The following document is adapted from a protocol paper originally published in Current Protocols in Bioinformatics:

Jou J, Gabdank I, Luo Y, Lin K, Sud P, Myers Z, Hilton JA, Kagda MS, Lam B, O'Neill E, Adenekan P, Graham K, Baymuradov UK, R Miyasato S, Strattan JS, Jolanki O, Lee JW, Litton C, Y Tanaka F, Hitz BC, Cherry JM. The ENCODE Portal as an Epigenomics Resource. Curr Protoc Bioinformatics. 2019 Dec;68(1):e89. doi: 10.1002/cpbi.89. PMID: 31751002; PMCID: PMC7307447.
Available on Pubmed

The ENCODE portal as an epigenomics resource

Introduction
The Encyclopedia of DNA Elements (ENCODE) project has the goal of identifying all functional elements in the human and mouse genomes. The genomic data produced as a part of ENCODE is hosted on the ENCODE portal (https://www.encodeproject.org), along with data from a number of related projects. The ENCODE portal is actively maintained by the ENCODE Data Coordination Center (DCC) to ensure access to the most up-to-date analysis results and metadata for all available datasets.

This article covers examples of 3 common tasks relating to ENCODE portal navigation using a web browser, with a final section that discusses programmatic access to the ENCODE portal and underlying database via REST API.

Table of Contents

Section 1: How to query the portal
The ENCODE portal hosts a large corpus of data. For most users, only a small subset of data is relevant to their interests. Various portal features can be used to narrow down search results.

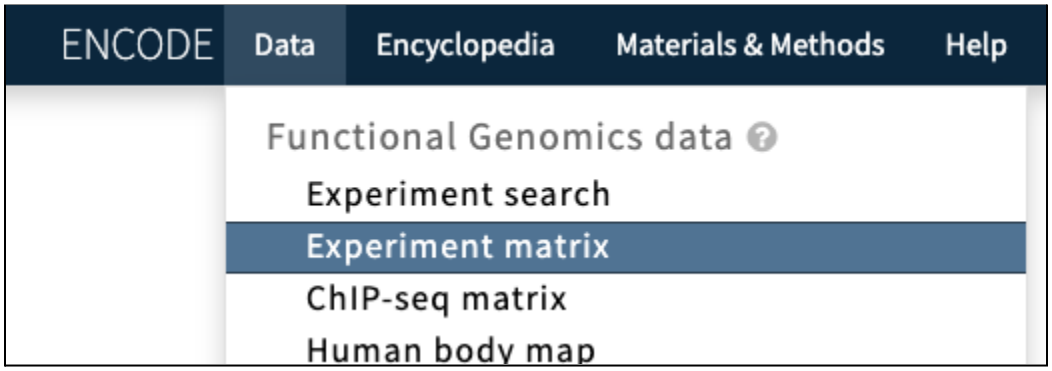
In this section of the guide, users will locate a specific TF ChIP-seq dataset on K562 targeting a cohesin-related target other than RAD21 by using search filters.

Requirements:
- Up-to-date web browser (Chrome, Microsoft Edge, Firefox, Safari)

Use the Matrix to navigate to a search page

1. Navigate to the ENCODE portal home page at https://www.encodeproject.org.
2. In the toolbar along the top of the page, click “Data” to open a drop-down menu with multiple options.

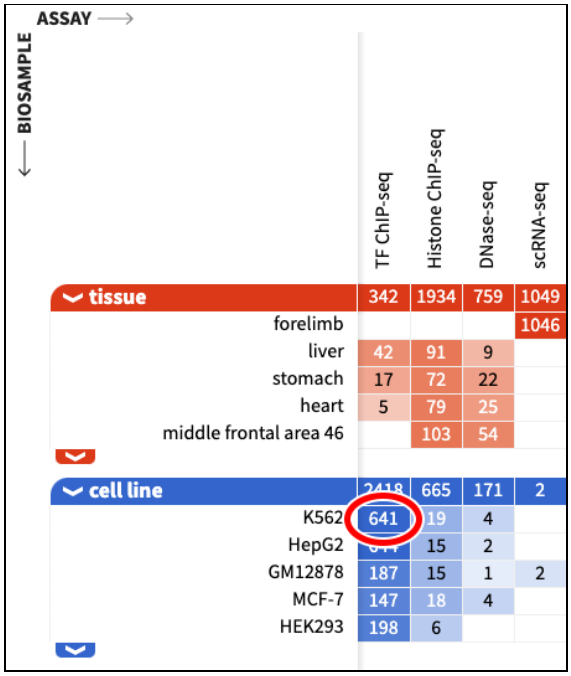
   Other menu options in the toolbar provide access to several key resources on the portal, separated by category.



