User Manual

Target ALS Data Portal

March 28, 2024
Version 1.0

Contact
dataportalsupport@targetals.org

Background

The Target ALS Data Portal is a collective platform providing an environment for collaboration and access of data. The goal is to provide an open access data collection platform for improving ALS research, paving the way to understand the heterogeneity of ALS leading to innovative diagnostic and treatment strategies to cure ALS.

Main deliverables are:

- Harmonized multi-omics data collection: the Post-mortem data collection contains sequencing data (RNA-Seq, semiquantitative Tdp43, WGS) and imaging (histopathology). The Natural History Study collection contains sequencing data (WGS), clinical and demographic data and data from digital technologies measuring speech and motor function at-home. The data are well-structured including meta-data and mutation status to allow multi-omics landscaping of ALS.
- Easy access and data screening: the data can be accessed through the data portal including easy-to-use search engine for individual requests
- Interoperability for data analysis: the data can be easily accessed through the Verily Workbench which allows for downstream analysis using Jupyter Notebook

Purpose

The purpose of the Target ALS Data Portal User Manual is to provide guidance for how to make use of the data through the portal and Verily Workbench.

Requesting Data Access

Access to the data requires acceptance of the Target ALS Data Portal Data Use Agreement. Your application will typically take three business days to review and approve. Once data access has been granted, users will have access to all data on the portal.

Submit your access request by completing the Google Form. You will need a Google-enabled email address to apply for data access. If you wish to request access using an institutional email address, follow these instructions to register your email with Google.

You will receive a notification via email once your request has been processed.

Getting Started

Once your request for access has been approved, you can access the data portal at dataportal.targetals.org and Log in by clicking the "Log In" button in the top right corner. Follow the prompts to sign in using the same Google-based identity that you used to submit the access request form.

Once logged in, you may view the data in any collection by clicking on the "Explore Collections" button (Figure 1).

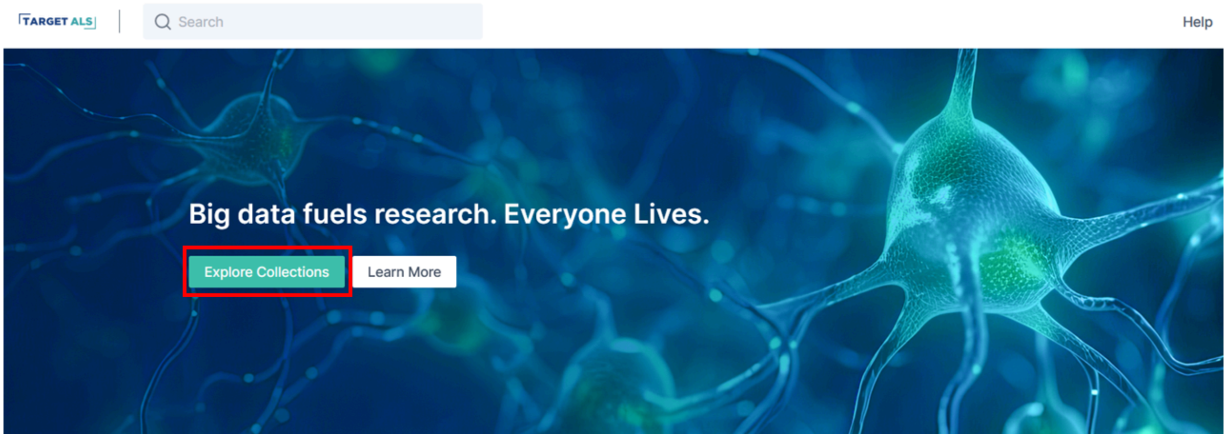
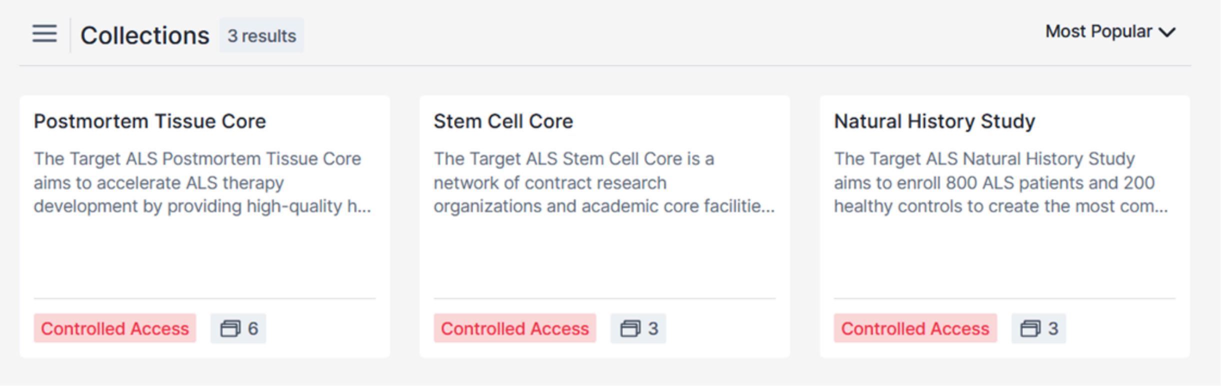


Figure 1 - Explorer entry page

Currently, the data portal contains three collections (Figure 2):

- Postmortem Tissue Core
- Stem Cell Core
- Natural History Study



Postmortem Tissue Core

The Target ALS Postmortem Tissue Core aims to accelerate ALS therapy development by providing high-quality h...

