



**STANFORD**  
SCHOOL OF MEDICINE



# ENCODEPROJECT.ORG

# USING THE ENCODE PORTAL

---

ENCODE DCC

ASHG ENCODE Workshop

October, 2014



**STANFORD**  
SCHOOL OF MEDICINE



J. Seth Strattan, PhD ENCODE DCC

# encodeproject.org: Motivation

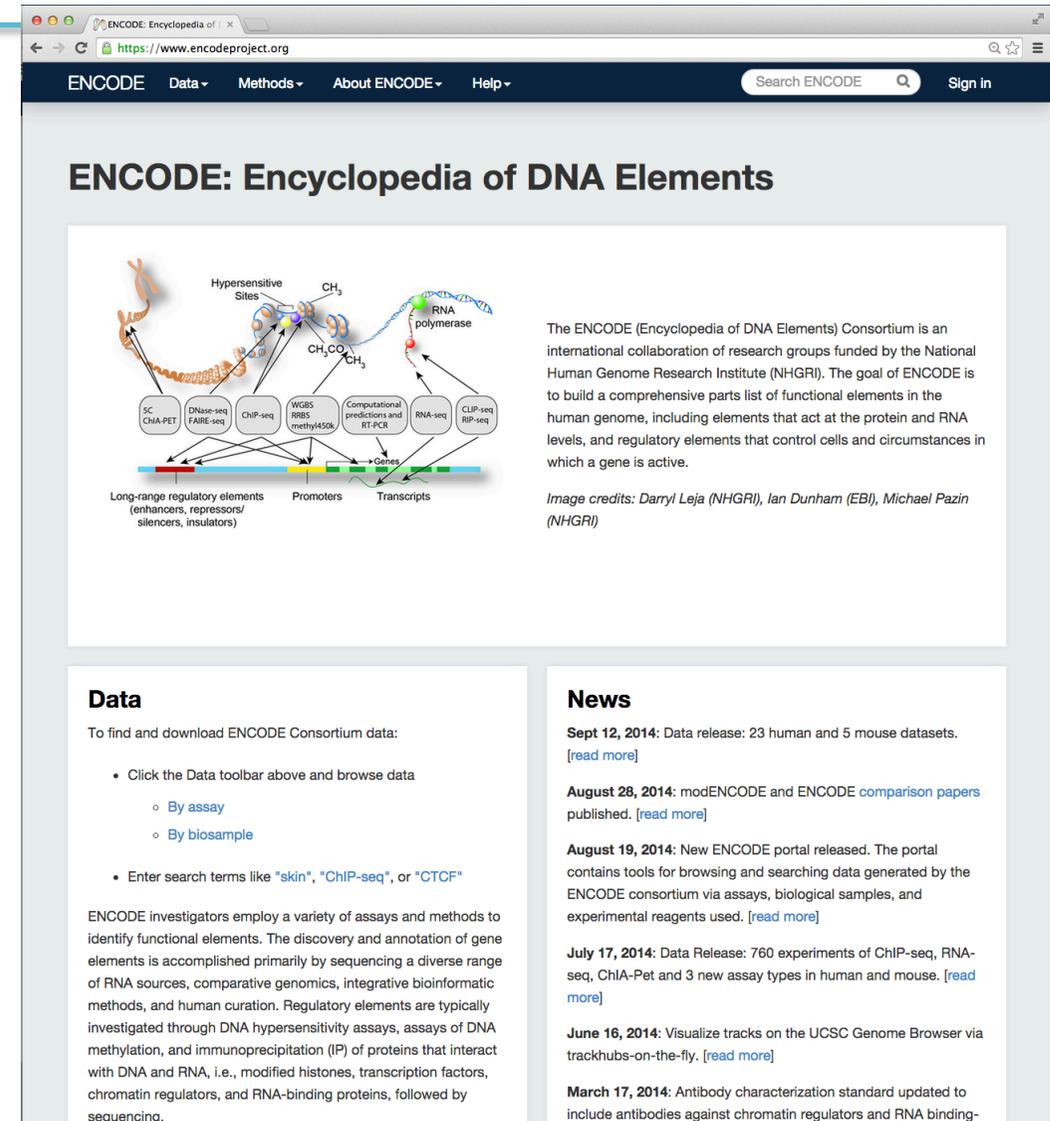
---

- ENCODE data can be difficult to find.
  - Different data are in different places.
  - Browser tracks, ftp sites, web sites.
- Metadata describing experiments can be hard to search.
- Where is the “official” list of everything ENCODE has done?
- Can I just download the ChIP data on ZNF143?



# encodeproject.org: Why did we build it?

- encodeproject.org is the ENCODE Data Coordination Center (DCC)'s new portal application.
  - Central source for all ENCODE metadata and data.
  - Updated continuously.
  - Easy to find the data you care about ...
  - ... and maybe discover data you didn't even know about.



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)

### Data

To find and download ENCODE Consortium data:

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

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

### News

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

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

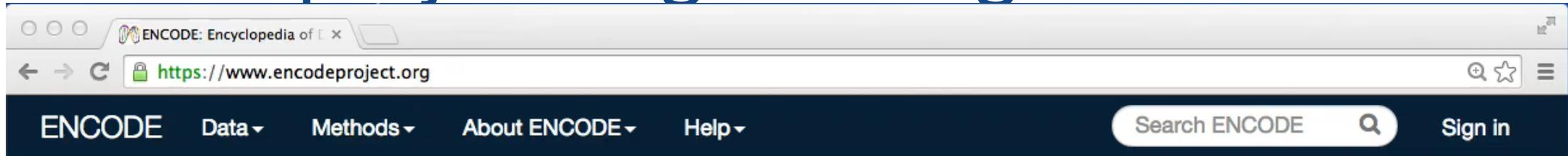
**August 19, 2014:** New ENCODE portal released. The portal contains tools for browsing and searching data generated by the ENCODE consortium via assays, biological samples, and experimental reagents used. [\[read more\]](#)