3. In the drop-down menu, click “Experiment Matrix” to navigate to the Experiment matrix page.

   The Y-axis of the Experiment Matrix lists biosample types, which refers to the biological material used, such as a cell culture or tissue sample. The X-axis lists various assays. Each cell indicates the number of experiments for a given combination of assay and biosample type.

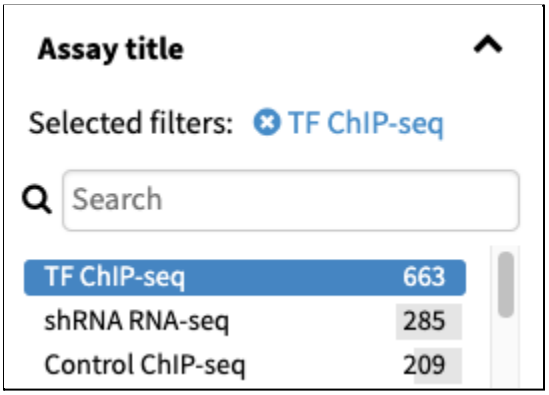
   Only a subsection of the matrix is visible upon page load. Click the arrows along the left side of the biosample category headers to reveal an extended list of biosample types. Scroll horizontally to view more available assays.
4. Click on the cell for transcription factor ChIP-seq (abbreviated as TF ChIP-seq) experiments on K562.



This link leads to a search page with a list of the experiments.

Filter search results using facets

5. Locate the “Assay title” facet and observe that the facet term “TF ChIP-seq” is highlighted in blue. This indicates that the search results have been filtered for experiments with an assay title of TF ChIP-seq. This filter is automatically applied when clicking a link from a matrix cell, as in Step 4.

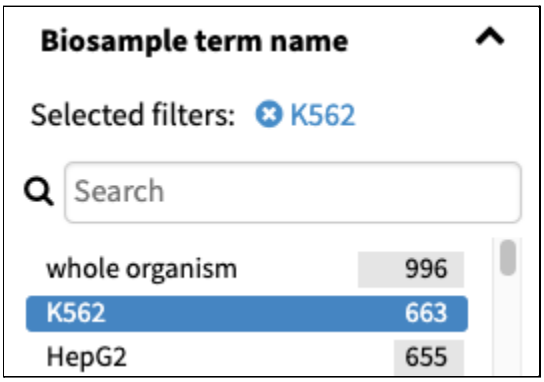


The sidebar on the left side of the search page is populated with multiple facets, which allow users to filter search results by different properties.

Clicking on a facet term applies that term as a filter. To remove the filter, click on the term a second time.

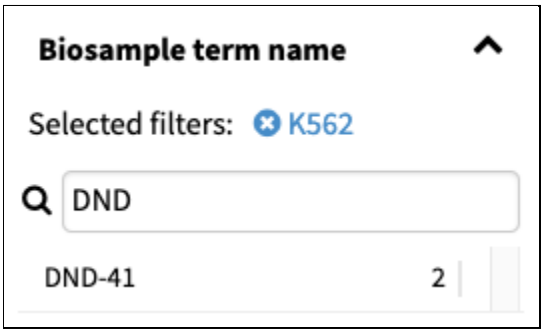
6. Scroll further down on the page and locate the “Biosample term name” facet. Observe that “K562” is already selected, indicating that the search results have been filtered for

experiments with a biosample term name of K562.



Selections can be made in more than one facet at a time. When such a selection is made, the combined filters possess an AND relationship. For example, the selection of TF ChIP-seq in the “Assay title” facet and K562 in the “Biosample term name” facet returns only experiments which are TF ChIP-seq assays AND use K562 cells as the biosample.

7. [Optional] Practice using typeahead search
    a. In the typeahead search box below the “Biosample term name” label, start typing “DND-41.” The list of terms below the search box will dynamically filter.



    b. Click “DND-41” in the “Biosample term name” facet.

        More than one facet term can be selected in a single facet simultaneously. Multiple selections in one facet represent an OR relationship between the selected facet terms. In this example, the selection of “K562” and “DND-41” terms means the returned experiments may be on K562 OR DND-41 cells.

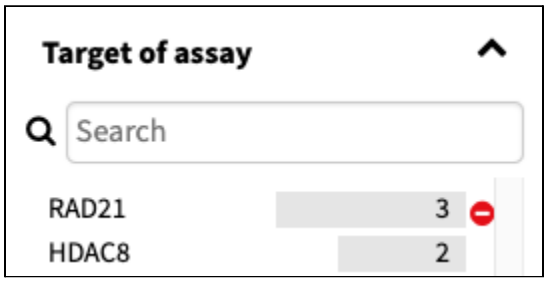
        Current URL: https://www.encodeproject.org/search/?type=Experiment&status=released&assay_title=TF+ChIP-seq&biosample_ontology.term_name=K562&biosample_ontology.term_name=DND-41

    c. Click “DND-41” in the facet term list a second time to remove the filter for DND-41 biosamples.

