## Explore Overview

Through the **Explore** tab on the navigation menu, users are able to browse available data sets in their assigned study(s), as well as any publicly released datasets. The data is pulled dynamically through a database query allowing for custom interfaces for each logged in user. The sections below outline details for the three different options under **Explore**: **Tree View**, **Data Overview** and **Data Files**. There is also a section on **Limited Use Features**, specifically, the **Manage Files** and **PI Team Files** tabs, which are only accessible to users that have uploaded files or the study's Lead Principal Investigator (PI), respectively.

Note: For all options under **Explore**, only active data files are counted. Data files that have been archived are not included.
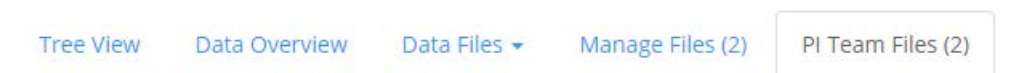
I.   Tree View

Users can explore data sets through a dynamic and comprehensive branching schematic depicting files for a study organized by file metadata (e.g., Species, Strain, Group, Experiment, Data Type and Tissue). By clicking on the nodes within the schematic, the branches will expand or collapse to reveal more or less information, respectively. When nodes are fully expanded, they appear white. When nodes are collapsed, they appear purple. The final node on a branch is the file count. The branching point labels (nodes) are dictated by the metadata selected during **Data File Upload**.

In Figure 1, one can see that the data files for the study called 'Cancer' were from *human* (**Species**) *lung* tissue *(**Tissue**).* Because these data are from human, there is only one **Strain**, denoted as *all* in the second node. The fourth nodes denote the three types of **Experiments** for the study: *mRNA_Microarray, methylation* and *miRNA_Microarray.* The next node denotes the **Data Types** for each **Experiment**: *analyzed*, *raw* and *normalized*. File counts on the last nodes indicate there is only one file of each in this study.

II.   Data Overview

Data Overview provides a tabular output of all available data for a single study by file counts. Figure 2 is an example of what a user would see. Similar to the **Tree View**, column headers align with metadata for each of the data files. Unlike the **Tree View**, files are not subdivided by the **Data Type**, only by **Experiment**. In addition, file size information is noted in brackets. This information can be very helpful when downloading files via the **Access** tab.
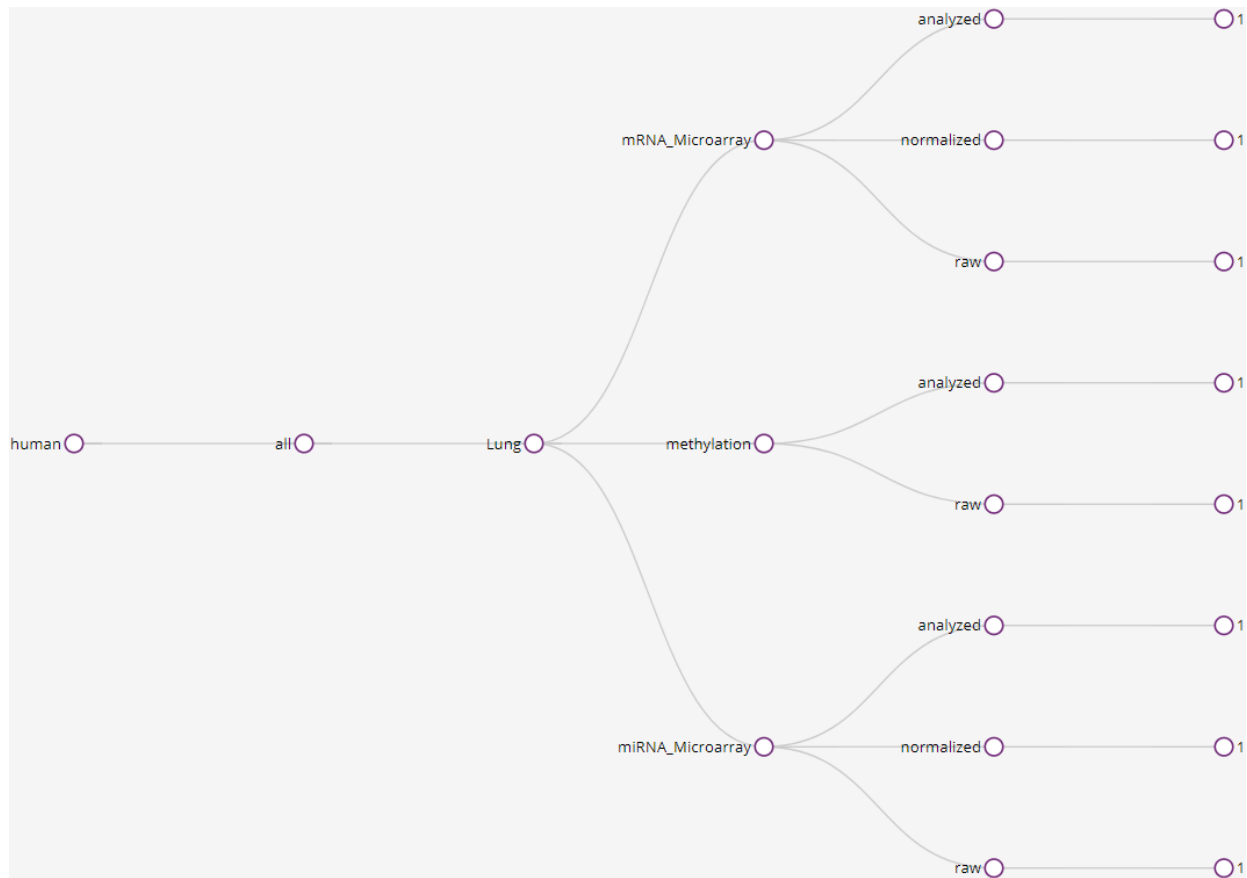
Figure 1. Tree View of the Cancer (Human) Public Data Files

