

ADBio

Active Data Biology

Instruction manual

By

Ryan Wilson

Table of Contents

Navigation.....	- 2 -
My Projects Page	- 5 -
Project Home Page	- 7 -
Heat Map View	- 9 -
Statistical tests popup.....	- 10 -
Enrichment tests.....	- 12 -
Pathway View	- 14 -
Canvas View.....	- 16 -
Create New Project.....	- 18 -

Navigation



1. Navigation bar
2. Home page link
3. Current project information
4. Visual Apps
5. User account / settings
6. Help/about

1. Navigation bar

The navigation bar is a fixed area on top of the screen used to navigate the website. You can click the different buttons and links to easily move around on ADBio.

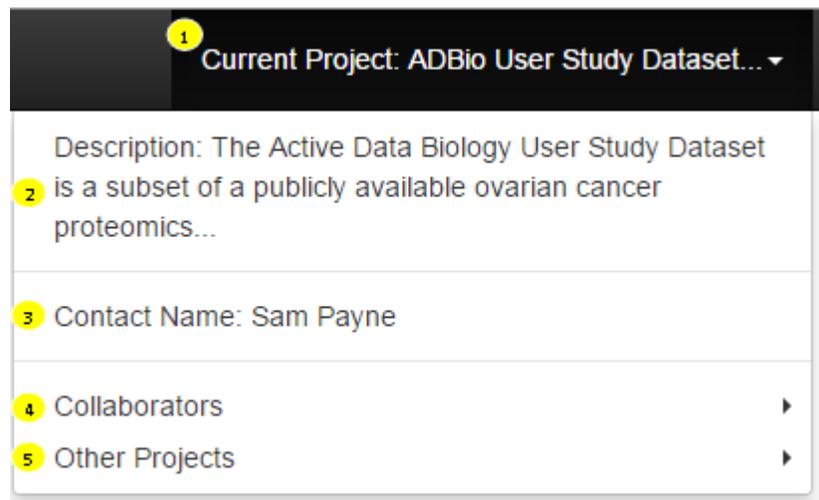
2. Home page link

The branding area is a link that moves you back to the index page, also helps you remember what website you're on.

3. Current project information

3.1. Clicking on the current project information link opens this dropdown.

3.2. This is a small description of the current project. Hover over this to show the full description in a tool tip. Clicking will take you to the project home page.



- 3.3. This is the contact for the current project. Clicking this will attempt to open an email service to create an email to them.
- 3.4. This is a list of the collaborators on this current project.
- 3.5. This is a list the other projects that you are a collaborator on; clicking one of them will load that project.

4. Visual Apps

Heat map view button

Clicking will open a new tab on the browser to the heat map view of that project.

Pathway view button

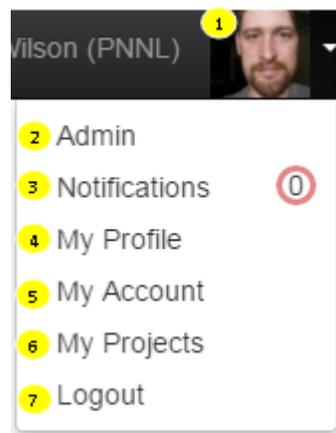
Clicking will open a popup that lists the pathways for this project. Clicking on one of the Kegg pathway will open a new tab on the browser to the pathway view with this projects data.

Canvas view button

Clicking will open a new tab on the browser to the canvas view of that project.

5. User account / settings

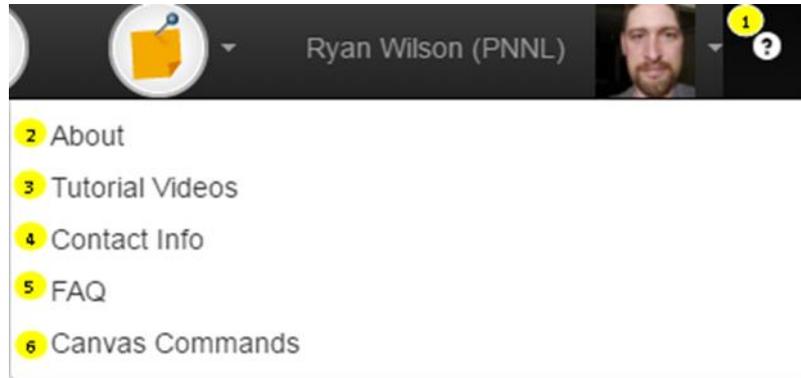
- 5.1. Clicking will open a dropdown menu with different settings
- 5.2. This is a link to the admin page. If you don't have admin access then this will not show up.
- 5.3. Notifications, clicking will open a view to manage your notifications
- 5.4. My Profile, clicking will open a popup allowing you to add or edit your information that is saved on the ADBio server. The login for you profile cannot be changed.
- 5.5. My Account, clicking will do the same thing as 6.
- 5.6. My projects, clicking will take you back to the "my projects" page.
- 5.7. Logout, clicking will log you out of ADBio. This doesn't log you out of Github.com so that next time you visit ADBio you will not have to use your user name and password. To log out completely, log out of ADBio then go to Github.com and sign out from there. The next time you want to visit ADBio you will need to enter your user name and password.