Stem Cell Core

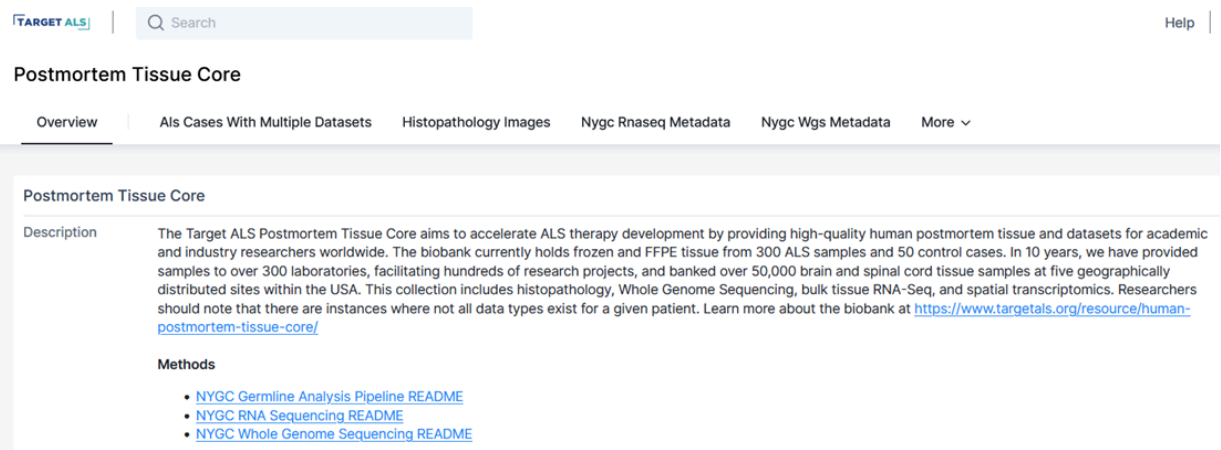
The Target ALS Stem Cell Core is a network of contract research organizations and academic core facilitie...

Natural History Study

The Target ALS Natural History Study aims to enroll 800 ALS patients and 200 healthy controls to create the most com...

Figure 2 - Data collection page

Clicking on the dataset links you to the dataset collection.



Postmortem Tissue Core

Postmortem Tissue Core

Description

The Target ALS Postmortem Tissue Core aims to accelerate ALS therapy development by providing high-quality human postmortem tissue and datasets for academic and industry researchers worldwide. The biobank currently holds frozen and FFPE tissue from 300 ALS samples and 50 control cases. In 10 years, we have provided samples to over 300 laboratories, facilitating hundreds of research projects, and banked over 50,000 brain and spinal cord tissue samples at five geographically distributed sites within the USA. This collection includes histopathology, Whole Genome Sequencing, bulk tissue RNA-Seq, and spatial transcriptomics. Researchers should note that there are instances where not all data types exist for a given patient. Learn more about the biobank at https://www.targetals.org/resource/human-postmortem-tissue-core/

Methods

- NYGC Germline Analysis Pipeline README
- NYGC RNA Sequencing README
- NYGC Whole Genome Sequencing README

Figure 3 - Data overview page

On each collection page, an **Overview** is provided outlining the available data tables (Figure 3). It contains a detailed description of the sequencing method and data processing.

Moreover, it provides a summary of basic statistics (Figure 4).

