STANFORD
SCHOOL OF MEDICINE

ENCODEPROJECT.ORG
USING THE ENCODE PORTAL

ENCODE DCC

ASHG ENCODE Workshop

October, 2014

STANFORD
SCHOOL OF MEDICINE

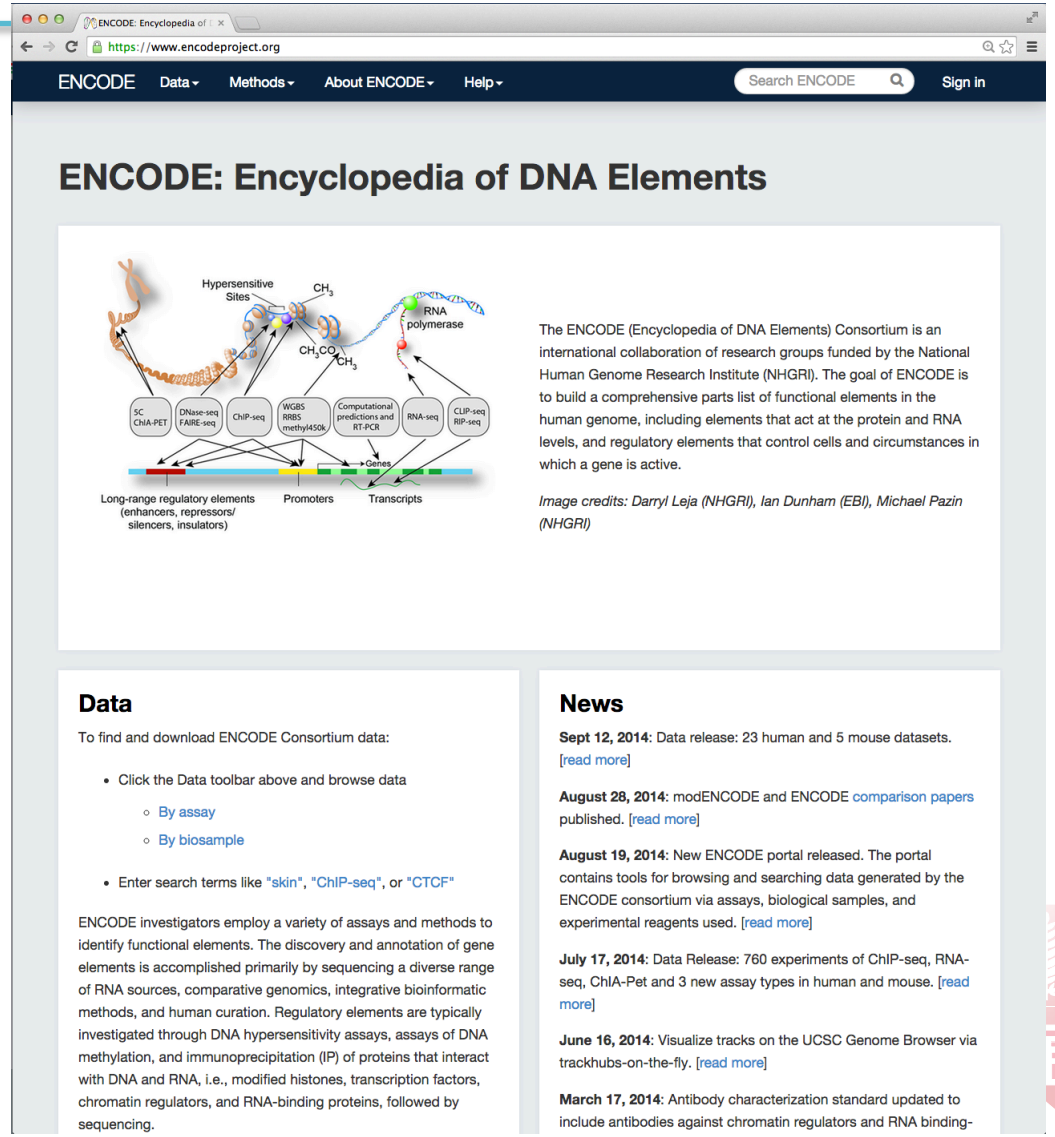
J. Seth Strattan, PhD ENCODE DCC

encodeproject.org: Motivation

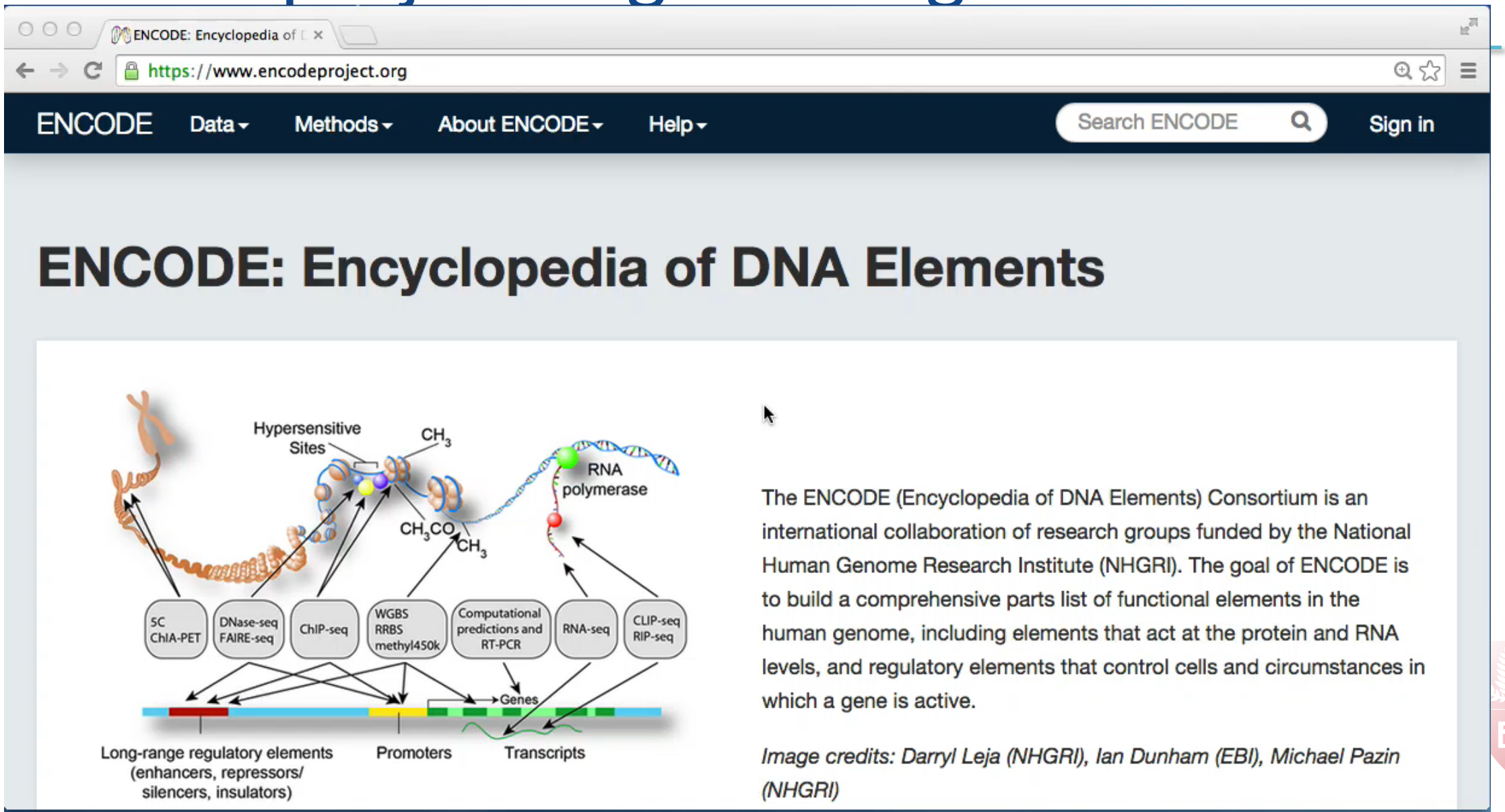
- ENCODE data can be difficult to find.
  - Different data are in different places.
  - Browser tracks, ftp sites, web sites.
- Metadata describing experiments can be hard to search.
