren, UCSD
Project: ENCODE

UCSC browser composite file set: Data from UCSC Genome Browser composite hg19-wgEncodeOpenChromChip FileSet
ENCSR343YJH released
Lab: Gregory Crawford, Duke
Project: ENCODE

CTCF (*Mus musculus*) Target
External resources:
[UniProtKB:Q61164](#)

eGFP-CTCF (*Drosophila melanogaster*) Target
External resources:
[FlyBase:FBgn0035769](#) [FlyBase:FBtr0076878](#) [RefSeq:NM_139852](#)
[UniProtKB:Q9VS55](#) [FlyBase:CG8591-RA](#) [FlyBase:CTCF-RA](#) [FlyBase:CTCF](#)
[FlyBase:CG8591](#)

CTCF (*Drosophila melanogaster*) Target
External resources:

An exercise

Click through to the Experiments search page...

2. How many of them are performed using stem cells?

- ANSWER: 5

The screenshot shows the ENCODE Experiments search page. The top navigation bar includes 'ENCODE', 'Data', 'Methods', 'About', and 'Help', along with a search bar containing 'CTCF'. The left sidebar contains several filter categories with horizontal bar charts and counts:

- Assay:** ChIP-seq (162), ChIA-PET (2)
- Project:** ENCODE (164)
- Experiment status:** released (162), revoked (2)
- Genome assembly (visualization):** hg19 (128), mm9 (27)
- Organism:** *Homo sapiens* (135), *Mus musculus* (28)
- Target of assay:** transcription factor (164)
- Biosample type:** immortalized cell line (95), primary cell (43), tissue (18), stem cell (5), in vitro differentiated cells (2)
- Organ:** skin of body (10), brain (7), blood vessel (6), lung (6)

The main content area shows 'Showing 25 of 164 results' with buttons for 'Visualize', 'Download', and 'View All'. The first five results are listed:

- ChIA-PET of K562 (*Homo sapiens*, adult 53 year)** Experiment ENCSR000CAC released. Target: CTCF, Lab: Yijun Ruan, GIS, Project: ENCODE.
- ChIP-seq of bone marrow (*Mus musculus*, adult 8 week)** Experiment ENCSR000CBL released. Target: CTCF, Lab: Bing Ren, UCSD, Project: ENCODE.
- ChIP-seq of testis (*Mus musculus*, adult 8 week)** Experiment ENCSR000CEF released. Target: CTCF, Lab: Bing Ren, UCSD, Project: ENCODE.
- ChIP-seq of small intestine (*Mus musculus*, adult 8 week)** Experiment ENCSR000CED released. Target: CTCF, Lab: Bing Ren, UCSD, Project: ENCODE.
- ChIP-seq of spleen (*Mus musculus*, adult 8 week)** Experiment ENCSR000CBY released. Target: CTCF, Lab: Bing Ren, UCSD, Project: ENCODE.

An exercise

Filter as shown...

- Filter: ChIP-seq, immortalized cell line, postnatal life stage
- Only one of these experiments has treatment information available- which one?

- ANSWER: ENCSR000DIR

The screenshot shows the ENCODE Data Portal interface. The top navigation bar includes 'ENCODE', 'Data', 'Methods', 'About', and 'Help'. A search bar on the right contains 'CTCF'. The left sidebar displays various filters with counts:

- Assay:** ChIP-seq (3)
- Project:** ENCODE (3)
- Experiment status:** released (3)
- Genome assembly (visualization):** mm9 (2), hg19 (1)
- Organism:** *Mus musculus* (2), *Homo sapiens* (1)
- Target of assay:** transcription factor (3)
- Biosample type:** immortalized cell line (3), primary cell (2)
- Life stage:** adult (40), unknown (36), child (11), postnatal (3), embryonic (1)
- Biosample treatment:** 17 β -estradiol (1)

The main content area shows 'Showing 3 of 3 results' with 'Visualize' and 'Download' buttons. The results list three experiments:

- ChIP-seq of G1E-ER4 (*Mus musculus*, postnatal 0 day)** Experiment ENCSR000DIR released. Target: CTCF. Treatment: 10 nM 17 β -estradiol (CHEBI:16469) for 24 hour. Lab: Ross Hardison, PennState. Project: ENCODE.
- ChIP-seq of G1E (*Mus musculus*, postnatal 0 day)** Experiment ENCSR000DIS released. Target: CTCF. Lab: Ross Hardison, PennState. Project: ENCODE.
- ChIP-seq of HFF-Myc (*Homo sapiens*, postnatal)** Experiment ENCSR000DUM released. Target: CTCF. Lab: John Stamatoyannopoulos, UW. Project: ENCODE.

An exercise

Select experiment ENCSR000DUM...