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

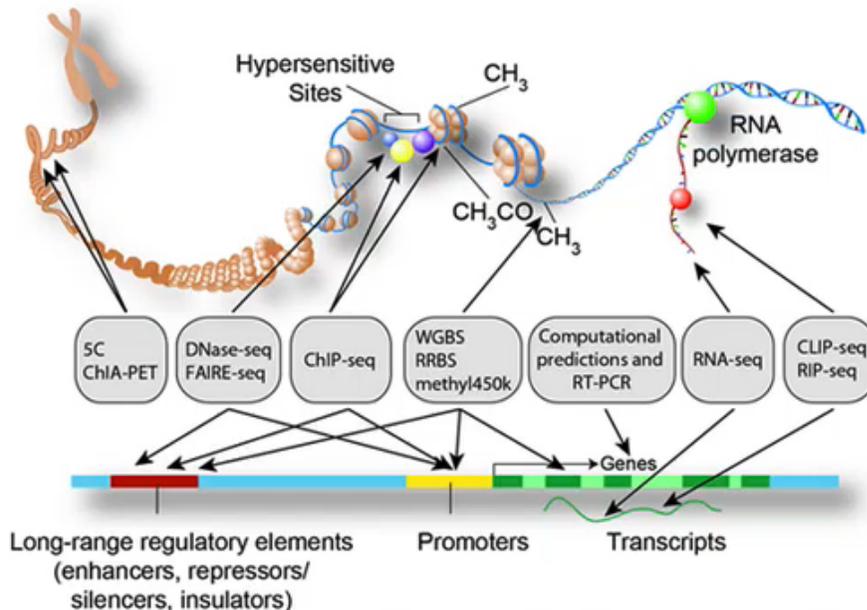
**June 16, 2014:** Visualize tracks on the UCSC Genome Browser via trackhubs-on-the-fly. [\[read more\]](#)

**March 17, 2014:** Antibody characterization standard updated to include antibodies against chromatin regulators and RNA binding-

# encodeproject.org: Getting Started



## ENCODE: Encyclopedia of DNA Elements

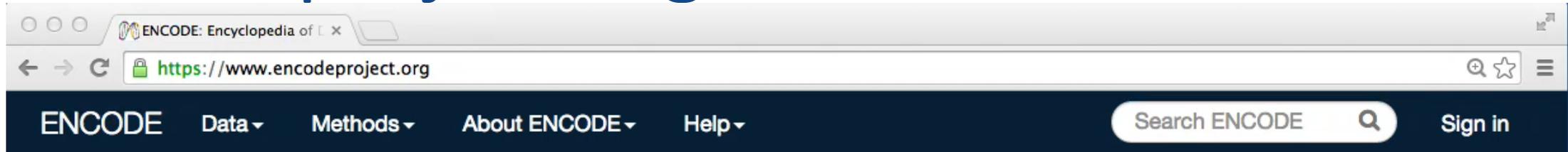


The ENCODE (Encyclopedia of DNA Elements) Consortium is an international collaboration of research groups funded by the National Human Genome Research Institute (NHGRI). The goal of ENCODE is to build a comprehensive parts list of functional elements in the human genome, including elements that act at the protein and RNA levels, and regulatory elements that control cells and circumstances in which a gene is active.

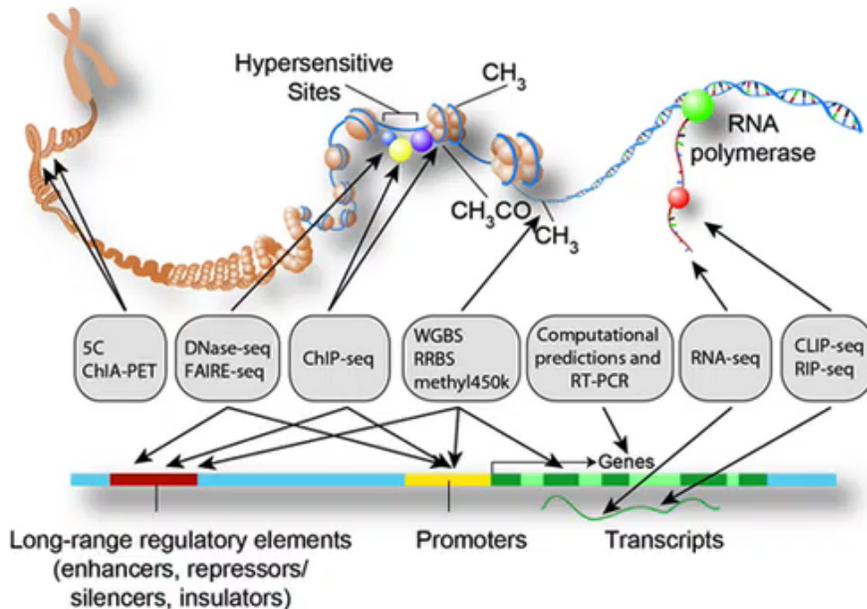
*Image credits: Darryl Leja (NHGRI), Ian Dunham (EBI), Michael Pazin (NHGRI)*



# encodeproject.org: The Menu Bar



## ENCODE: Encyclopedia of DNA Elements

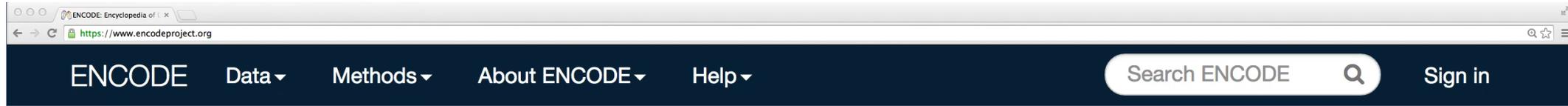


The ENCODE (Encyclopedia of DNA Elements) Consortium is an international collaboration of research groups funded by the National Human Genome Research Institute (NHGRI). The goal of ENCODE is to build a comprehensive parts list of functional elements in the human genome, including elements that act at the protein and RNA levels, and regulatory elements that control cells and circumstances in which a gene is active.

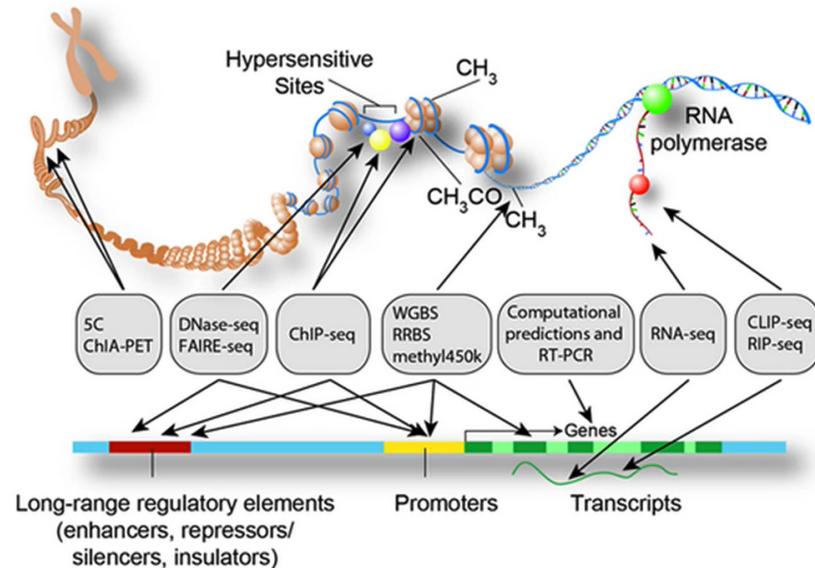
*Image credits: Darryl Leja (NHGRI), Ian Dunham (EBI), Michael Pazin (NHGRI)*