- Where is the “official” list of everything ENCODE has done?
- Can I just download the ChIP data on ZNF143?

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.



encodeproject.org: Getting Started



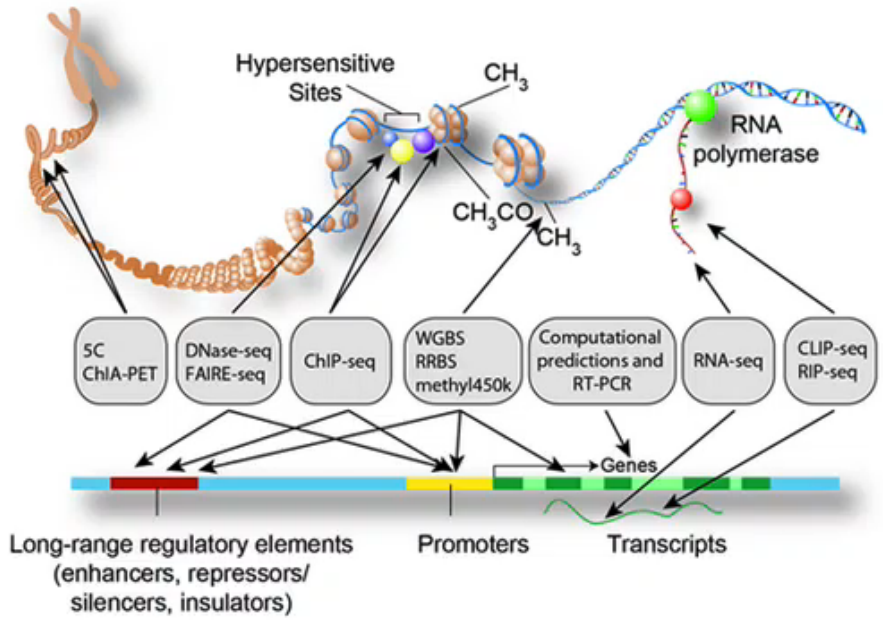
encodeproject.org: The Menu Bar

ENCODE: Encyclopedia of

https://www.encodeproject.org

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.

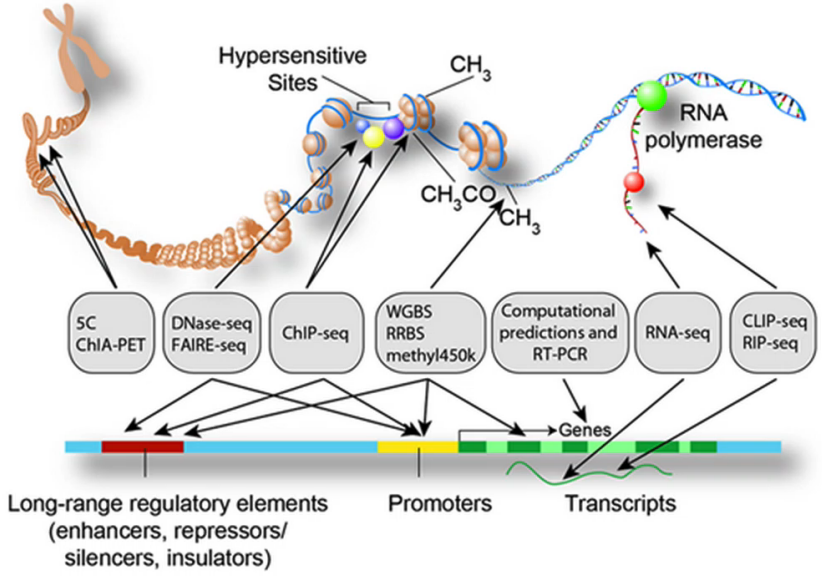
Image credits: Darryl Leja (NHGRI), Ian Dunham (EBI), Michael Pazin (NHGRI)