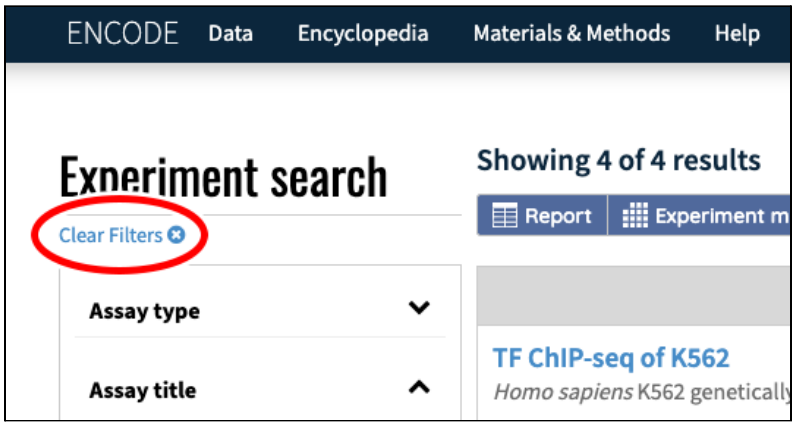
Current URL:
https://www.encodeproject.org/search/?type=Experiment&status=released&assay_title=TF+ChIP-seq&biosample_ontology.term_name=K562

8. Locate the “Target category” facet. Scroll through the list of terms and click on “cohesin.”

   Current URL:
   https://www.encodeproject.org/search/?type=Experiment&status=released&assay_title=TF+ChIP-seq&biosample_ontology.term_name=K562&target.investigated_as=cohesin
9. Locate the “Target of assay” facet. Hover the cursor over the “RAD21” facet term so that a red icon appears to the right. Click on the red icon to exclude experiments targeting RAD21.



   Current URL:
   https://www.encodeproject.org/search/?type=Experiment&status=released&assay_title=TF+ChIP-seq&biosample_ontology.term_name=K562&target.investigated_as=cohesin&target.label%21=RAD21
10. [Optional] Use “Clear filters” to remove applied filters.