# encodeproject.org: Faceted Browsing



## 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.

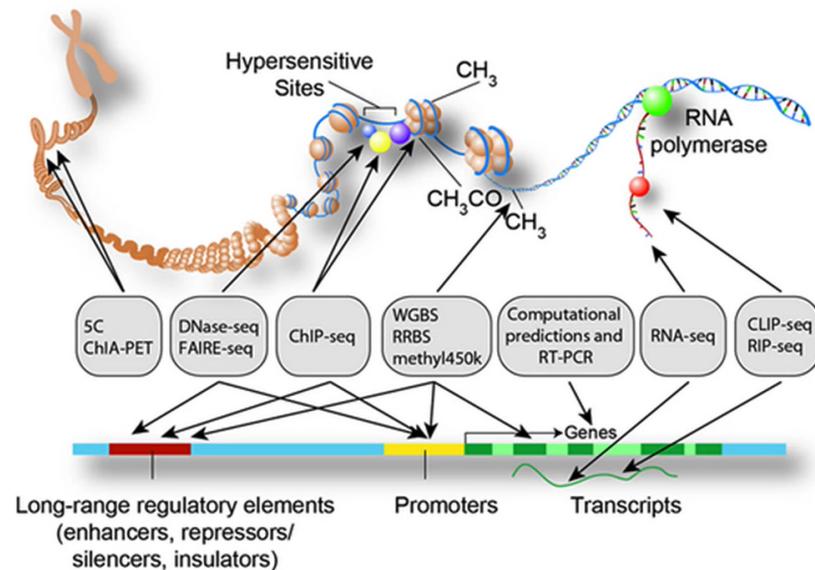
*Image credits: Darryl Leja (NHGRI), Ian Dunham (EBI), Michael Pazin (NHGRI)*



# encodeproject.org: The Search Box



## ENCODE: Encyclopedia of DNA Elements

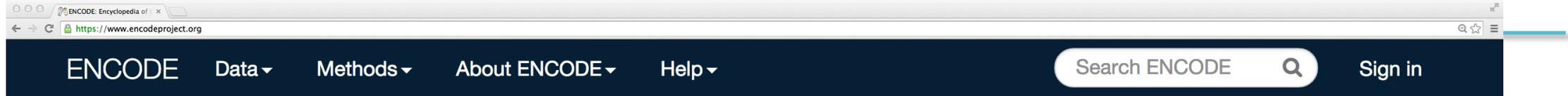


The ENCODE (Encyclopedia of DNA Elements) Consortium is an international collaboration of research groups funded by the National Human Genome Research Institute (NHGRI). The goal of ENCODE is to build a comprehensive parts list of functional elements in the human genome, including elements that act at the protein and RNA levels, and regulatory elements that control cells and circumstances in which a gene is active.

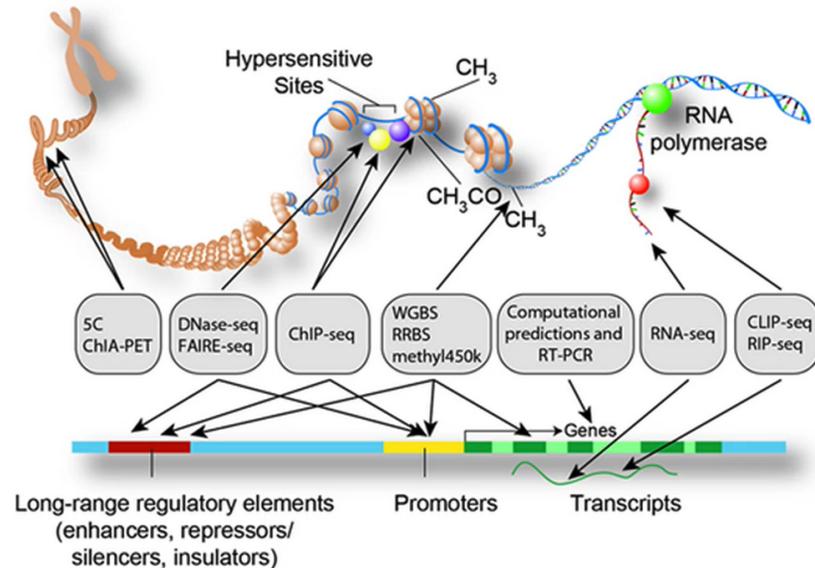
*Image credits: Darryl Leja (NHGRI), Ian Dunham (EBI), Michael Pazin (NHGRI)*