encodeproject.org: Faceted Browsing

https://www.encodeproject.org

ENCODE Data Methods About ENCODE Help Search ENCODE Sign in

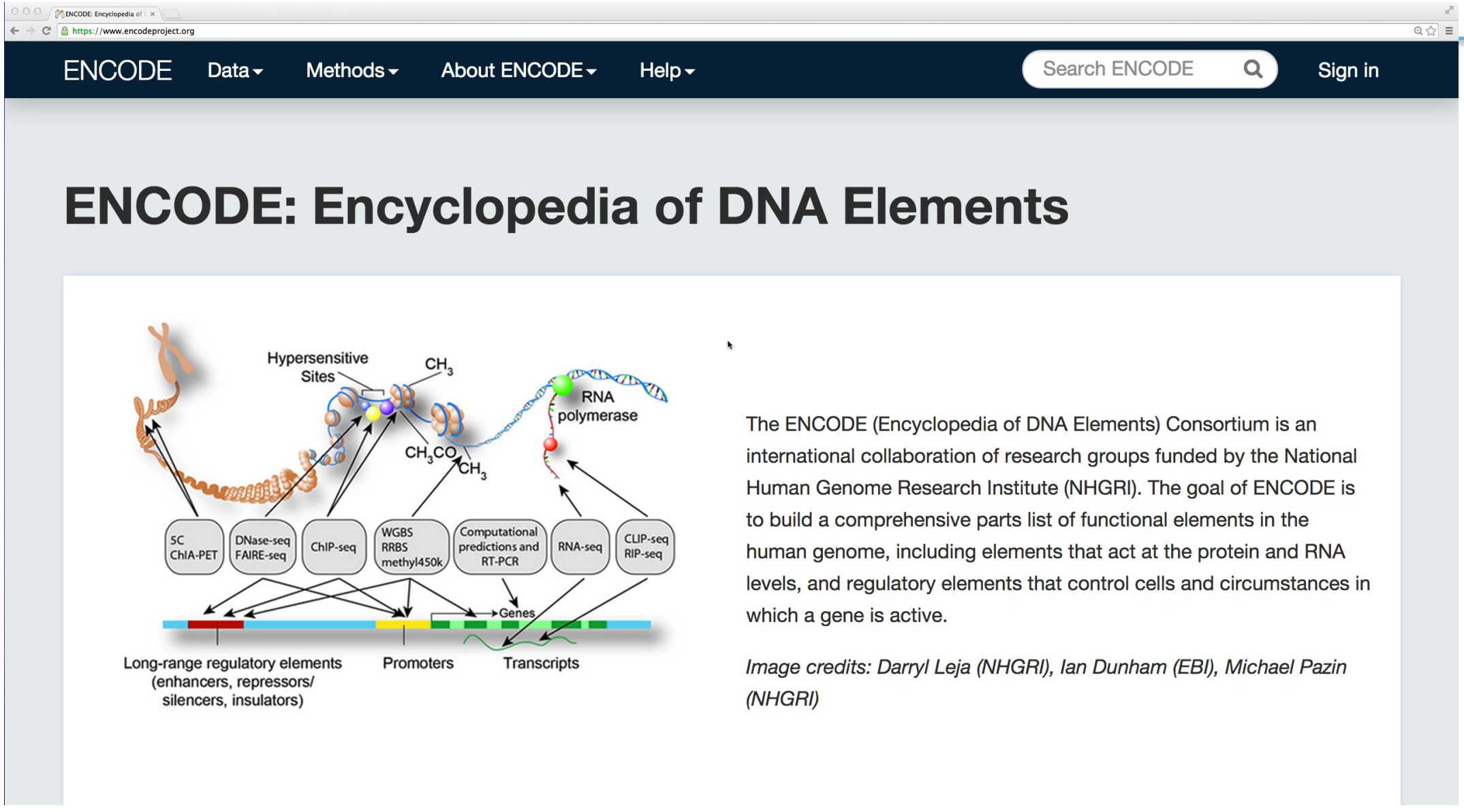
ENCODE: Encyclopedia of DNA Elements



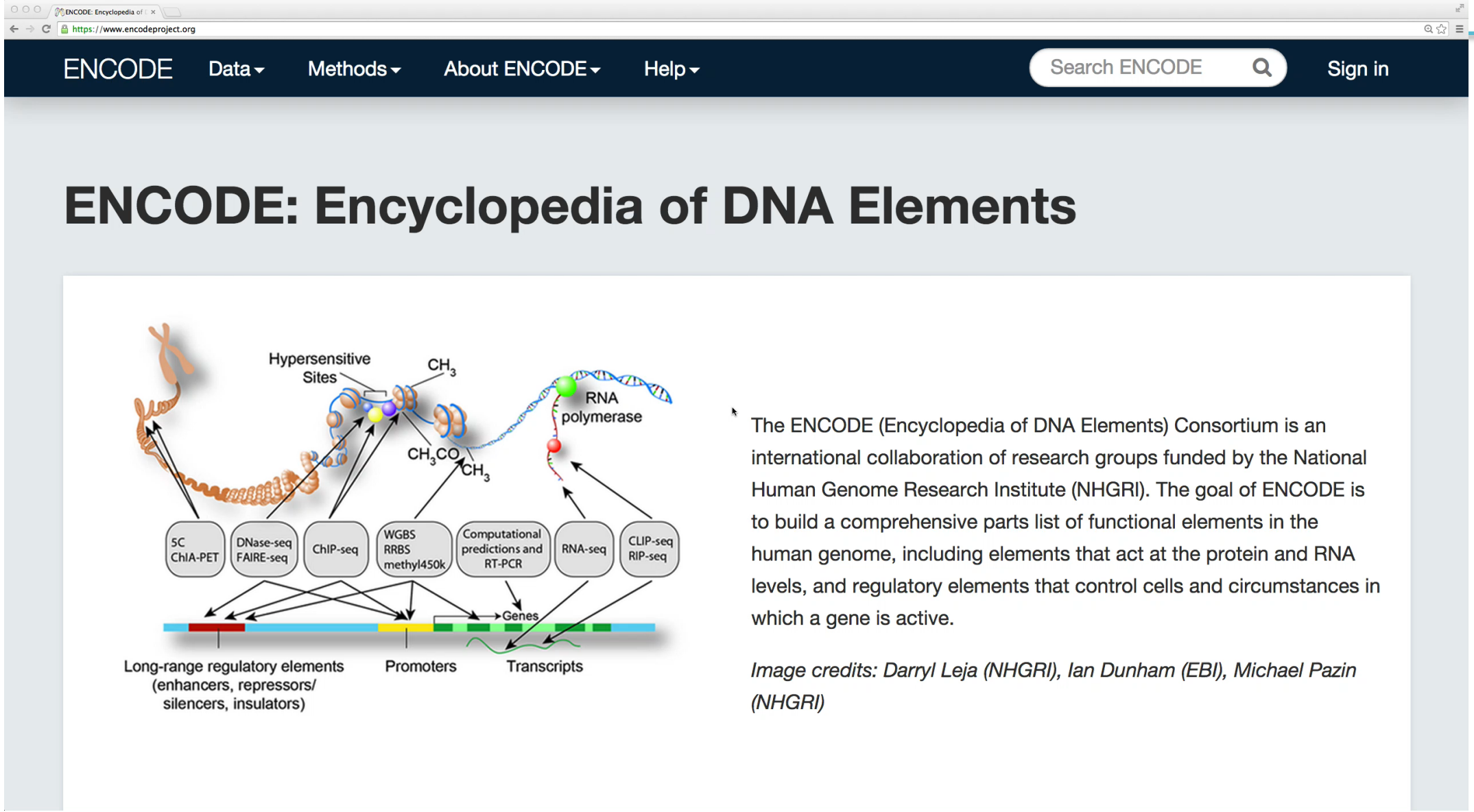
The ENCODE (Encyclopedia of DNA Elements) Consortium is an international collaboration of research groups funded by the National Human Genome Research Institute (NHGRI). The goal of ENCODE is to build a comprehensive parts list of functional elements in the human genome, including elements that act at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