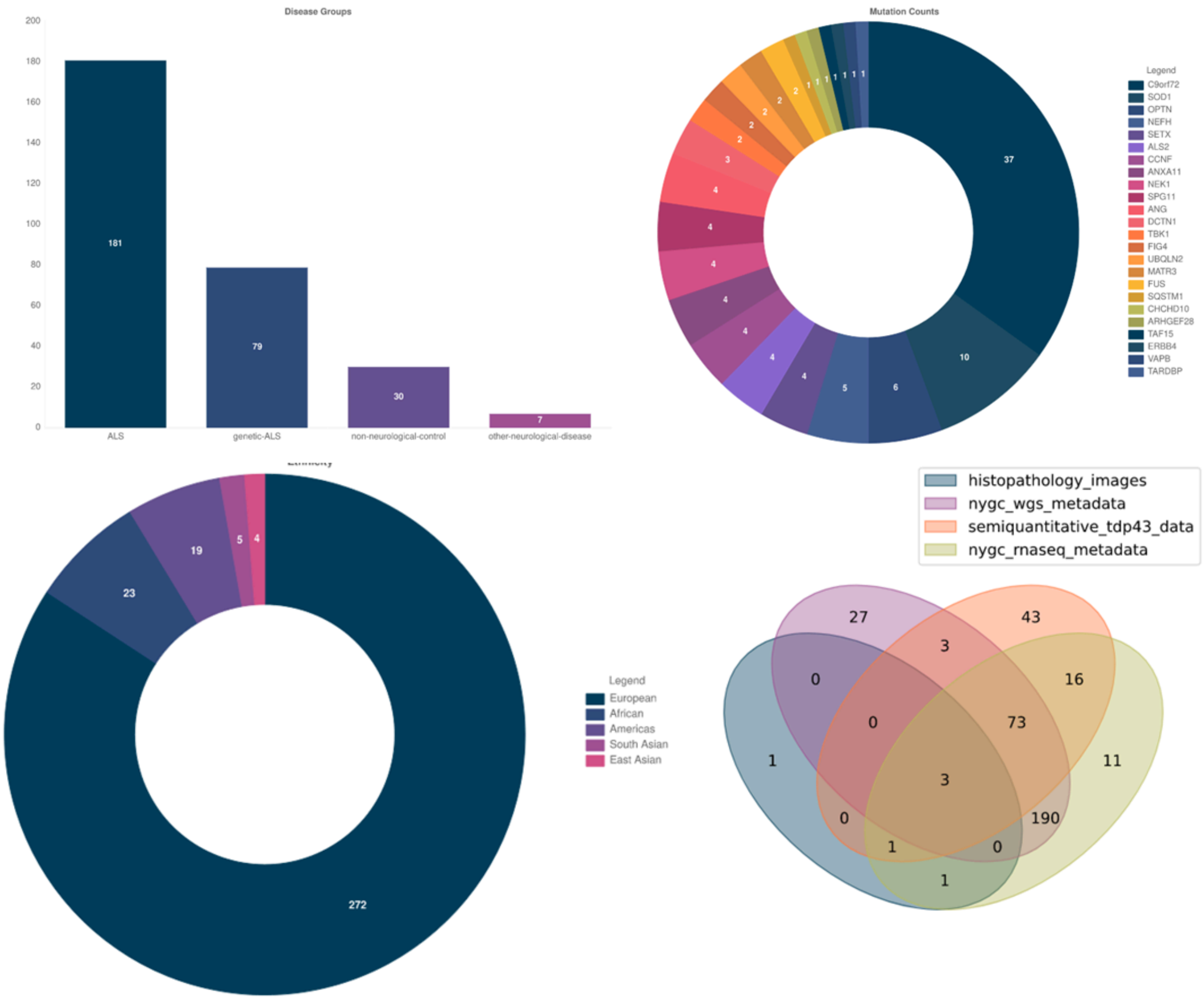


Figure 4 - Overview of basic statistics

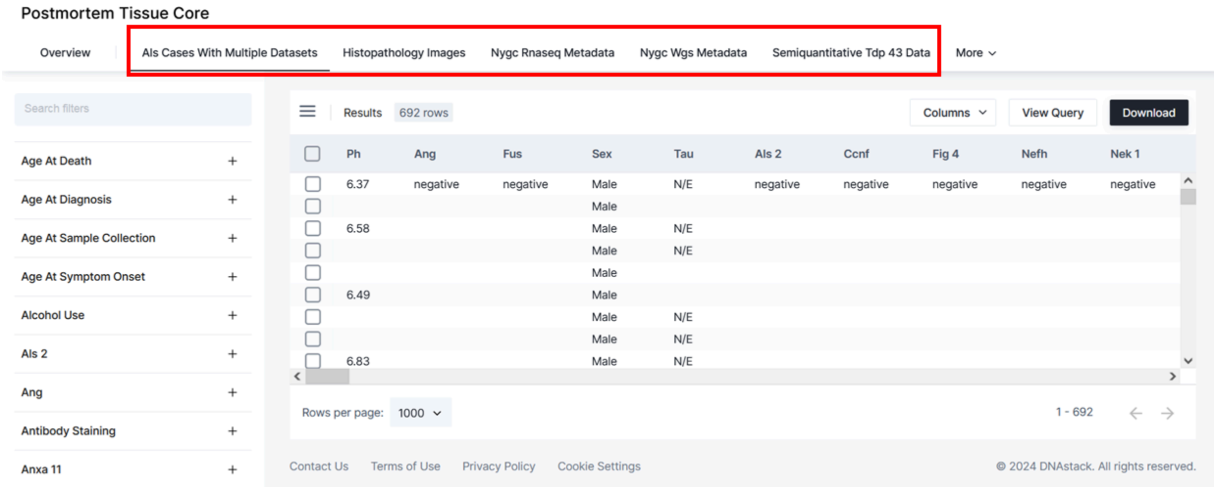
The Postmortem Tissue Core collection currently contains unique samples:

- Histopathology: 6
- RNASeq: 295
- Semiquantitative Tdp43: 139
- WGS: 296

We further provide a curated dataset of overlapping patients in 3 of 4 datasets, resulting in 692 patients (“Als Cases With Multiple Datasets”).

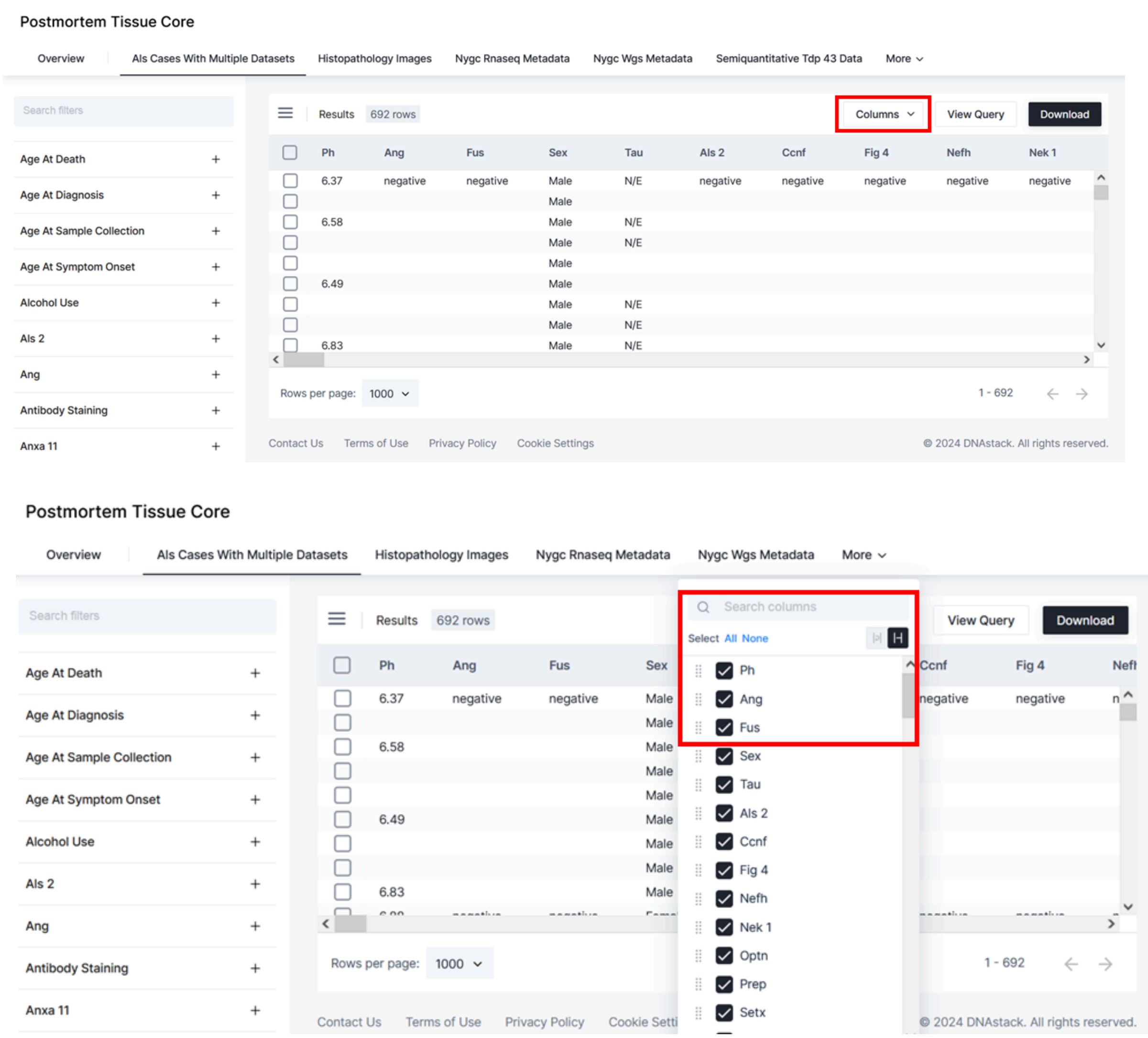
Users can click on any of the table tabs to get to the table explorer view for that table (Figure 5). Here exemplified for “Als Cases With Multiple Datasets”, users will be able to:

- Review the entire table (Figure 5)
- Rearrange or hide columns to customize the data you wish to review (Figure 6)
- Apply search filters (using the left sidebar) on any column to filter for specific subsets/cohorts (Figure 7).



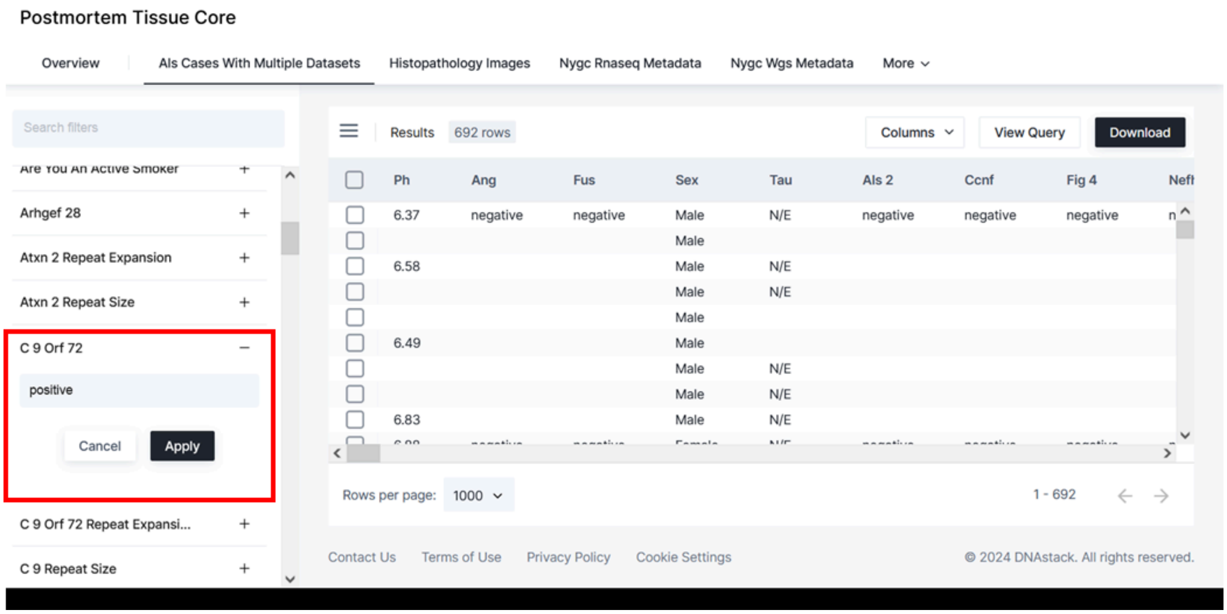
☐	Ph	Ang	Fus	Sex	Tau	Als 2	Ccnf	Fig 4	Nefh	Nek 1
☐	6.37	negative	negative	Male	N/E	negative	negative	negative	negative	negative
☐				Male						
☐	6.58			Male	N/E					
☐				Male	N/E					
☐				Male						
☐	6.49			Male						
☐				Male	N/E					
☐				Male	N/E					
☐	6.83			Male	N/E					

Figure 5 - Detailed view on the datasets



☐	Ph	Ang	Fus	Sex	Tau	Als 2	Ccnf	Fig 4	Nefh	Nek 1
☐	6.37	negative	negative	Male	N/E	negative	negative	negative	negative	negative
☐				Male						
☐	6.58			Male	N/E					
☐				Male	N/E					
☐				Male						
☐	6.49			Male						
☐				Male	N/E					
☐				Male	N/E					
☐	6.83			Male	N/E					