# encodeproject.org: The Search Box



## ENCODE: Encyclopedia of DNA Elements

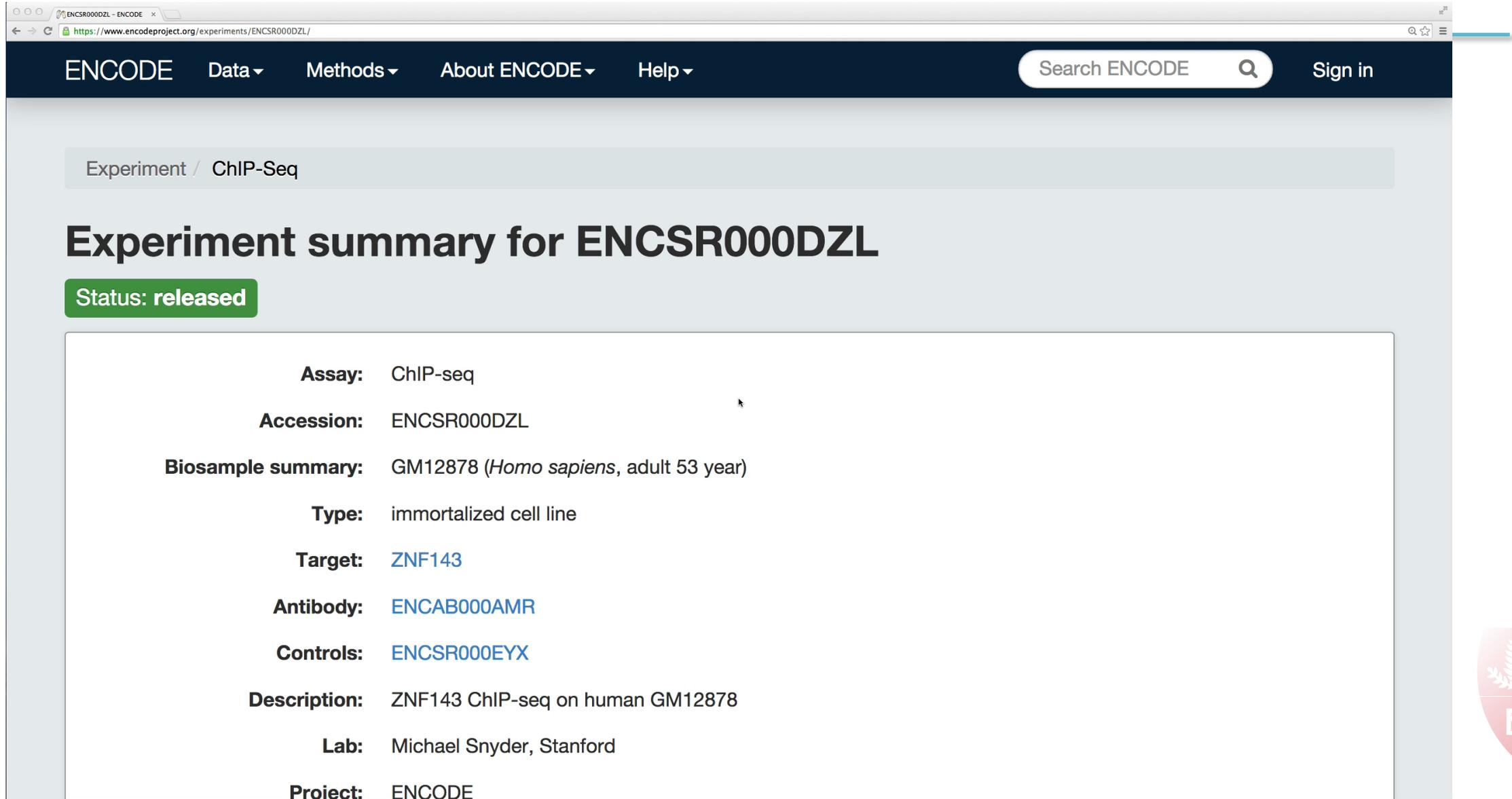


The ENCODE (Encyclopedia of DNA Elements) Consortium is an international collaboration of research groups funded by the National Human Genome Research Institute (NHGRI). The goal of ENCODE is to build a comprehensive parts list of functional elements in the human genome, including elements that act at 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)*



# encodeproject.org: Links to data

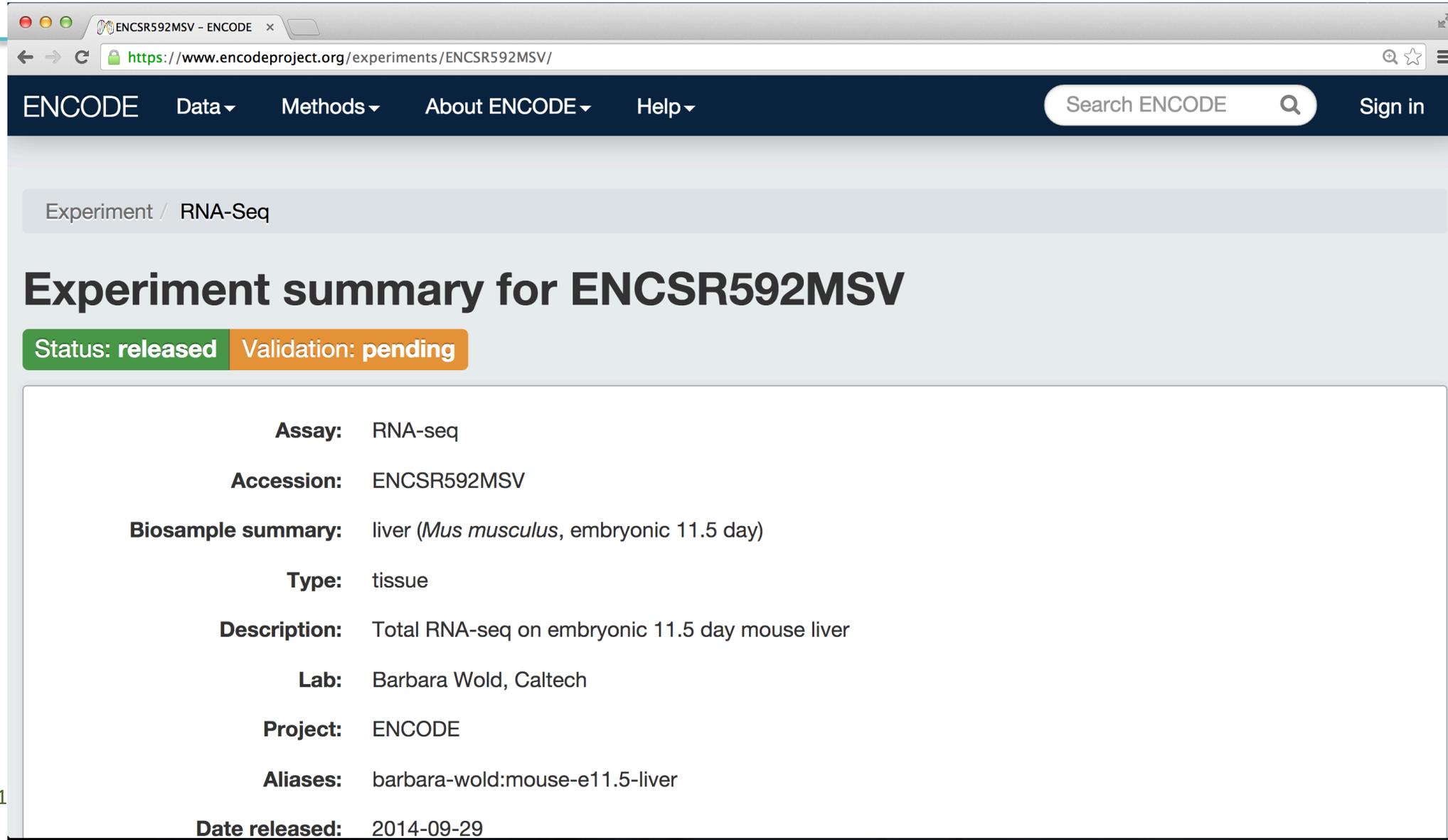


The screenshot shows a web browser window with the URL <https://www.encodeproject.org/experiments/ENCSR000DZL/>. The page features a dark blue navigation bar with the ENCODE logo and menu items: Data, Methods, About ENCODE, and Help. A search bar labeled "Search ENCODE" and a "Sign in" link are also present. Below the navigation bar, the breadcrumb "Experiment / ChIP-Seq" is displayed. The main heading is "Experiment summary for ENCSR000DZL", followed by a green badge indicating "Status: released". The experiment details are listed in a table format:

<b>Assay:</b>	ChIP-seq
<b>Accession:</b>	ENCSR000DZL
<b>Biosample summary:</b>	GM12878 ( <i>Homo sapiens</i> , adult 53 year)
<b>Type:</b>	immortalized cell line
<b>Target:</b>	<a href="#">ZNF143</a>
<b>Antibody:</b>	<a href="#">ENCAB000AMR</a>
<b>Controls:</b>	<a href="#">ENCSR000EYX</a>
<b>Description:</b>	ZNF143 ChIP-seq on human GM12878
<b>Lab:</b>	Michael Snyder, Stanford
<b>Project:</b>	ENCODE



# encodeproject.org: “Rapid Release”



The screenshot shows a web browser window with the URL <https://www.encodeproject.org/experiments/ENCSR592MSV/>. The page features a dark blue navigation bar with the ENCODE logo and menu items: Data, Methods, About ENCODE, and Help. A search bar labeled "Search ENCODE" and a "Sign in" link are also present. Below the navigation bar, a breadcrumb trail reads "Experiment / RNA-Seq". The main heading is "Experiment summary for ENCSR592MSV". Two status boxes are displayed: "Status: released" in green and "Validation: pending" in orange. The experiment details are listed in a table format:

<b>Assay:</b>	RNA-seq
<b>Accession:</b>	ENCSR592MSV
<b>Biosample summary:</b>	liver ( <i>Mus musculus</i> , embryonic 11.5 day)
<b>Type:</b>	tissue
<b>Description:</b>	Total RNA-seq on embryonic 11.5 day mouse liver
<b>Lab:</b>	Barbara Wold, Caltech
<b>Project:</b>	ENCODE
<b>Aliases:</b>	barbara-wold:mouse-e11.5-liver
<b>Date released:</b>	2014-09-29



# encodeproject.org: “Rapid Release”

The screenshot shows a web browser window with the URL <https://www.encodeproject.org/experiments/ENCSR592MSV/>. The page features a dark blue navigation bar with the ENCODE logo and menu items: Data, Methods, About ENCODE, and Help. A search bar labeled "Search ENCODE" and a "Sign in" button are also present. The main content area is titled "Assay details" and lists the following information:

- Nucleic acid type:** RNA
- Lysis method:** Ambion mirVana
- Extraction method:** Ambion mirVana
- Fragmentation method:** Illumina/Nextera tagmentation
- Size range:** >200
- Size selection method:** SPRI beads
- Platform:** [HiSeq 2000](#)

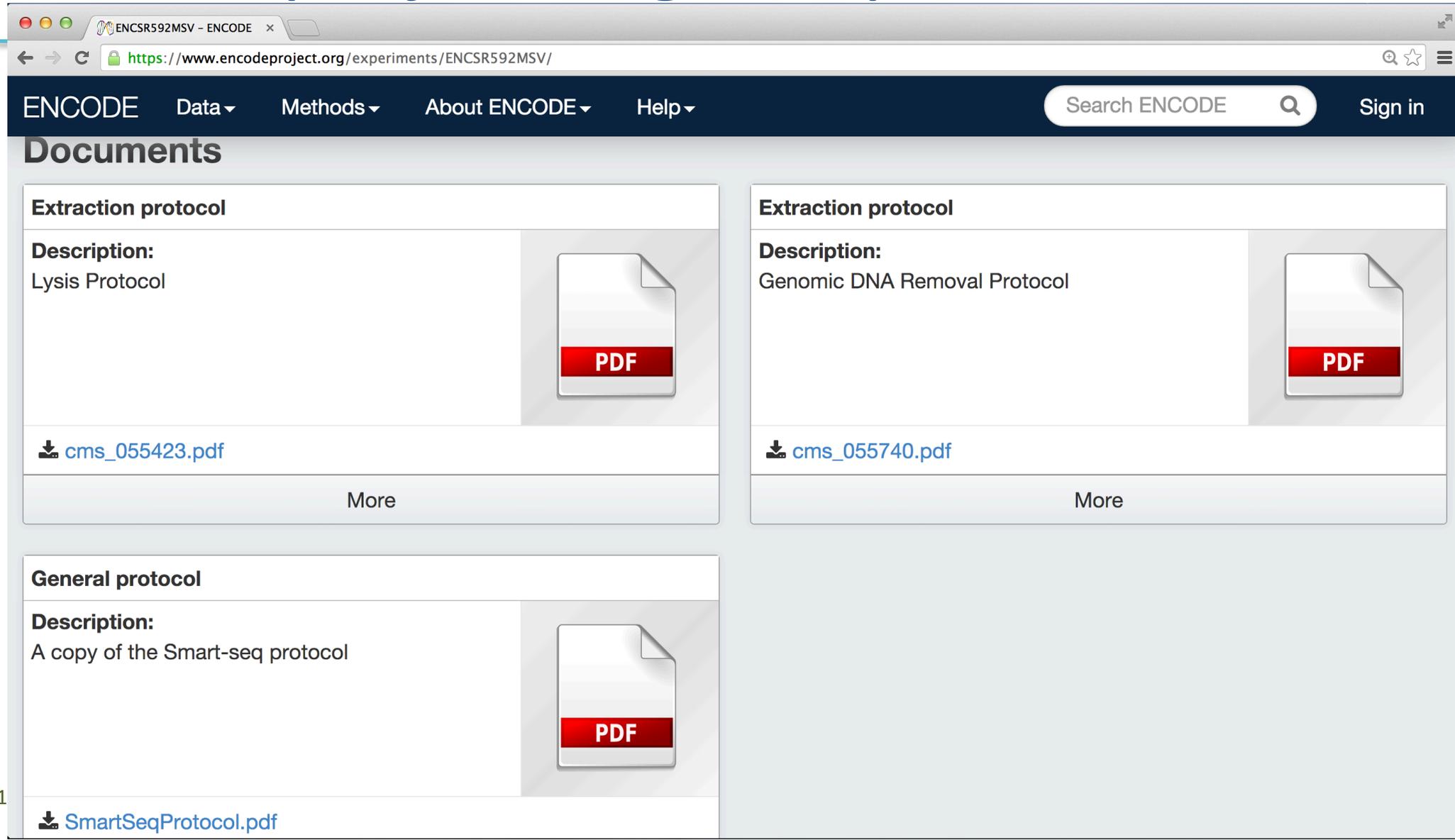
Below the assay details is a "Documents" section with two entries:

- Extraction protocol**  
**Description:** Lysis Protocol
- Extraction protocol**  
**Description:** Genomic DNA Removal Protocol

Each document entry includes a file icon and a red bar at the bottom, indicating a download or view action.