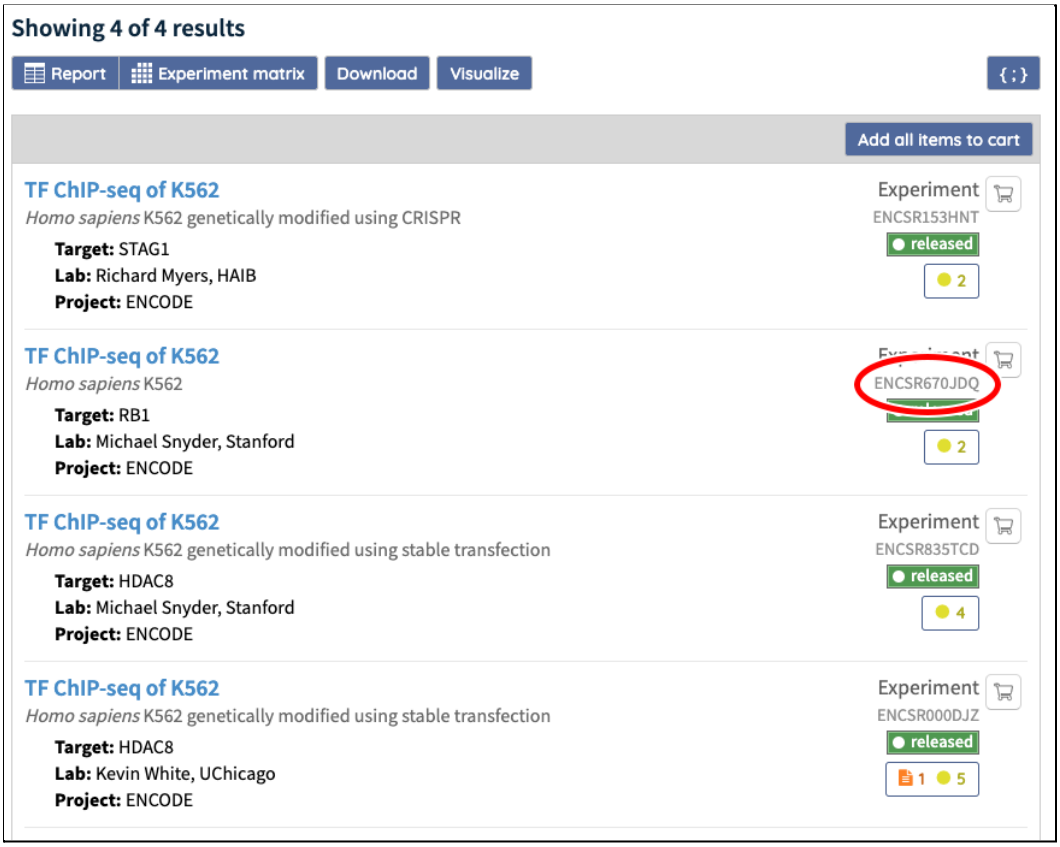


    a. Scroll to the top of the page. At the top of the facet sidebar below the words “Experiment search,” click “Clear filters” to remove all selected filters.
    b. Click the back button of the browser to undo the previous action. This will return to the previous query state.

11. Review the list of search results.

    Each result has a brief summary describing the dataset in question. Where applicable, the following information appears below the short title:
    a. Biosample summary
    b. Target
    c. Lab
    d. Project

    The accession, which is a unique and persistent identifier for each ENCODE dataset, appears on the right side of each search result, along with other details.

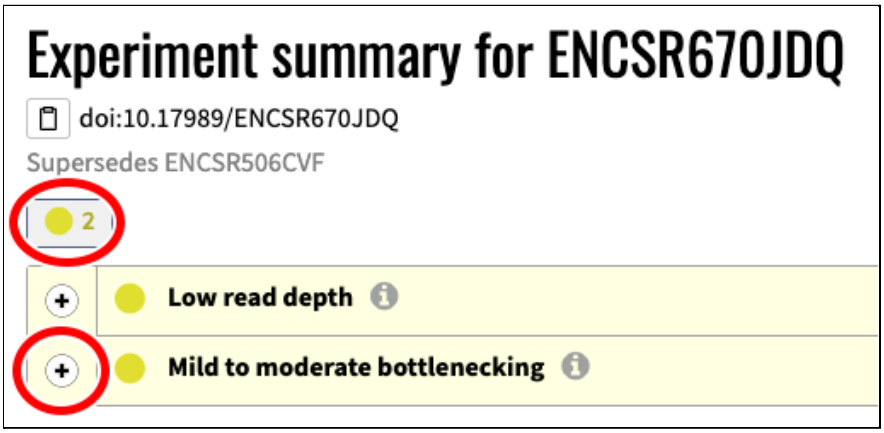


Visit a single experiment page

12. Click the short title corresponding to experiment ENCSR670JDQ to go to its experiment summary page.

    This page displays experimental metadata not shown on the search result page, including links to related experimental components, data files, and documents.
13. Click the audit button located below the page title, “Experiment summary for ENCSR670JDQ,” to display a list of audits. The plus symbol to the left of each audit

expands the audit to display the additional details.

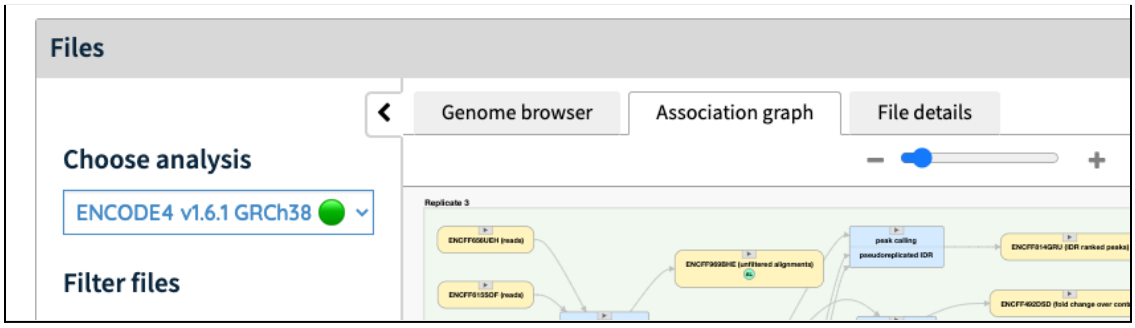


14. Scroll down the page to the Summary and Attribution sections.

    These sections contain general information about the experiment, such as the biosample, assay, target, a link to a control experiment, and information about the lab that performed the experiment.
15. Scroll down the page to the Replicates section.