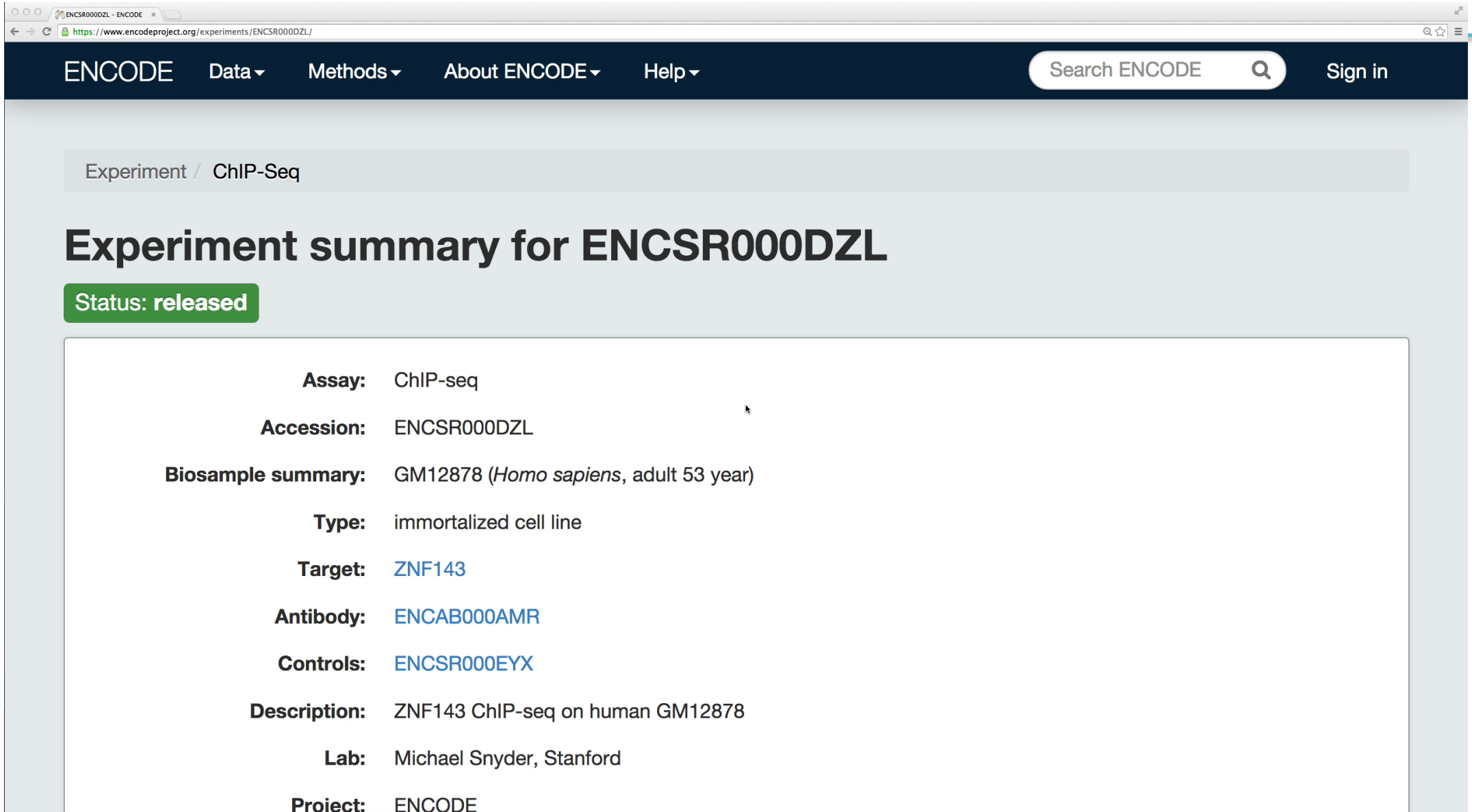
encodeproject.org: The Search Box



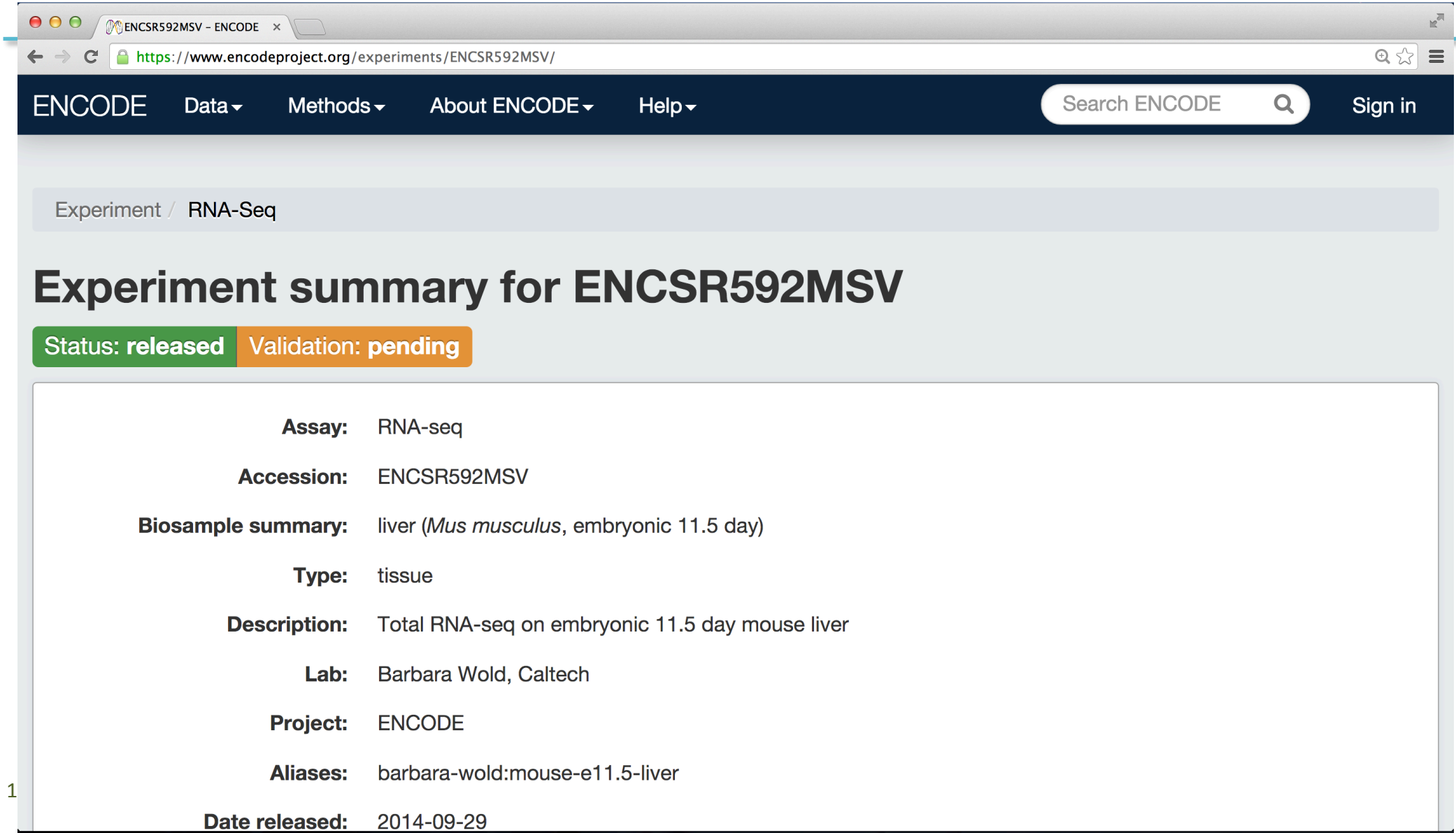
encodeproject.org: The Search Box



encodeproject.org: Links to data



encodeproject.org: “Rapid Release”



encodeproject.org: “Rapid Release”