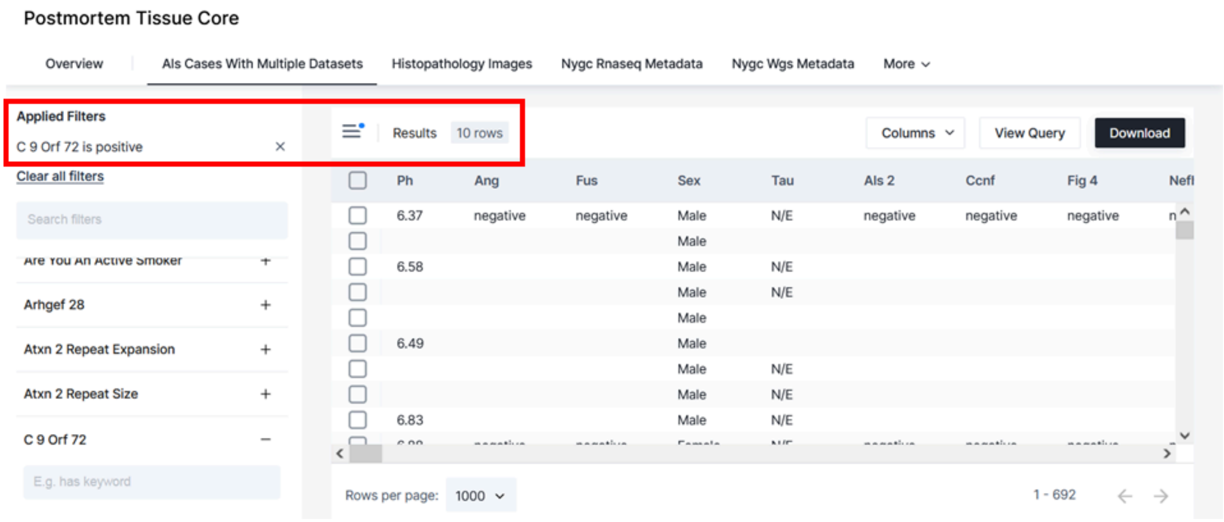
Figure 6 - Customize the columns on the datasets



☐	Ph	Ang	Fus	Sex	Tau	Als 2	Ccnf	Fig 4	Nefl
☐	6.37	negative	negative	Male	N/E	negative	negative	negative	n
☐				Male					
☐	6.58			Male	N/E				
☐				Male	N/E				
☐				Male					
☐	6.49			Male					
☐				Male	N/E				
☐				Male	N/E				
☐	6.83			Male	N/E				

Figure 7 - Customize search filters on the datasets

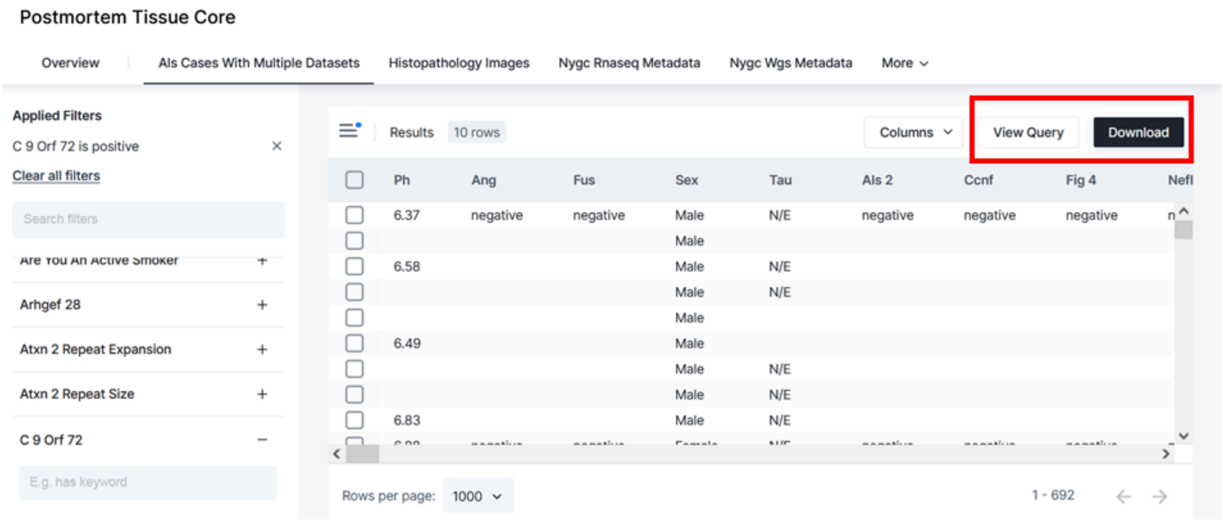
Users can apply individual search filters on the left sidebar, e.g. filtering for all patients with a mutation in C9Orf72 (Figure 7). Applying for instance "positive" in the C9Orf72 filter results in ten patients (Figure 8).



☐	Ph	Ang	Fus	Sex	Tau	Als 2	Ccnf	Fig 4	Nefl
☐	6.37	negative	negative	Male	N/E	negative	negative	negative	n
☐				Male					
☐	6.58			Male	N/E				
☐				Male	N/E				
☐				Male					
☐	6.49			Male					
☐				Male	N/E				
☐				Male	N/E				
☐	6.83			Male	N/E				