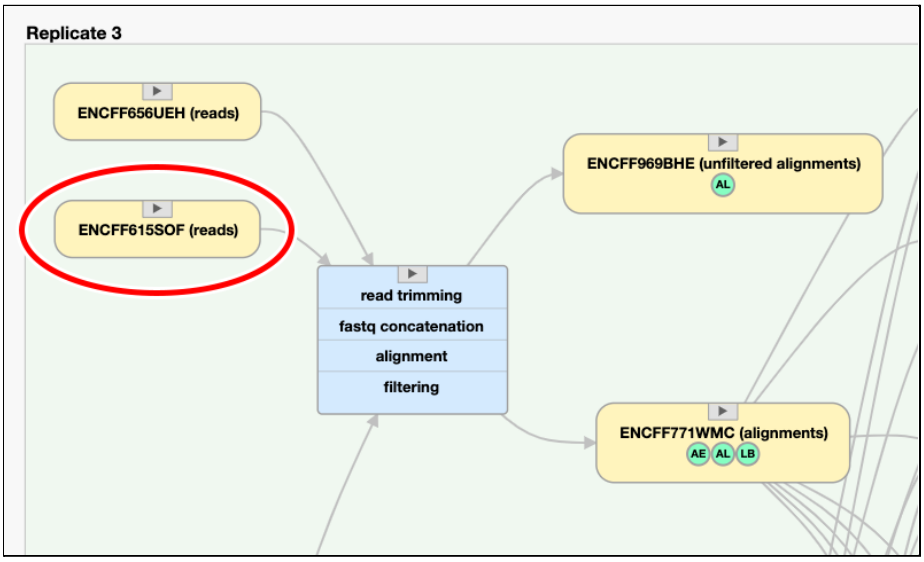
    This section contains information about the replicate(s) of the experiment, and links to biosamples, genetic modifications, and antibodies used when applicable.
16. Scroll down the page to the Files section.

    The Files section is divided into three tabs: Genome browser, Association graph, and File details. By default, the Files section displays the Association graph, which shows data provenance and derivation of downstream processed files. Use the slider above the graph to zoom in or out of the graph.

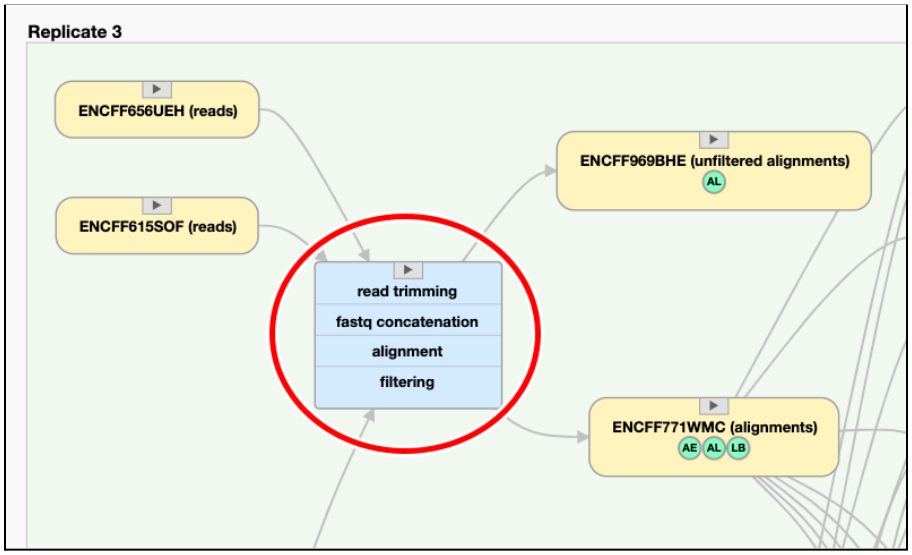


17. Click on a yellow File node in the Association graph to view a pop-up containing the file's unique accession as well as other metadata such as the file size, output type, and

submission date.