ENCSR592MSV – ENCODE

https://www.encodeproject.org/experiments/ENCSR592MSV/

ENCODE Data Methods About ENCODE Help Search ENCODE Sign in

Assay details

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

Documents

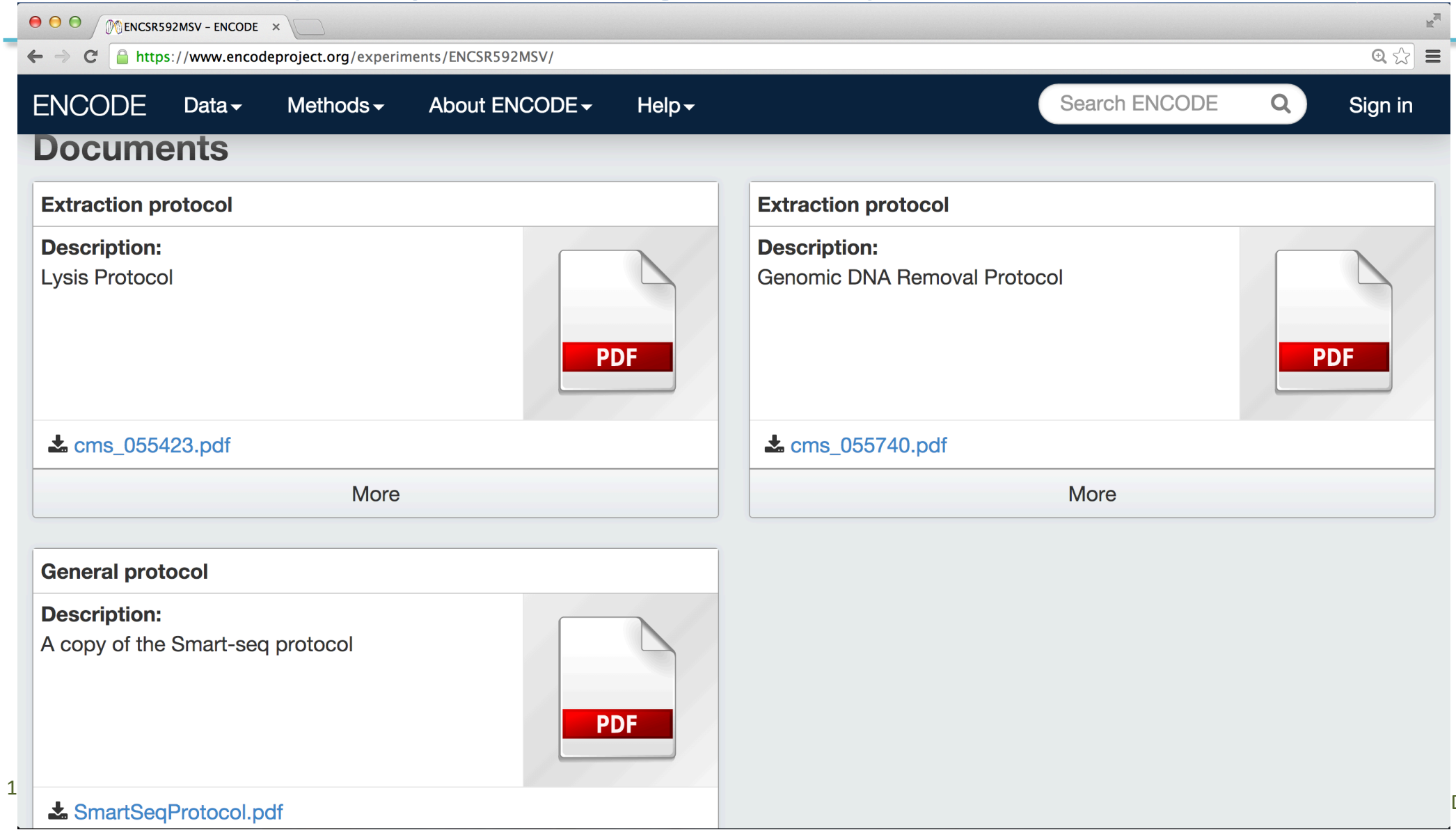
Extraction protocol

Description:
Lysis Protocol