STRAIN	TISSUE	METHYLATION	MIRNA MICROARRAY
all	Lung	2 files [3.84 MB]	3 files [4.68 MB]

Figure 2. Excerpt of Data Overview Matrix for Cancer (Human) Study

The following tools seen in **Data Overview** are also available in the **Data Files** and **Limited Use Features:**

- For studies with more complex data matrices, information can be resorted using the blue arrows on the right-hand side of each of the column headers.

- To change the size of the table, select from the *Show 10, 25, 50* or *100 entries* options.

- To export the data, select one of three formats, *Excel*, *CSV* or *PDF*, or simply *Copy* the information, by clicking the respective button from the table just above the matrix.

SysBioCube

III.   Data Files

Using the **Data Files** feature, a user is provided with more detailed file level browsing whereby metadata are displayed based on experimental categories. Metadata fields displayed include *Strain, Tissue,* and *Data Type, Groups.* Additional, the SysBioCube records information when a file is uploaded, which is also displayed in this matrix: *Uploaded* (i.e., the date the file was uploaded)*, File Name*, and *File Size* (See Figure 3).
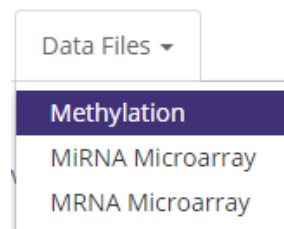
To get started select the *Experiment* of interest from the **Data Files** dropdown menu (example above on right), which is automatically populated with a list of all *Experiments* within the study.

Strain	Tissue	Data Type	Groups	Upload Date	File Name	File Size	File Actions
all	Lung	raw	Lung Cancer	2016-09-21	jhu-usc.edu_GBM.HumanMethylation27.2.lvl-1.TCGA-02-0001-01C-01D-0186-05.txt	2.04 MB	🔍📷
all	Lung	analyzed	Lung Cancer	2016-09-21	jhu-usc.edu__HumanMethylation27__TCGA-02-0001-01C-01D-0186-05__methylation_analysis.txt	1.8 MB	🔍📷

Figure 3. Data Files Matrix for Methylation Experiments from Cancer (Human) Study

 Unique to the **Data Files** feature, users are able to preview and download individual data files. The preview feature allows for visualization of the first 20 rows and all columns or first page of a file, depending on the file type. To preview, click the magnifying glass icon under *File Actions*. For those files not preview compatible, the icon is not visible. To download a file, click the camera icon and the file will download within the standard Download Folder for the internet browser in use. It should be noted that larger files or bandwidth restrictions on the user's end will impact the speed of a download.

IV.   Limited Use Features

Some registered users will see up to two additional tabs under **Explore**. These tabs are only active when a user has uploaded files to the study or if a user is the designated lead PI for a study.

A. Manage Files

The **Manage Files** tab will provide a visualization of all the files that a registered user has uploaded to the study in a table format similar to that seen under **Data Files**. The last three columns in the table will appear as they do in the image to the right: *Uploaded* (i.e., date file was uploaded and user who uploaded file), *Archived* (if file has been archived, the date the file was archived and the user who archived it will be displayed here, otherwise this cell will appear blank) and *File Actions* (icons indicated what actions can be applied to the given file). In the example above, the bottom file was archived at 0906 hours on 28 June 2018 by User, Jessica Calzola. The file can be restored by clicking the red arrow under the *File Actions* column. The first file (first row) has not been archived, so the arrow under *File Actions* remains green indicating the ability to still be archived. Overall, this tab allows the user to have fine-grained file management control.

SysBioCube

3

B. PI Team Files

The Lead PI for a study will also be able to see the **PI Team Files** tab. This tab will provide visualization of <u>all</u> of the files that have been uploaded to the study by <u>all</u> users under the purview of a Lead PI in a table with a format identical to that of the **Manage Files** tab. Similar to the **Manage Files** tab, the last three columns in the table will provide the Lead PI with fine-grained file management control to archive and restore files. Overall, these features ensure that the onus to maintain the integrity of the study data is on the Lead PI throughout the duration of the study.