6. Help/about

- 6.1. Help button, clicking will open a dropdown menu for help.
- 6.2. About, clicking will open a popup with information about this website and program.
- 6.3. Tutorial Videos, clicking this will open a popup with a few videos you can watch to help you get familiar with ADBio.

6.4. Contact Info, clicking this will open a popup that shows information about PNNL and contact information to contact the creators and project lead.



6.5. Frequently Asked Questions (FAQ). This

is the FAQ page it is a resource that allows you to quickly find answers to questions you might have about ADBio.

6.6. Canvas Commands, while in the Canvas view this option will be available. Clicking this will load a popup with all the commands and keyboard shortcuts that are used on the canvas.

My Projects Page

1 Show 10 entries 2 NEW 3 IMPORT 4 OPEN 5  6 Search:

Project Title	Repository	Owner	Description	Collaborators
7 ADBio User Study Dataset	ADB-User-Study	PNNL-Comp-Mass-Spec	The Active Data Biology User Study Dataset is a subset of a publicly available ovarian cancer proteomics...	Ryan Wilson Sam Payne Joon-Yong Lee
7 Transcriptome and Proteomics of a Light-Dark Synchronized Bacterial Cell Cycle	Prochlorococcus	ActiveDataBio	This data is taken from a publication by Jacob Waldbauer from the Chisholm group at MIT. (http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0043432)....	Ryan Wilson Sam Payne Joon-Yong Lee

8 Showing 1 to 2 of 2 entries 9

1. Length
2. Create New
3. Import From Github
4. Open Project
5. Share Project
6. Search
7. Project Information
8. Table Information
9. Pagination

The first page after logging in is the “My Projects Page”. Here is a table view of all the projects that you are a collaborator on.

1. Length

The length element of the table changes how many rows are visible per page.

2. Create New

The NEW button is how users create a new project with their own dataset in a repository on Github.com. This requires some files to be created by the user before the process of creating a new project can be completed. You can find more information about creating a new project in the “create new project” section.

3. Import From Github

The IMPORT button is how users create a new project with all the required files already in a Github repository. You can find more information about this in the “create new project” section.

4. Open Project

The OPEN button will load the index page for the selected project. Double clicking on the project will also do this operation.

5. Share Project

The Share button allows the admin of a repository to search for a Github user name and add them as a collaborator to the selected project.

6. Search

This allows the user to search the table for the inputted keywords.

7. Project Information

Each row in the table is project information that displays project name, repository, owner, description, and a few collaborators. These rows can be sorted by clicking the different column icon in the header. Owner and repository is how Github.com organizes all the files that they store. Owner can be a user or an organization. Repository is the name given as the project folder. In a similar setup as a computer hard drive “c:” each user/organization will have a folder “c:\user”, “c:\organization” and each project will have a folder “c:\user\project1”, “c:\user\project2”.

8. Table Information

This show how many rows are showing out of a total number of rows.

9. Pagination

If the length option is set so there are more rows than what can be shown the pagination will allow you to load more pages.

Project Home Page

1 ADBio User Study Dataset

2 Description

The Active Data Biology User Study Dataset is a subset of a publicly available ovarian cancer proteomics experiment. This dataset contains a proteomic survey of 174 tumors and provides quantitative abundance measurement for 3589 proteins. Proteins are cross-annotated by KEGG and can be view in the Pathway viewer. The tumors are annotated with a variety of clinical information including treatment and outcome. The patients are anonymized and shown only as a randomized identifier.

3 Contact Information



Sam Payne

4 Other Collaborators(27)

  Ryan Wilson	  TomMetz
  Sam Payne	  cevansmolina
  Joon-Yong Lee	  rmimira

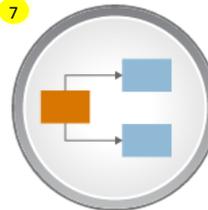
6



HeatMap Viewer

[Go to heatmap »](#)

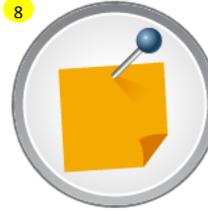
7



Pathway Viewer

[Go to pathway »](#)

8



Canvas Viewer

[Go to canvas »](#)

1. Title
2. Description
3. Contact Information
4. Other Collaborators
5. Share
6. Heat Map View
7. Pathway View
8. Canvas View

1. Title

This is the title of the project.

2. Description

This is the full description provide for this project.

3. Contact Information

This is who to contact about this project, clicking the email icon will attempt to create a new email to this person. (Not every computer has the email setup for use with the browser)

4. Other Collaborators

This is a list of all the other collaborators and a total number of them. Clicking on the email icon next will attempt to create a new email to this person. (Not every computer has the email setup for use with the browser)

5. Share

The Share button will allow admins of the project to search Github.com for a user name and add them as a collaborator to this project.

6. Heat Map View

Clicking the image or the button will open the heat map view for this project.