Figure 8 - Customized search filters on C9Orf7s results in 10 samples

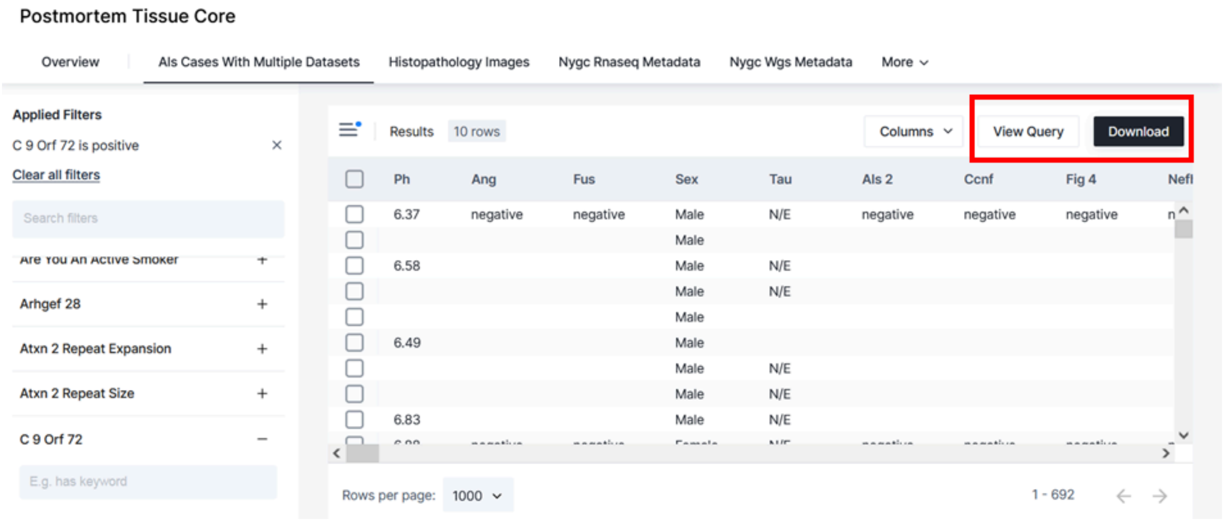
Users have the options to download the dataset (”Download” button), or transfer the dataset (”View Query” button) to the integrated Verily workbench (Figure 9).



☐	Ph	Ang	Fus	Sex	Tau	Als 2	Ccnf	Fig 4	Nefl
☐	6.37	negative	negative	Male	N/E	negative	negative	negative	n
☐				Male					
☐	6.58			Male	N/E				
☐				Male	N/E				
☐				Male					
☐	6.49			Male					
☐				Male	N/E				
☐				Male	N/E				
☐	6.83			Male	N/E				

Figure 9 - Options for further analysis of the datasets

By clicking on the ”View Query” button, users can view the command and for instance copy the python command for downstream analysis using Jupyter Notebook to the Verily workbench (Figure 10).



☐	Ph	Ang	Fus	Sex	Tau	Als 2	Ccnf	Fig 4	Nefl
☐	6.37	negative	negative	Male	N/E	negative	negative	negative	n
☐				Male					
☐	6.58			Male	N/E				
☐				Male	N/E				
☐				Male					
☐	6.49			Male					
☐				Male	N/E				
☐				Male	N/E				
☐	6.83			Male	N/E				

Figure 10 - Users can copy the command for further analysis to the Verily Workbench

Analyzing Data

Verily Workbench for conducting analysis of data collections to allow for access to large-volume data and facilitate collaboration in a shared environment.

Accessing Verily Workbench

Reference Workspace

A reference workspace has been setup within Verily Workbench to help users get started. The reference workspace can be duplicated to create a personal workspace that is connected to the Target ALS Data Portal for easy analysis using the development environment available in the personal workspace.

Verily Workbench Setup

Establishing a Personal Workspace

A personal workspace must be established to allow for analysis of data. To establish a personal workspace, select the Duplicate button from the reference workspace.

Please email workbench-support@verily.com to establish your spend profile.

Understanding Cloud Costs

Verily Workbench's Cloud Cost Management guide provides information to help you manage your cloud costs, including:

- Summary list of Workbench activities that DO and do NOT generate cloud charges
- Quick tips for managing common cloud costs
- Detailed explanation of the most common cloud costs for Workbench users

Please follow the best practices outlined in Verily’s Cloud Cost Management guide to avoid generating unwanted cloud costs.

Accessing Data Navigating Data on Verily Workbench

The data available through the Target ALS Data Portal includes raw data (e.g. – FASTQs) and curated data (processed data derived from raw data). Curated Data To review curated data, navigate to your personal workspace then select the Resources tab.

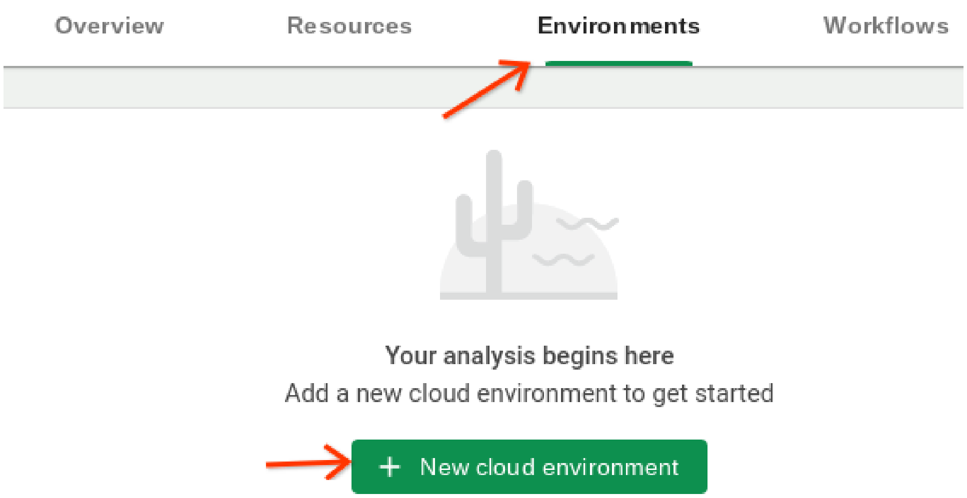
Raw Data

Raw data is available in **Requester Pays buckets**, meaning that the requestor (i.e., you) must assume the cost for data transfer when requesting raw data. Target ALS requires users to establish a Google Cloud Billing Account owned by you (i.e., the same Google Identity you use to log in to the Target ALS Data Portal) to access the raw data.