# encodeproject.org: “Rapid Release”



The screenshot shows a web browser window with the URL <https://www.encodeproject.org/experiments/ENCSR592MSV/>. The page features a dark blue navigation bar with the ENCODE logo and menu items: Data, Methods, About ENCODE, and Help. A search bar labeled "Search ENCODE" and a "Sign in" link are also present. The main content area is titled "Documents" and displays three document cards. Each card includes a title, a description, a PDF icon, and a download link. The first card is for an "Extraction protocol" (Lysis Protocol) with file [cms\\_055423.pdf](#). The second card is for an "Extraction protocol" (Genomic DNA Removal Protocol) with file [cms\\_055740.pdf](#). The third card is for a "General protocol" (Smart-seq protocol) with file [SmartSeqProtocol.pdf](#). Each card also has a "More" button at the bottom.

ENCODE Data Methods About ENCODE Help Search ENCODE Sign in

## Documents

**Extraction protocol**

**Description:**  
Lysis Protocol



[cms\\_055423.pdf](#)

More

**Extraction protocol**

**Description:**  
Genomic DNA Removal Protocol



[cms\\_055740.pdf](#)

More

**General protocol**

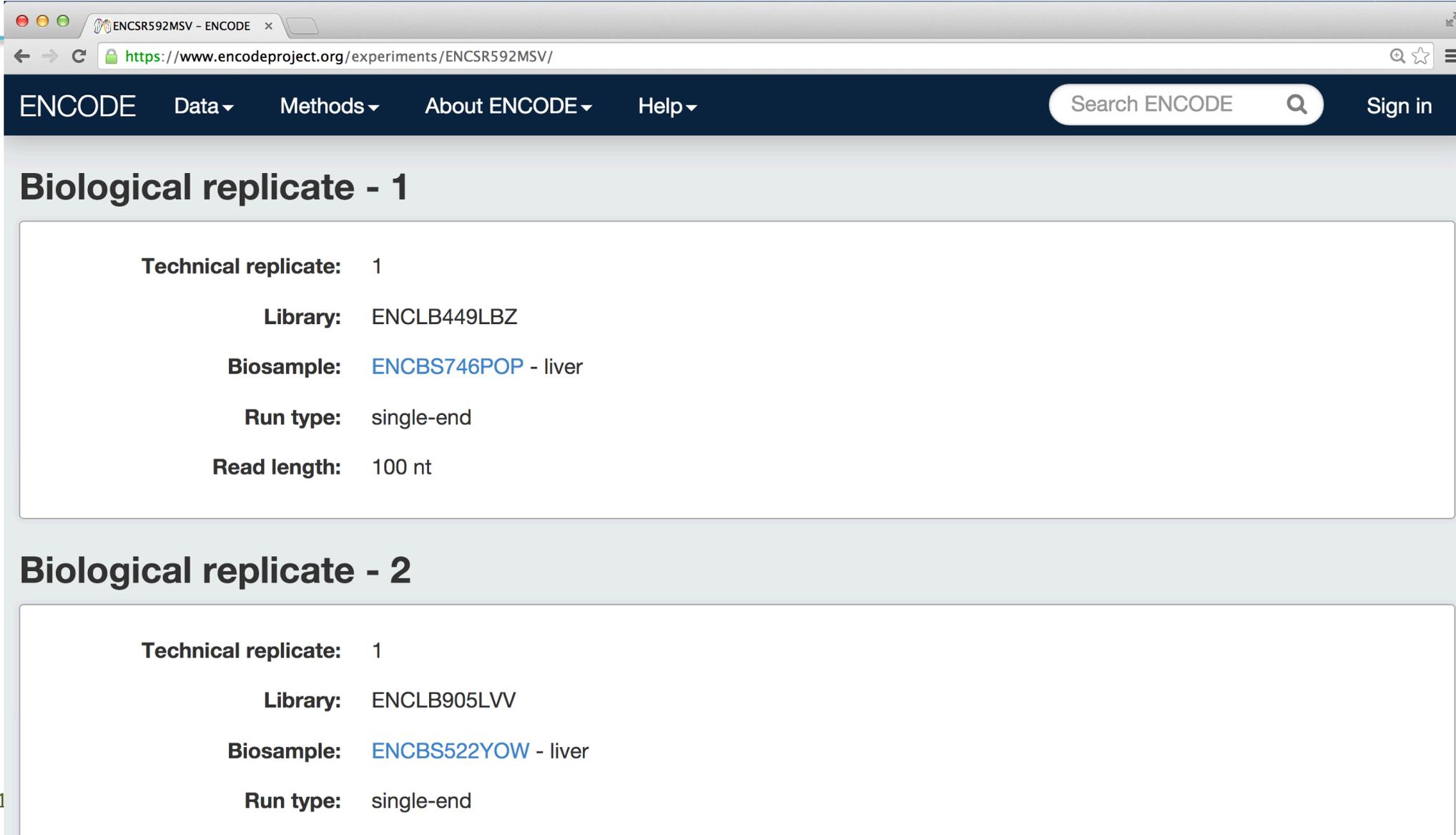
**Description:**  
A copy of the Smart-seq protocol



[SmartSeqProtocol.pdf](#)



# encodeproject.org: “Rapid Release”



The screenshot shows a web browser window with the URL <https://www.encodeproject.org/experiments/ENCSR592MSV/>. The page features a dark blue navigation bar with the ENCODE logo and menu items: Data, Methods, About ENCODE, and Help. A search bar labeled "Search ENCODE" and a "Sign in" link are also present. The main content area is divided into two sections, each titled "Biological replicate - 1" and "Biological replicate - 2". Each section contains a list of technical details for a specific experiment.

**Biological replicate - 1**

- Technical replicate: 1
- Library: ENCLB449LBZ
- Biosample: [ENCBS746POP](#) - liver
- Run type: single-end
- Read length: 100 nt

**Biological replicate - 2**

- Technical replicate: 1
- Library: ENCLB905LVV
- Biosample: [ENCBS522YOW](#) - liver
- Run type: single-end