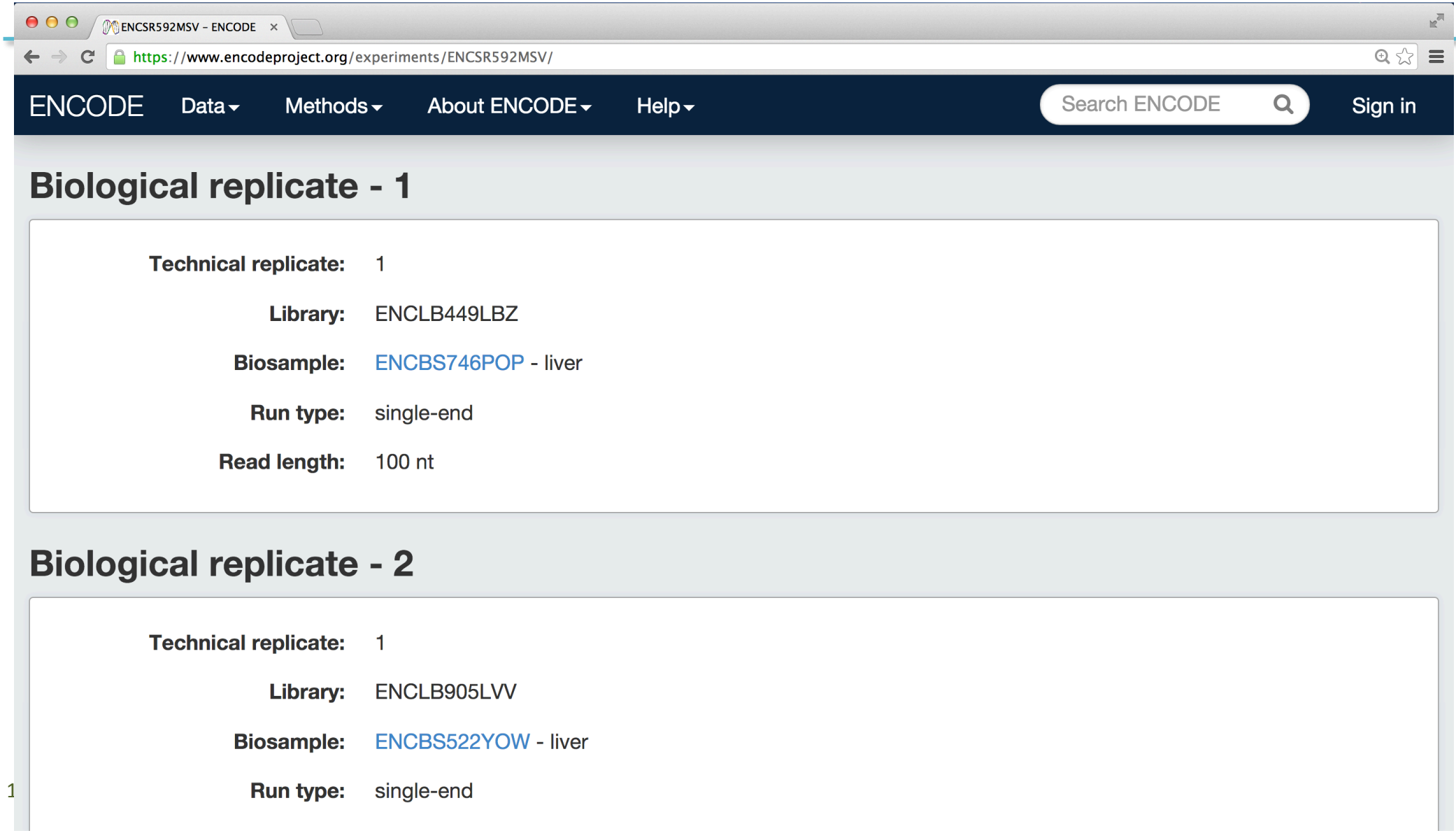
Extraction protocol

Description:
Genomic DNA Removal Protocol

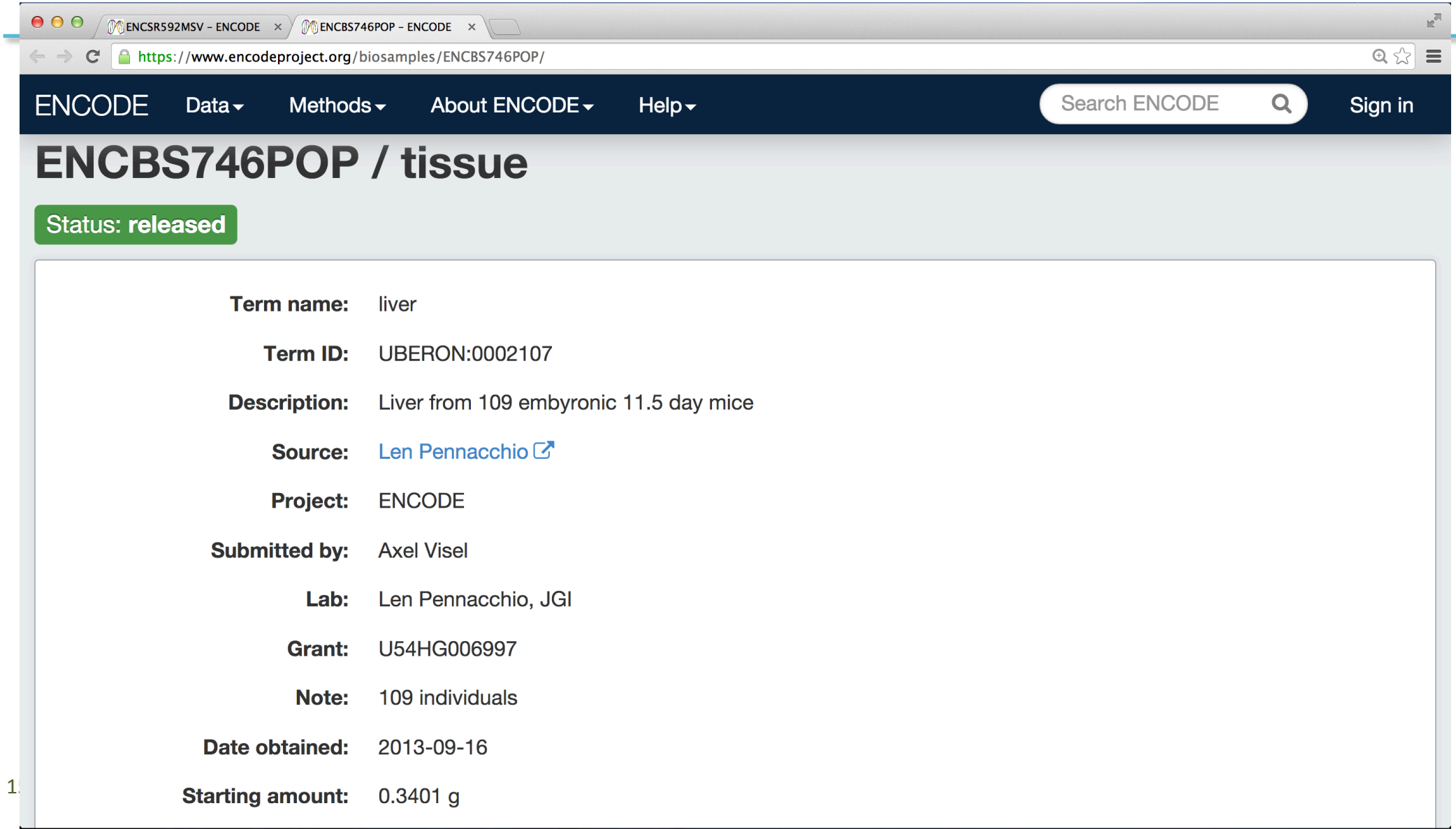
encodeproject.org: “Rapid Release”



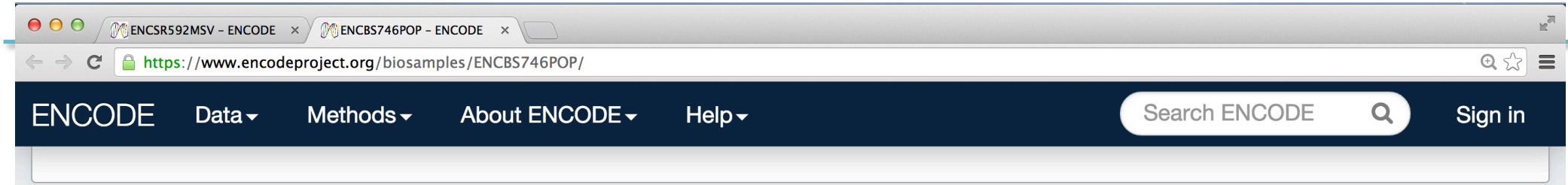
encodeproject.org: “Rapid Release”



encodeproject.org: “Rapid Release”



encodeproject.org: “Rapid Release”