To keep costs minimal, please follow the best practices provided in Verily’s Cloud Cost Management guide.

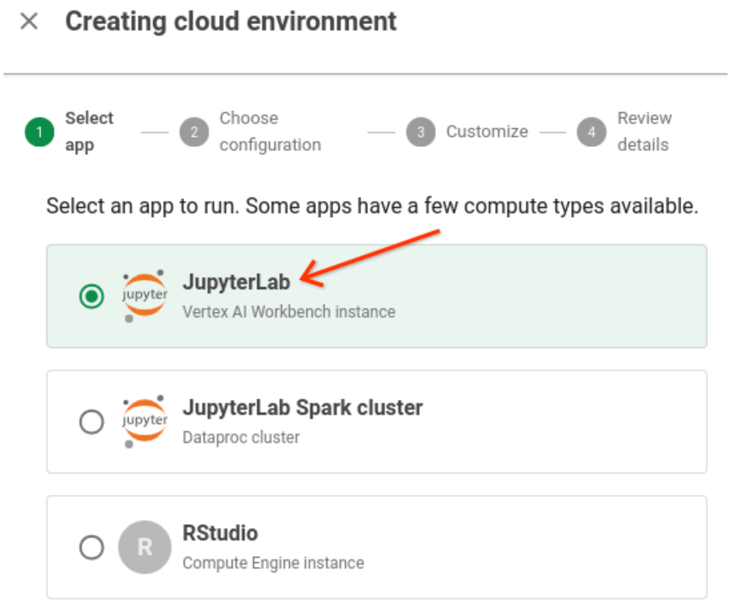
Using Cloud Environments for Analysis

Cloud Environments in Workbench consist of a virtual machine and a persistent disk, with some useful libraries and tools pre installed. They are ideal for conducting analyses using Python or R. To get started, navigate to the "Environments" tab in your duplicated workspace and create a new Cloud Environment (Figure 11). Choose the JupyterLab app and select the default configurations (Figure 12).



Your analysis begins here
Add a new cloud environment to get started

Figure 11 - Users can create their own cloud environments in Workbench to run an analysis


Creating cloud environment

1 Select app — 2 Choose configuration — 3 Customize — 4 Review details

Select an app to run. Some apps have a few compute types available.

● **JupyterLab**
Vertex AI Workbench instance

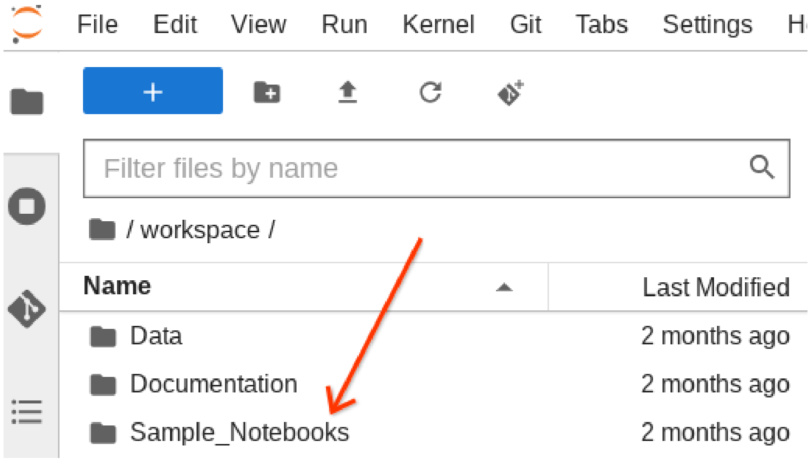
○ **JupyterLab Spark cluster**
Dataproc cluster

○ **RStudio**
Compute Engine instance

Figure 12 - JupyterLab app can run Python and R notebooks

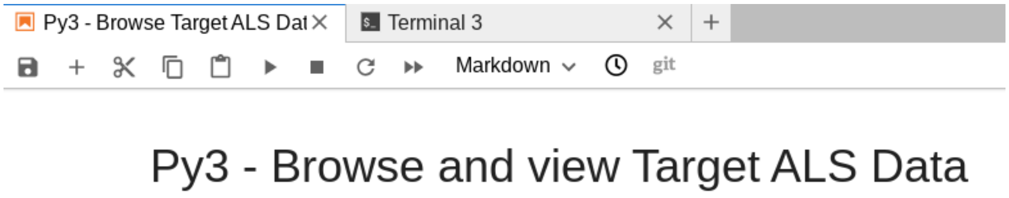
The Cloud Environment will take a few minutes to start up. Once it is running, click on the Cloud Environment to launch JupyterLab.

The JupyterLab environment is set up to automatically mount the cloud storage buckets in your workspace. This means you can use the file browser to navigate the data in the storage buckets. For example, navigate to the workspace/Sample_Notebooks directory and open one of the example notebooks there (Figure 13). They will guide you through running an example analysis in Python or R (Figure 14).



Name	Last Modified
Data	2 months ago
Documentation	2 months ago
Sample_Notebooks	2 months ago

Figure 13 - The reference workspace includes sample notebooks



Py3 - Browse and view Target ALS Data

Figure 14 -The Python version of the Browse and view Target ALS Data sample notebook

Analyzing Image Data

The Target ALS Data Portal provides image files for you to inspect and analyze. Image data can be found in the following collections and tables.