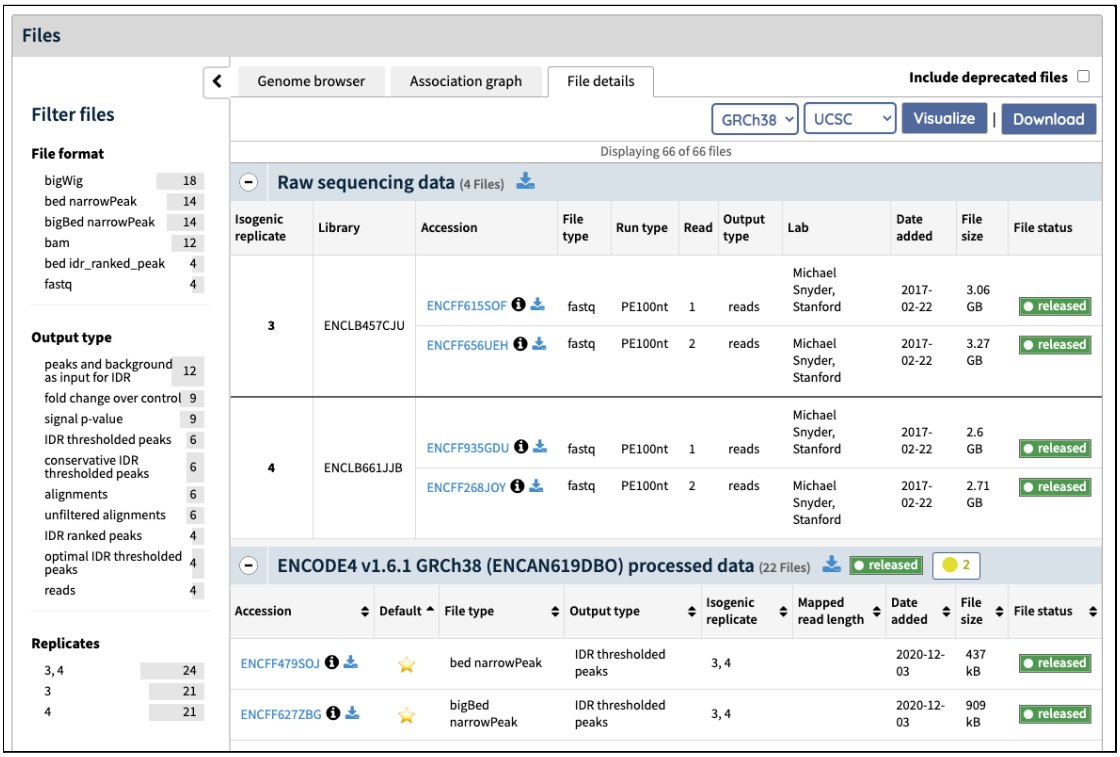
18. Click on a blue Pipeline Step node in the Association graph to view information about the relevant software, inputs and outputs, and purpose of the given pipeline analysis step.



19. Click the “File details” tab to view a table of all files linked to this experiment.

    Each row of the table contains information about the file’s output type, reference genome assembly for mapping, and submission date. Click the blue icon next to each file’s

accession to download the given file.



20. Scroll down to the Documents section.

    Experiments may also have attached documents describing the experimental or computational protocols. Click the link for a particular document to download it.

Section 2: How to batch download data

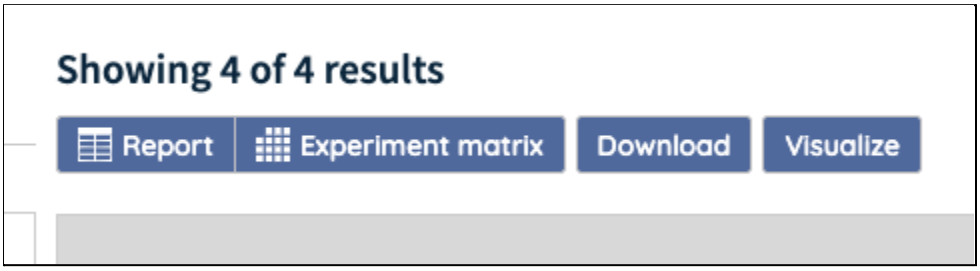
After identifying experiments of interest, users can download data associated with these experiments. Batch downloading provides a quick and simple way to download multiple files.

In this section, users will learn how to download multiple files from a search result page using the `curl` utility.

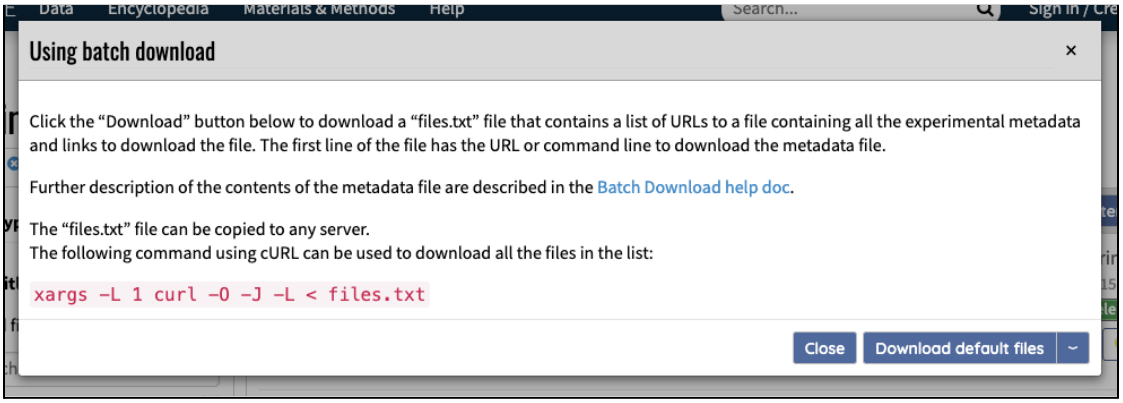
Requirements:

- Up-to-date web browser (Chrome, Microsoft Edge, Firefox, Safari)
- Command-line terminal
- Command-line UNIX utilities:
    - curl: https://curl.haxx.se/
    - xargs
- A text editor

1. Navigate to a search result page. Consult the “How to query the portal” steps 1-11 of this article for guidance.
2. Click the “Download” button located above the list of results. A pop-up containing detailed instructions appears.