5. What is the accession for the donor of the biosample for Replicate 1 of the human experiment?

The screenshot shows the ENCODE Data Portal interface. The top navigation bar includes 'ENCODE', 'Data', 'Methods', 'About', and 'Help', with a search bar containing 'CTCF'. The left sidebar displays filters for Assay (ChIP-seq), Project (ENCODE), Experiment status (released), Genome assembly (mm9, hg19), Organism (Mus musculus, Homo sapiens), Target of assay (transcription factor), Biosample type (immortalized cell line, primary cell), Life stage (adult, unknown, child, postnatal, embryonic), and Biosample treatment (17β-estradiol). The main content area shows 'Showing 3 of 3 results' for 'ChIP-seq of G1E-ER4 (Mus musculus, postnatal 0 day)', 'ChIP-seq of G1E (Mus musculus, postnatal 0 day)', and 'ChIP-seq of HFF-Myc (Homo sapiens, postnatal)'. The first result is highlighted with a red box, and a red arrow points to the 'HFF-Myc' result.

Assay	Count
ChIP-seq	3

Project	Count
ENCODE	3

Experiment status	Count
released	3

Genome assembly (visualization)	Count
mm9	2
hg19	1

Organism	Count
Mus musculus	2
Homo sapiens	1

Target of assay	Count
transcription factor	3

Biosample type	Count
immortalized cell line	3
primary cell	2

Life stage	Count
adult	40
unknown	36
child	11
postnatal	3
embryonic	1

Biosample treatment	Count
17β-estradiol	1

Showing 3 of 3 results	Experiment
ChIP-seq of G1E-ER4 (<i>Mus musculus</i>, postnatal 0 day) Target: CTCF Treatment: 10 nM 17β-estradiol (CHEBI:16469) for 24 hour Lab: Ross Hardison, PennState Project: ENCODE	ENCSR000DIR released
ChIP-seq of G1E (<i>Mus musculus</i>, postnatal 0 day) Target: CTCF Lab: Ross Hardison, PennState Project: ENCODE	ENCSR000DIS released
ChIP-seq of HFF-Myc (<i>Homo sapiens</i>, postnatal) Target: CTCF Lab: John Stamatoyannopoulos, UW Project: ENCODE	ENCSR000DUM released

An exercise

Select experiment ENCSR000DUM...

Within ENCSR000DUM, select the accession for the biosample under Biological replicate 1

5. What is the accession for the donor of the biosample for Replicate 1 of the human experiment?

ENCODE Data Methods About Help Search...

PDF

[wgEncodeUwTfbs.release3.html.pdf](#)

More

Biological replicate - 1

Technical replicate: 1

Library: ENCLB340XFO

Biosample: **ENCBS411AMC** - (*Homo sapiens*, postnatal male)

Biological replicate - 2

Technical replicate: 1

Library: ENCLB203MUP

Biosample: ENCBS553ENC - (*Homo sapiens*, postnatal male)

Files linked to ENCSR000DUM

Visualize Data

Raw data

Accession	File type	Biological replicate	Technical replicate	Read length	Run type	Paired end	Mapping assembly	Lab	Date added
ENCF001HOQ	fastq	1	1	36 nt	single-			John	2011-06-

An exercise

Select experiment ENCSR000DUM...

Within ENCSR000DUM, select the accession for the biosample under Biological replicate 1

Scroll down to the Donor Information section of the biosample object page

5. What is the accession for the donor of the biosample for Replicate 1 of the human experiment?

- ANSWER: ENCDO232AAA

The screenshot shows the ENCODE website interface. At the top, there is a navigation bar with 'ENCODE' and dropdown menus for 'Data', 'Methods', 'About', and 'Help'. A search bar is located on the right. Below the navigation bar, the breadcrumb trail reads 'BIOSAMPLES / IMMORTALIZED CELL LINE / HOMO SAPIENS / HFF-MYC'. The main heading is 'ENCBS411AMC / immortalized cell line'. A green badge indicates 'Status: released'. The main content area lists the following details:

- Term name:** HFF-Myc
- Term ID:** EFO:0005710
- Description:** Foreskin fibroblast cells expressing canine cMyc
- Source:** [Beverly Torok-Storb](#)
- Project:** ENCODE
- Submitted by:** John Stamatoyannopoulos
- Lab:** John Stamatoyannopoulos, UW
- Award PI:** John Stamatoyannopoulos, UW
- Aliases:** john-stamatoyannopoulos:15073, john-stamatoyannopoulos:DS15073-biosample
- External resources:** [UCSC-ENCODE-cv:HFF-Myc](#)

Below this list, there is a section titled 'Derived from biosample' with the accession [ENCBS309AAA](#). At the bottom, the 'Donor information' section is visible, with the 'Accession' field containing [ENCDO232AAA](#), which is highlighted with a red box. The 'Species' field is partially visible as 'Homo sapiens'.

The ENCODE Portal: Recap

1. Find tutorials, contact details, and other information on how to use the ENCODE portal
2. Use faceted browsing to select and narrow data to items of interest, including searches for data from partner projects
3. Search for data of interest using keyword searches and identifiers
4. Navigate the metadata of a single object and retrieve ENCODE data one file at a time
5. Hands-on, independent practice at searching and navigating the portal

<https://www.encodeproject.org/help/getting-started/>

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

Any questions?