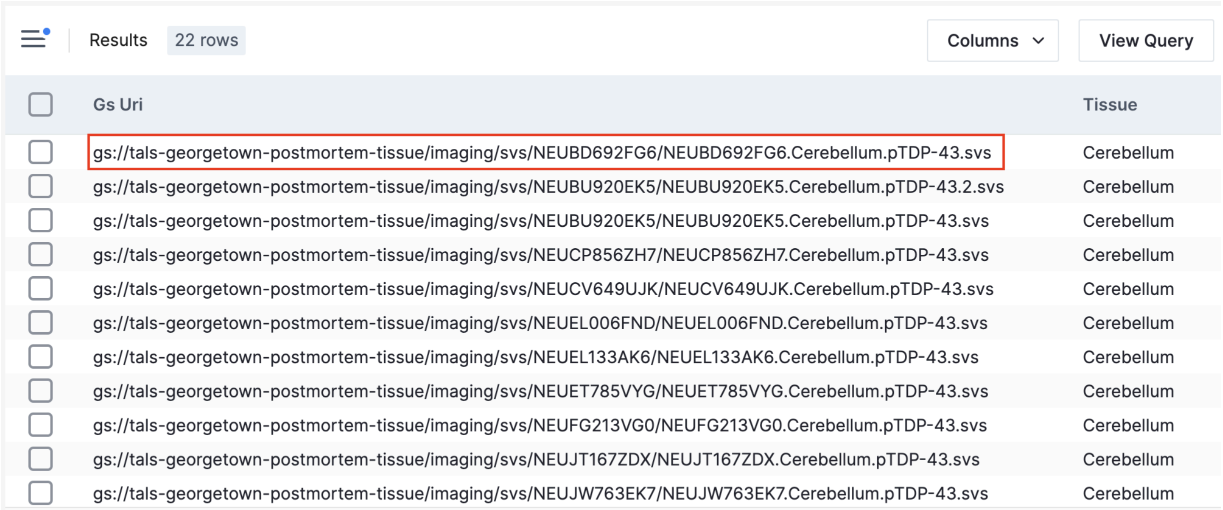
- Collection: **Postmortem Tissue Core**
    - Table: **Histopathology Images**

As these images are often in custom file formats, the appropriate software must be installed on your local machine. Image files must also be downloaded to your local machine to be viewed. Target ALS does not officially endorse any particular software, however the community has installed the following software to view images:

- For SVS (.svs) files:
    - QuPath
    - Aperio Imagescope

Downloading Image Files

The data portal hosts tables containing metadata about image files. You may create your own cohorts of patient images using the filters provided. Once you have created your cohort, you will notice that each record has a value in the Gs Uri field (Figure 15). This represents the Google Cloud Storage URI where the file is located. To download this file, you will need to download and install the Google Cloud CLI (gcloud) on your local machine.



☐	Gs Uri	Tissue
☐	gs://tals-georgetown-postmortem-tissue/imaging/svs/NEUBD692FG6/NEUBD692FG6.Cerebellum.pTDP-43.svs	Cerebellum
☐	gs://tals-georgetown-postmortem-tissue/imaging/svs/NEUBU920EK5/NEUBU920EK5.Cerebellum.pTDP-43.2.svs	Cerebellum
☐	gs://tals-georgetown-postmortem-tissue/imaging/svs/NEUBU920EK5/NEUBU920EK5.Cerebellum.pTDP-43.svs	Cerebellum
☐	gs://tals-georgetown-postmortem-tissue/imaging/svs/NEUCP856ZH7/NEUCP856ZH7.Cerebellum.pTDP-43.svs	Cerebellum
☐	gs://tals-georgetown-postmortem-tissue/imaging/svs/NEUCV649UJK/NEUCV649UJK.Cerebellum.pTDP-43.svs	Cerebellum
☐	gs://tals-georgetown-postmortem-tissue/imaging/svs/NEUEL006FND/NEUEL006FND.Cerebellum.pTDP-43.svs	Cerebellum
☐	gs://tals-georgetown-postmortem-tissue/imaging/svs/NEUEL133AK6/NEUEL133AK6.Cerebellum.pTDP-43.svs	Cerebellum
☐	gs://tals-georgetown-postmortem-tissue/imaging/svs/NEUET785VYG/NEUET785VYG.Cerebellum.pTDP-43.svs	Cerebellum
☐	gs://tals-georgetown-postmortem-tissue/imaging/svs/NEUFG213VG0/NEUFG213VG0.Cerebellum.pTDP-43.svs	Cerebellum
☐	gs://tals-georgetown-postmortem-tissue/imaging/svs/NEUJT167ZDX/NEUJT167ZDX.Cerebellum.pTDP-43.svs	Cerebellum
☐	gs://tals-georgetown-postmortem-tissue/imaging/svs/NEUJW763EK7/NEUJW763EK7.Cerebellum.pTDP-43.svs	Cerebellum

Figure 15 - Google Storage URI (GS URI) for image files in the data portal

Install the gcloud CLI, follow the [documentation and instructions]() for installing on your system. Once installed, open a terminal application and run:

Unset
```
gcloud auth login
```

A browser window should open, prompting you to login to authenticate your user identity with the gcloud CLI. You **must** login using the same identity you use for the Data Portal and Verily Workbench.

Once logged in, you can now use the `gcloud storage cp` command to copy image files. The format of the command is:

Unset
```
gcloud storage cp {source_uri} {destination_folder}
```

Where {source_uri} is the Gs Uri value from the table, and `{destination_folder}` is the local folder the file will be copied to. For example:

Unset

```
gcloud storage cp gs://tals-georgetown-postmortem-tissue/imaging/svs/NEUBD692FG6/NEUBD692FG6.Cerebellum.pTDP-43.svs .
```

You are now ready to load the file using the image viewing software you previously installed.