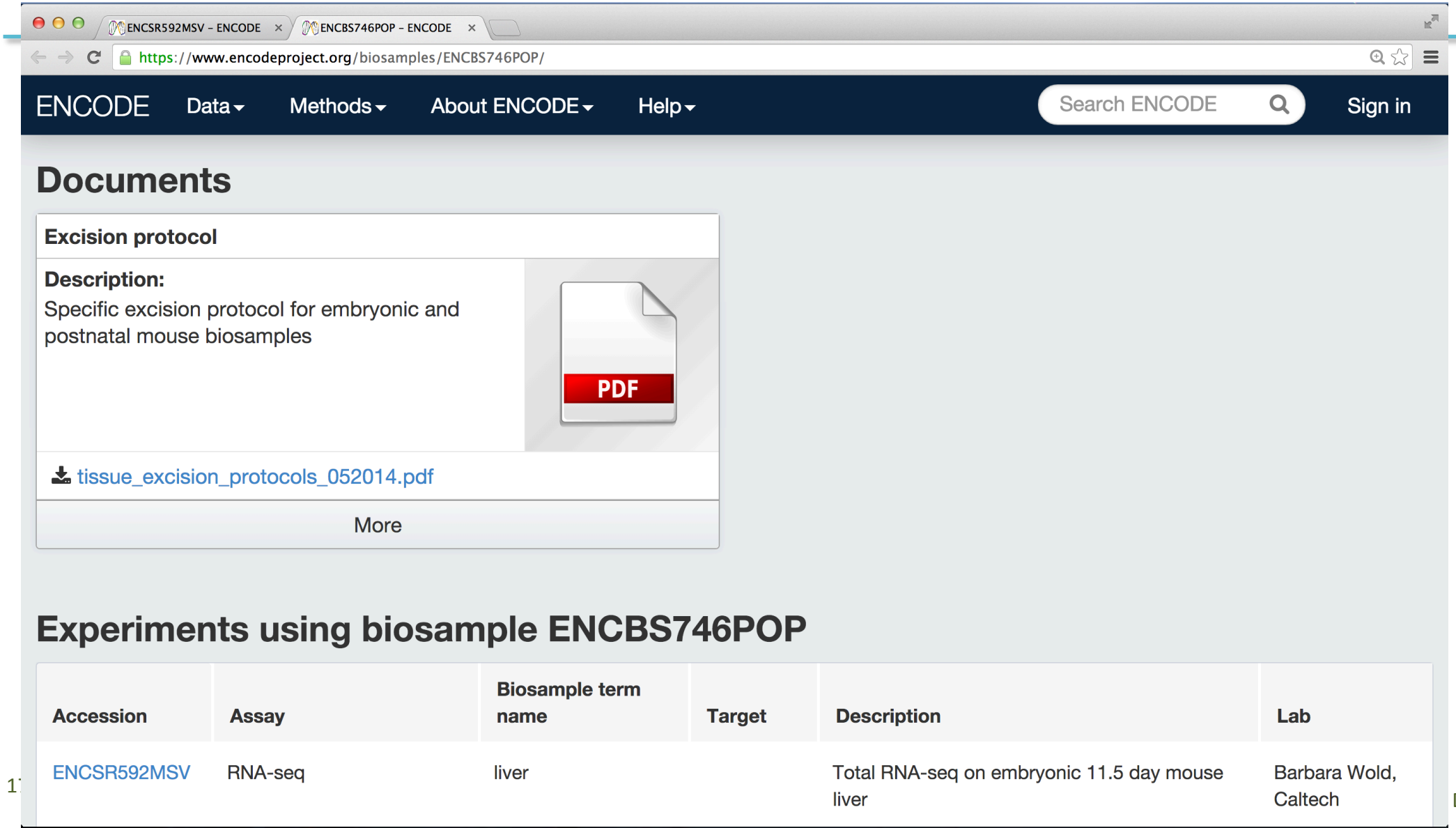


Strain information

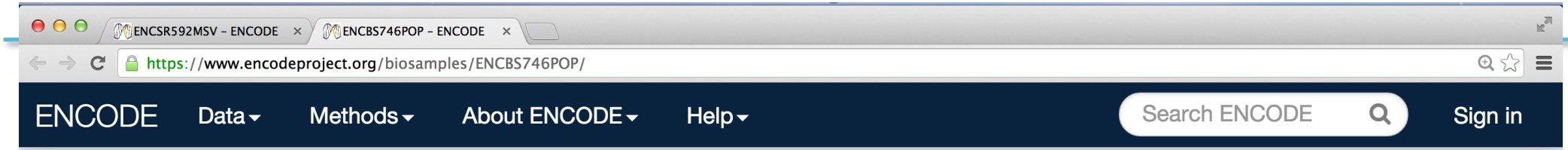
Accession:	ENCDO956IXV
Aliases:	ENCODE:B6NCrl
Species:	*Mus musculus*
Life stage:	Embryonic
Age:	11.5 day
Sex:	Mixed
Health status:	Healthy
Strain background:	C57BL/6
Strain name:	B6NCrl

Documents

encodeproject.org: “Rapid Release”



encodeproject.org: “Rapid Release”

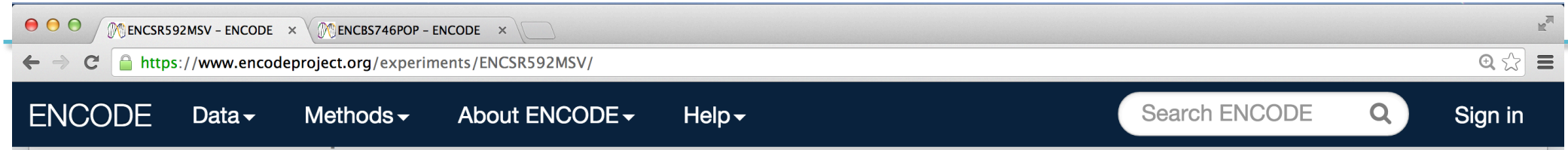


Experiments using biosample ENCBS746POP

Accession	Assay	Biosample term name	Target	Description	Lab
ENCSR592MSV	RNA-seq	liver		Total RNA-seq on embryonic 11.5 day mouse liver	Barbara Wold, Caltech
ENCSR932BNP	ChIP-seq	liver	H3K36me3	H3K36me3 ChIP-seq on embryonic 11.5 day mouse liver	Bing Ren, UCSD
ENCSR033PGF	whole genome bisulfite sequencing	liver		Whole genome bisulfite-seq on embryonic 11.5 day mouse liver	Joe Ecker, Salk
ENCSR573MJR	ChIP-seq	liver	Control	Input control ChIP-seq on embryonic 11.5 day mouse liver	Bing Ren, UCSD
ENCSR244KCW	ChIP-seq	liver	H3K27me3	H3K27me3 ChIP-seq on embryonic 11.5 day mouse liver	Bing Ren, UCSD
ENCSR948RYN	ChIP-seq	liver	H3K9ac	H3K9ac ChIP-seq on embryonic 11.5 day mouse liver	Bing Ren, UCSD
ENCSR058DQA	ChIP-seq	liver	H3K27ac	H3K27ac ChIP-seq on embryonic 11.5 day	Bing Ren, UCSD

DCC

encodeproject.org: “Rapid Release”



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
ENCFF002ERQ	fastq	reads		(1, 1)	Diane Trout	2014-07-22	Download	pending
ENCFF002ERR	fastq	reads		(2, 1)	Diane Trout	2014-07-22	Download	pending
ENCFF002ERS	fastq	reads		(2, 1)	Diane Trout	2014-07-22	Download	pending
ENCFF002ERT	fastq	reads		(1, 1)	Diane Trout	2014-07-22	Download	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.