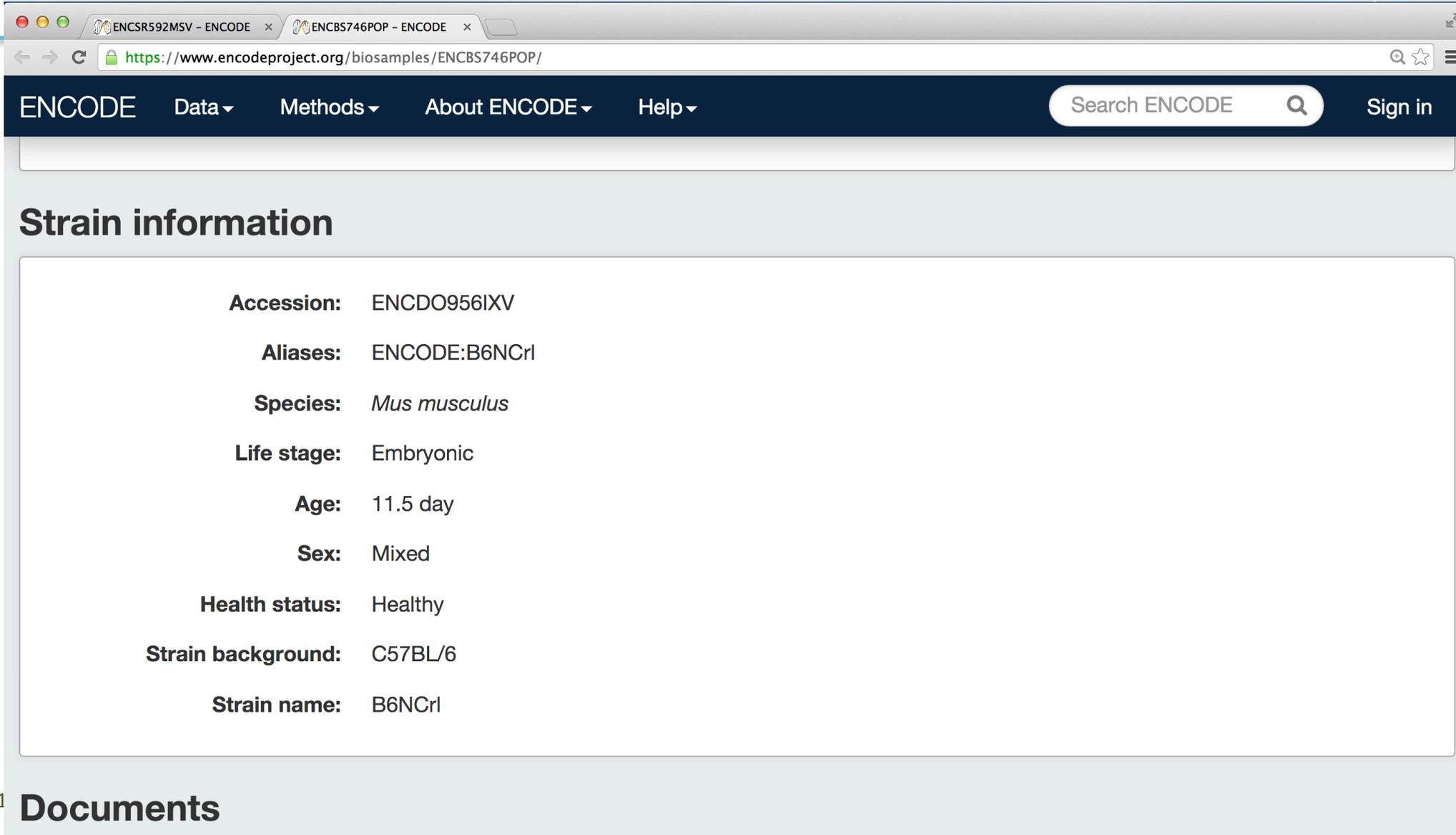
# encodeproject.org: “Rapid Release”

The screenshot shows a web browser window with two tabs: 'ENCSR592MSV - ENCODE' and 'ENCBS746POP - ENCODE'. The active tab is 'ENCBS746POP - ENCODE' with the URL 'https://www.encodeproject.org/biosamples/ENCBS746POP/'. The page header includes 'ENCODE' and navigation links for 'Data', 'Methods', 'About ENCODE', and 'Help'. A search bar labeled 'Search ENCODE' and a 'Sign in' link are also present. The main content area displays the title 'ENCBS746POP / tissue' and a green badge indicating 'Status: released'. Below this, a list of metadata is shown:

<b>Term name:</b>	liver
<b>Term ID:</b>	UBERON:0002107
<b>Description:</b>	Liver from 109 embryonic 11.5 day mice
<b>Source:</b>	<a href="#">Len Pennacchio</a>
<b>Project:</b>	ENCODE
<b>Submitted by:</b>	Axel Visel
<b>Lab:</b>	Len Pennacchio, JGI
<b>Grant:</b>	U54HG006997
<b>Note:</b>	109 individuals
<b>Date obtained:</b>	2013-09-16
<b>Starting amount:</b>	0.3401 g



# encodeproject.org: “Rapid Release”



The screenshot shows a web browser window with two tabs: 'ENCSR592MSV - ENCODE' and 'ENCBS746POP - ENCODE'. The address bar shows the URL 'https://www.encodeproject.org/biosamples/ENCBS746POP/'. The navigation bar includes 'ENCODE', 'Data', 'Methods', 'About ENCODE', and 'Help', along with a search box and a 'Sign in' link. The main content area is titled 'Strain information' and lists the following details:

<b>Accession:</b>	ENCDO956IXV
<b>Aliases:</b>	ENCODE:B6NCrI
<b>Species:</b>	<i>Mus musculus</i>
<b>Life stage:</b>	Embryonic
<b>Age:</b>	11.5 day
<b>Sex:</b>	Mixed
<b>Health status:</b>	Healthy
<b>Strain background:</b>	C57BL/6
<b>Strain name:</b>	B6NCrI

Below the strain information, the 'Documents' section is partially visible.



# encodeproject.org: “Rapid Release”

The screenshot shows a web browser with two tabs: 'ENCSR592MSV - ENCODE' and 'ENCBS746POP - ENCODE'. The address bar shows the URL 'https://www.encodeproject.org/biosamples/ENCBS746POP/'. The navigation bar includes 'ENCODE', 'Data', 'Methods', 'About ENCODE', and 'Help', along with a search bar and a 'Sign in' link.

## Documents

**Excision protocol**

**Description:**  
Specific excision protocol for embryonic and postnatal mouse biosamples



[tissue\\_excision\\_protocols\\_052014.pdf](#)