3. In the lower right corner of the pop-up, click the “Download default files” button to download a text file named “files.txt.” This file contains a list of URLs of all files of each experiment returned in the search.



4. Open “files.txt” in a text editor.
5. Enter the first line of the file into a web browser to download a .tsv titled “metadata.tsv.”

   “Metadata.tsv” is a tab-delimited file containing metadata on all files listed in the “files.txt” file. This includes information about the file itself and its dataset, download links, and other properties. Users can utilize this metadata to filter the list of files to those of interest.

6. The remaining lines in “files.txt” are download links for every file associated with the experiments returned in the initial search. Use command-line utilities such as curl to download every link in the text file. An example of such a command is:

```
xargs -L 1 curl -O -L < files.txt
```

Section 3: How to visualize data

The ENCODE portal supports visualization of analysis results from one or multiple experiments at once using the external UCSC Genome Browser or Juicebox browser. The ENCODE portal

has also recently introduced embedded visualization of tracks directly on individual experiment pages using the Valis genome browser.

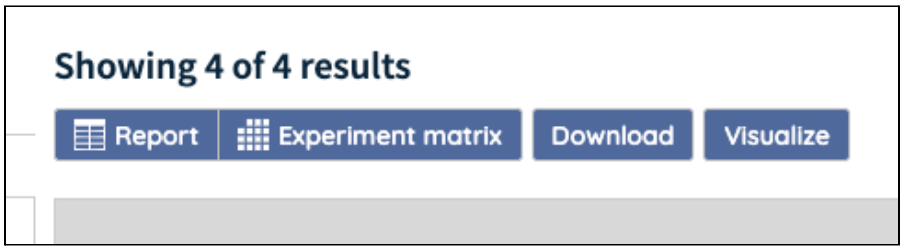
In this section, users will learn how to visualize data from multiple search results and from a single experiment page.

Requirements:

- Up-to-date web browser (Chrome, Microsoft Edge, Firefox, Safari)

Visualize results from multiple experiments from a search page

1. Navigate to a search page by following “How to query the portal” steps 1-11.
2. Click the “Visualize” button at the top of the page.



A pop-up appears with different options grouped by the available reference genome assemblies.
Due to genome browser limitations, there must be fewer than 100 results for visualization. If more than 100 results are returned, the pop-up will display the following warning message: “Filter to 100 to visualize.”

3. In the “GRCh38” category, click “UCSC” to visualize the data from the experiments using the UCSC Genome Browser in a new tab. Close the new tab when done visualizing the tracks.

Visualize results from a single experiment

4. To visualize results from a single experiment, navigate to an experiment summary page (“How to query the portal” step 12). Then, scroll down to the Files section.
5. Click on the “Genome browser” tab to visualize tracks using the embedded Valis genome browser.

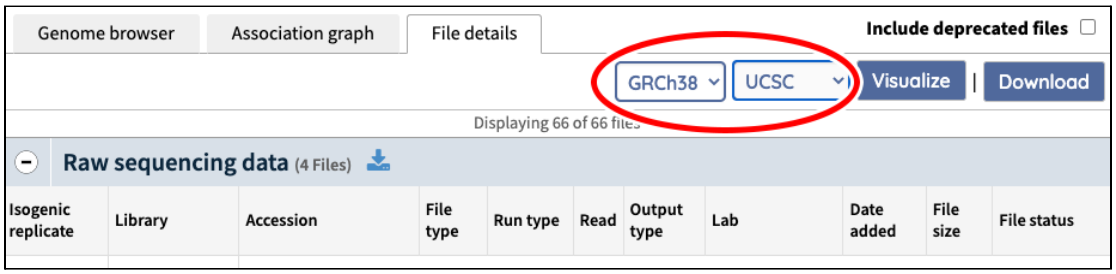
By default, all visualizable tracks will be shown. If available for the given reference genome assembly, gene and/or genome tracks are also displayed.

6. Click and drag to move along the genome. Alternatively, scroll left or right while hovering the mouse cursor over the tracks.
   Scroll up or down to change the zoom level of the tracks.
7. Select “signal p-value” below the “Output type” header to view 6 signal tracks.

