FIND AND VISUALIZE ENCODE DATA
WWW.ENCODEPROJECT.ORG

J. Seth Strattan, PhD
ENCODE Data Coordinating Center (DCC)
Keystone Chromatin and Epigenetics
March, 2016
ENCODE DCC Delivers ENCODE Data

Sample → Library → Primary Data → Processed Data

Sample Library Primary Data Processed Data

ENCODE DCC

ENCODE Files

AWS S3 Bucket

J. Seth Strattan, PhD ENCODE DCC
ENCODE DCC Delivers ENCODE Metadata

Sample → Library → Primary Data → Processed Data
The ENCODE Portal is the **canonical source** for ENCODE metadata and data.

The Portal also documents **ENCODE standards** like antibody standards, data release.

The Portal links to **documentation and tutorials**.

Use the Portal to **browse, search, and visualize** what ENCODE has done.

Understand **ENCODE data analyses**, and how you can replicate them.

**REST API** for programmatic **search and download** of ENCODE metadata and data.
www.encodeproject.org ... Getting Started

Help Menu Documents the Portal

• *Getting Started*: Portal organization
  - [https://www.encodeproject.org/help/getting-started](https://www.encodeproject.org/help/getting-started)

• *Tutorials*:
  - [https://www.encodeproject.org/tutorials](https://www.encodeproject.org/tutorials)

• Links to slides from today’s workshop

• *Contact*:
  - encode-help@lists.stanford.edu
  - @encodedcc
www.encodeproject.org ... Getting Started

- [https://www.encodeproject.org/help/getting-started](https://www.encodeproject.org/help/getting-started)
- How to browse and facet.
- How to access one file at a time.
- Bulk download of files.
- ENCODE data model and data organization.
Faceted Browsing of Assays

- Click on Data ... Assays
- [https://www.encodeproject.org/search/?type=Experiment](https://www.encodeproject.org/search/?type=Experiment)
- In the center is a list of all ENCODE experiments.
- To the left are facets you can choose to filter the list.
Keyword search powered by ontologies

- Search for “skin” ... select Experiments.
- Finds “skin of body” and “keratinocyte”
www.encodeproject.org ... The Matrix

Click the matrix icon near “results”
www.encodeproject.org ... The Matrix

The same list in matrix form

- See the search results as a matrix.
- Facts work here too.
- Select ChIP-seq ... histone.
Skin ChIP-seq ... histone

- Cells in the matrix are clickable.
- Produces a list of experiments that match the selected criteria.
Use metadata to find data

- Filter: RNAseq, ENCODE, mm10 assembly
- Select an experiment: [https://www.encodeproject.org/experiments/ENCSR466KZY/](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
www.encodeproject.org ... Metadata to Data

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
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.
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.
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.
Use metadata to find data:

- Search for “H3K9ac neural tube”.
- Select an experiment, for example
  
  https://www.encodeproject.org/experiments/ENCSR087PLZ/

- Note metadata on protocols, replicates.
- Graph: files are related by processing steps.
- Download from the graph or a list.
- Click on “Visualize Data” to visualize the results of this experiment.
Use metadata to find data:

- Search for “H3K9ac neural tube”
- Facet on ChIP-seq; mouse; mm10 assembly
- Select an experiment, for example
  [https://www.encodeproject.org/experiments/ENCSR087PLZ/](https://www.encodeproject.org/experiments/ENCSR087PLZ/)
- Note metadata on protocols, replicates
- Graph: files are related by processing steps
- Download from the graph or a list
- Click on “Visualize Data”... mm10 to visualize the results of this experiment.
Visualize the Experiment

Adjust the browser settings to display fold-over-signal in "full"

UCSC Genome Browser on Mouse Dec. 2011 (GRCm38/mm10) Assembly

chr12:55,552,346-57,515,345 1,963,000 bp. enter position, gene symbol or search terms

Scale chr12:
ENCCF2222VYU
ENCCF3333VUE
ENCCF0999KDO
ENCCF9760G0
27.9911
ENCCF9871IK
22.9791
ENCCF2677NH
24.0769
ENCCF3222VUW

ChIP-seq of neural tube - ENCSR687FLZ (Target - H3K4ac)
ChIP-seq of neural tube - ENCSR687FLZ - ENCCF9871IK fold change over control rep 1
ChIP-seq of neural tube - ENCSR687FLZ - ENCCF2677NH fold change over control rep 2
ChIP-seq of neural tube - ENCSR687FLZ - ENCCF3222VUW fold change over control pooled from reps [1, 2]
Find Several Experiments

Use metadata to find data:

- Search for “H3K9ac neural tube”
- Get a list of several experiments
- Click on “Visualize Data” to visualize all the experiments matching this search.
Stage-dependent H3K9ac signal present at Pax9 in neural tube at e11.5, e13.5.
Find & Download Several Experiments

Use metadata to find data:

- Search for “H3K9ac neural tube”
- Get a list of several experiments
- Click on “Download” to download selected metadata and complete links to data.
Download Several Experiments

Use metadata to find data:

• Search for “H3K9ac neural tube”
• Get a list of several experiments
• Click on “Download” to download selected metadata and complete links to data.
Download Several Experiments

• “Download” produces a file with a list of links to all the files for all the experiments in your search.

• You can iterate through the list in your own script.

• Or:
  
xargs -n 1 curl -O -L < files.txt

• The first link is to a file called metadata.tsv that contains metadata you need to interpret what each file is.
Download Several Experiments

- metadata.tsv: Each line contains metadata on a file from the download package.
Programmatic access via the ENCODE 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

```
#!/usr/bin/env python
import requests
URL = 'https://www.encodeproject.org/experiments/ENCSR236EGS/?format=json'
response = requests.get(URL)
experiment = response.json()
print experiment['accession']
print experiment['description']
```
Programmatic access via the ENCODE REST API

```python
#!/usr/bin/env python

import requests

URL = 'https://www.encodeproject.org/experiments/ENCSR236EGS/?format=json'

response = requests.get(URL)

experiment = response.json()

print(experiment['accession'])
print(experiment['description'])
```

jseth:Keystone Epigenomics 2015 jseth$ ./GET_object.py
ENCSR236EGS
RNA-seq on a dissected area of layer V from an 8 month old male wild type C57Bl6 mouse
jseth:Keystone Epigenomics 2015 jseth$
Programmatic access via the ENCODE REST API

```python
#!/usr/bin/env python

import requests

URL = ('https://www.encodeproject.org/search/?'
  'type=experiment&'
  'assay_term_name=ChIP-seq&'
  'replicates.library.biosample.donor.orgism.scientific_name=Homo sapiens&'
  'target.investigated_as=transcription factor&'
  'replicates.library.biosample.biosample_type=in vitro differentiated cells&'
  'format=json')

response = requests.get(URL)

search_result = response.json()]['@graph']

# extract and print the target for each experiment
print '
'.join(['experiment['target']['label'] for experiment in search_result])
```
Programmatic access via the ENCODE REST API

```python
#!/usr/bin/env python
import requests

URL = ('http://example.com/REST API endpoint

response = requests.get(URL)
search_result = response.json() ['@graph']

#extract and print the target for each experiment
print '
'.join([experiment['target']['label'] for experiment in search_result])
```
The ENCODE Portal: Recap

- Interactive access to ENCODE metadata via faceted browsing and search
- Interactive retrieval of ENCODE data one file at a time
- Batch download of ENCODE metadata and data files
- Programmatic access using the ENCODE REST API

Next: ENCODE’s Integrative Results and More Visualization
Contributors

ENCODE Data Coordinating Center
Mike Cherry, PI, Stanford
Jim Kent, co-PI, UCSC
Eurie Hong, Project Manager
**Pipeline Developers**
Ben Hitz, WGBS, Software Lead
Tim Dreszer, RNA-seq, DNase-seq
J. Seth Strattan, ChIP-seq

**Portal Developers**
Laurence Rowe
Nikhil Podduturi
Forrest Tanaka

**Data Wranglers**
Esther Chan
Jean Davidson
Venkat Malladi
Cricket Sloan
J. Seth Strattan

**QA & Biochemistry Assistance**
Brian Lee
Marcus Ho
Aditi Narayanan

**Support Staff**
Stuart Miyasato
Matt Simison
Zhenhua Wang

ENCODE Data Analysis Center
Zhiping Weng, PI, University of Massachusetts
Mark Gerstein, co-PI, Yale

**Methylation**
Junko Tsuji, U Mass
Eric Mendenhall, U Alabama, HAIB

**RNA-seq**
Alex Dobin, CSHL
Carrie Davis, CSHL
Rafael Irizarry, Harvard
Xintao Wei, UConn
Brent Gravely, UConn
Colin Dewey, U Wisconsin
Roderic Guigó, CRG
Sarah Djebari, CRG

**ChIP-seq**
Anshul Kundaje, Stanford
Nathan Boley, Stanford
Jin Lee, Stanford

@encodedcc
encode-help@lists.stanford.edu
https://github.com/ENCODE-DCC/