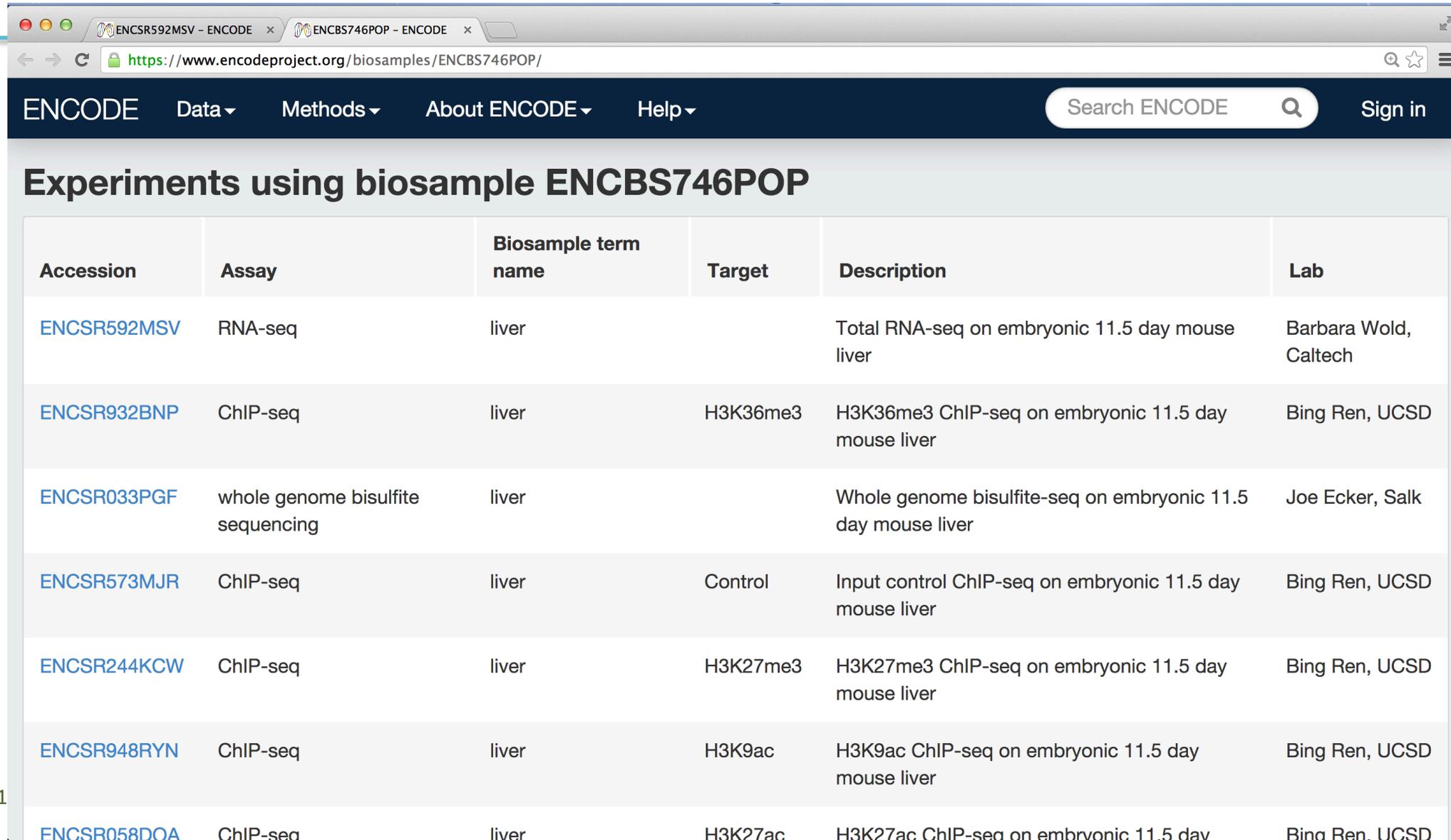
More

## Experiments using biosample ENCBS746POP

Accession	Assay	Biosample term name	Target	Description	Lab
<a href="#">ENCSR592MSV</a>	RNA-seq	liver		Total RNA-seq on embryonic 11.5 day mouse liver	Barbara Wold, Caltech



# encodeproject.org: “Rapid Release”



The screenshot shows a web browser window with two tabs: 'ENCSR592MSV - ENCODE' and 'ENCBS746POP - ENCODE'. The address bar shows the URL 'https://www.encodeproject.org/biosamples/ENCBS746POP/'. The page header includes the ENCODE logo, navigation menus for 'Data', 'Methods', 'About ENCODE', and 'Help', a search bar labeled 'Search ENCODE', and a 'Sign in' link. The main content area is titled 'Experiments using biosample ENCBS746POP' and contains a table with the following data:

Accession	Assay	Biosample term name	Target	Description	Lab
<a href="#">ENCSR592MSV</a>	RNA-seq	liver		Total RNA-seq on embryonic 11.5 day mouse liver	Barbara Wold, Caltech
<a href="#">ENCSR932BNP</a>	ChIP-seq	liver	H3K36me3	H3K36me3 ChIP-seq on embryonic 11.5 day mouse liver	Bing Ren, UCSD
<a href="#">ENCSR033PGF</a>	whole genome bisulfite sequencing	liver		Whole genome bisulfite-seq on embryonic 11.5 day mouse liver	Joe Ecker, Salk
<a href="#">ENCSR573MJR</a>	ChIP-seq	liver	Control	Input control ChIP-seq on embryonic 11.5 day mouse liver	Bing Ren, UCSD
<a href="#">ENCSR244KCW</a>	ChIP-seq	liver	H3K27me3	H3K27me3 ChIP-seq on embryonic 11.5 day mouse liver	Bing Ren, UCSD
<a href="#">ENCSR948RYN</a>	ChIP-seq	liver	H3K9ac	H3K9ac ChIP-seq on embryonic 11.5 day mouse liver	Bing Ren, UCSD
<a href="#">ENCSR058DOA</a>	ChIP-seq	liver	H3K27ac	H3K27ac ChIP-seq on embryonic 11.5 day mouse liver	Bing Ren, UCSD



# encodeproject.org: “Rapid Release”

The screenshot shows a web browser window with two tabs: 'ENCSR592MSV - ENCODE' and 'ENCBS746POP - ENCODE'. The address bar shows the URL 'https://www.encodeproject.org/experiments/ENCSR592MSV/'. The navigation bar includes 'ENCODE', 'Data', 'Methods', 'About ENCODE', and 'Help', along with a search bar and a 'Sign in' button. The main content area displays the following information:

**Run type:** single-end  
**Read length:** 100 nt

**Files linked to ENCSR592MSV**

Accession	File type	Output type	Paired end	Associated replicates	Added by	Date added	File download	Validation status
ENCF002ERQ	fastq	reads		(1, 1)	Diane Trout	2014-07-22	<a href="#">Download</a>	pending
ENCF002ERR	fastq	reads		(2, 1)	Diane Trout	2014-07-22	<a href="#">Download</a>	pending
ENCF002ERS	fastq	reads		(2, 1)	Diane Trout	2014-07-22	<a href="#">Download</a>	pending
ENCF002ERT	fastq	reads		(1, 1)	Diane Trout	2014-07-22	<a href="#">Download</a>	pending



# encodeproject.org: Summary

---

- An easy-to-use portal to ENCODE metadata and data.
- Continuously updated with ENCODE “rapid release” data.
- Supports faceted browsing and search.
- Links to visualize and download the data.
- REST API for programmatic access ... see help docs or ask us.