   The sidebar in the Files section contains filters to switch reference genome assemblies or select specific tracks. Click the left-facing arrow near the top of the sidebar to collapse it. Click “Clear all filters” to remove all selected filters.
8. Enter “RBFOX1” into the search box located above the browser tracks and select the option that comes up. The genome browser will automatically refresh and load the location of the requested gene.

   The chromosome and coordinates are displayed below the search box. To manually edit the values, click the pencil icon to the right of the coordinates.

9. Click on the file details tab of the Files section. Use the drop-downs in the upper right corner of the tab to select your desired assembly and external genome browser.



10. Click the “Visualize” button to the right of the drop-down menus to open the external genome browser in a new tab.

Section 4: How to build queries and access the database programmatically

The ENCODE portal is a user-friendly interface for querying the ENCODE database. Queries can also be sent directly to the database using the Representational State Transfer (REST) API, bypassing the portal’s user interface. The REST API accepts URLs in a typical HTTP format, and returns metadata records as JSON objects. Search methods described in “How to query the portal” also make use of the REST API, but are “wrapped” in the portal’s user interface for users who prefer to interact visually.

Facets (see step 5 of “How to query the portal”) represent a subset of commonly used query parameters. However, any property of any object type can potentially be used as a query parameter. These properties and object types are listed comprehensively on the ENCODE portal at the following link: https://www.encodeproject.org/profiles/?

In this section, users will learn to construct a query to return search results for Biosamples of eye tissue from adult donors.

Requirements:

- Up-to-date web browser (Chrome, Microsoft Edge, Firefox, Safari)
- Command-line terminal
- Command-line UNIX utilities:
    - curl: https://curl.haxx.se/
    - Xargs

Optional tools:

- Python 3.5 or higher

1. Write down the first part of a query:
   `https://www.encodeproject.org/search/?`
2. Append the parameter `type=Biosample` to the query.

   All parameters take the format `property=value`. This example uses the “`type`”

property with a value of “Biosample.” All objects possess the “type” property, which indicates what category the object belongs to.

Current query: https://www.encodeproject.org/search/?type=Biosample

This query would return a list of all biosample objects on the portal.

3. Append another parameter to the query by first adding “&” to the end of the query. Then add the “`life_stage=adult`” parameter to the end.

   Add another parameter, “`biosample_ontology.organ_slims=eye`”, to filter by organ.

   Current query:
   https://www.encodeproject.org/search/?type=Biosample&life_stage=adult&biosample_ontology.organ_slims=eye

   Additional syntax for query building are listed in a table following this section.
4. Request the query using the following command, which uses curl to fetch a URL, in the command line terminal:

   ```
   $ curl -H "Accept: application/json"
   https://www.encodeproject.org/search/?type=Biosample&life_stage=adult&biosample_ontology.organ_slims=eye
   ```

   The terminal displays the results for the biosample search as a JSON object.
5. [Optional] Request the query using the Python requests library. The same JSON object is returned, and can be parsed using the json library.

   Below is a sample script. The script uses the `get` function of the requests library to retrieve the URL in question. Then, the script uses the json library to parse the response and print the response to the terminal.

```
import requests, json
headers = {'accept': 'application/json'}
response = requests.get(
'https://www.encodeproject.org/search/?type=Biosample&
life_stage=adult&biosample_ontology.organ_slims=eye',
headers=headers)

search_results = response.json()
```

```
print(json.dumps(biosample, indent=4))
```

Syntax for query building		
Syntax	Parameter example	Description
=	assay_title=TF ChIP-seq	The equal symbol (=) connects a property to its value.
&	assay_title=DNase-seq&assembly=mm10	The ampersand (&) symbol joins multiple parameters together.
!=	assembly!=hg19	!= represents “not equals” and acts as a negation. In the example, the parameter would filter for objects with reference genome assemblies which are not hg19.
*	treatment=*	The wildcard (*) means the parameter can have any value.
.	biosample_ontology.term_name=HepG2	The data model of the ENCODE portal allows for certain objects to be embedded in others. Objects are able to access the properties of the objects embedded in them. The period joins properties and sub-properties to form the “path” to an embedded property, akin to the forward slash (/) in file directory paths. In this example, the biosample_ontology object is embedded in experiment objects, and this parameter accesses the term_name property of biosample_ontology.
format	format=json	Format is a special property. Appending this to the query returns the page as a raw JSON object. The default value is HTML.
frame	frame=embedded	Frame is a special property indicating how the ENCODE database should return a requested object. Possible values include: ● Object: No objects are embedded

		● Embedded: All objects are embedded ● Raw: Object links are in UUID (Universally Unique Identifier) format, rather than the default @id format