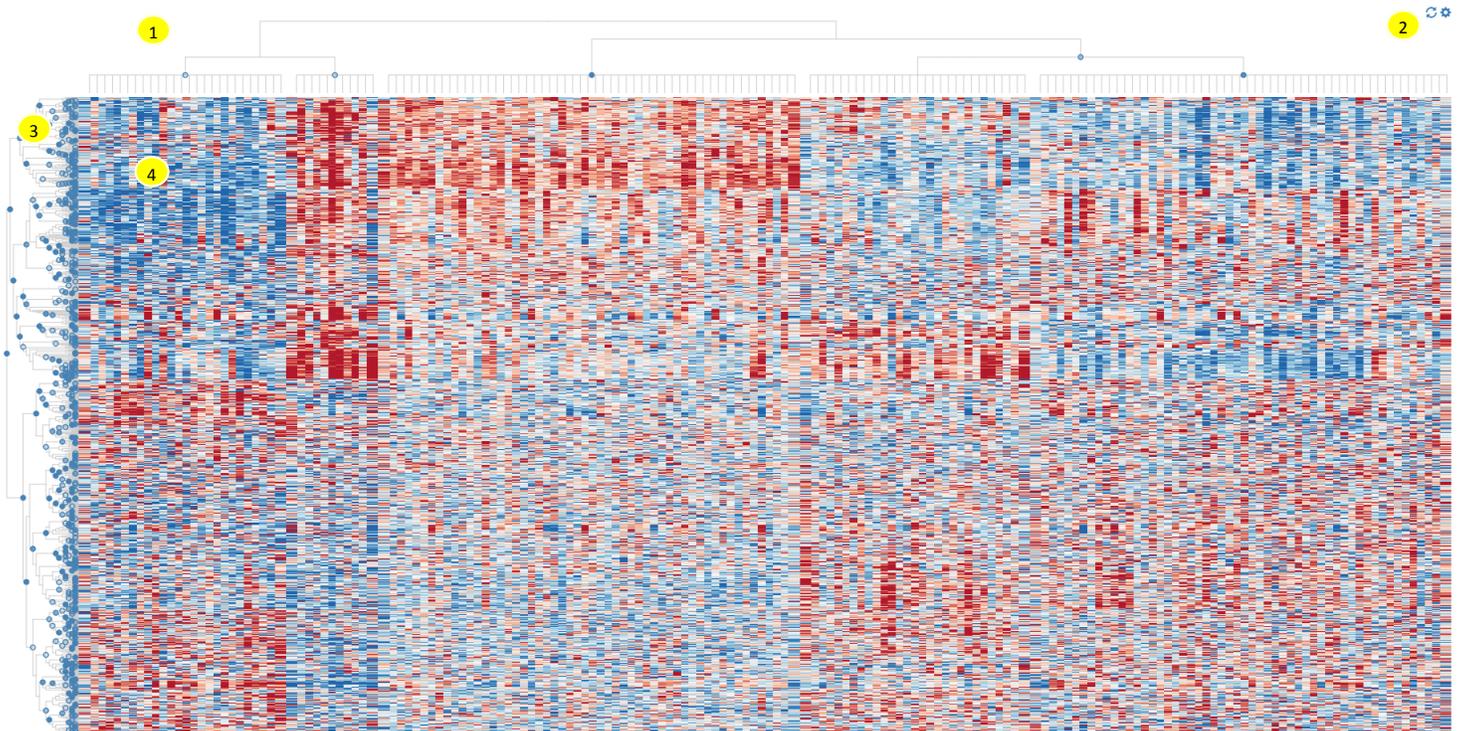
7. Pathway View

Clicking the image or the button will open a popup with a list of all the pathways associated with this project, clicking on a Kegg pathway will open a new tab on the browser to the pathway view with this projects data.

8. Canvas View

Clicking the image or the button will open the canvas view for this project.

Heat Map View



1. Sample dendrogram
2. Refresh/Settings
3. Protein dendrogram
4. Heat Map

1. Sample dendrogram

This is the dendrogram for the samples collected in the data set. The blue circles are nodes of the grouping. Clicking on one of them opens a popup with statistical tests for the data.

2. Refresh/Setting

The refresh button will reset the heat map to its default zoom ratio. The settings button allows the user to change the color of the heat map.

3. Protein dendrogram

This is the dendrogram for the proteins listed in this dataset. Clicking on a node opens a popup that shows enrichment tests for a group of proteins.

4. Heat Map

This is the heat map of the dataset for the project. This heat map is showing protein abundance with overabundance in red and under abundance in blue. Clicking and dragging

on the heat map will draw a highlight square that will zoom the heat map to that box after you release the button. Double clicking on the heat map will reset the heat map to the default zoom level.

Statistical tests popup

Metadata	Description	p-value
tumor_residual_disease	tumor residual disease	8.57e-4

1. Name
2. Pin to canvas button
3. Window controls
4. Data Tabs

1. Name

This is the name of that blue circle we call a node. You can change this allowing you to make the nodes have a better readability.

2. Pin to canvas button

The pin to canvas button is used to pin this statistical test to the canvas. It will use the name as the label on the pin.

3. Window controls

These are basic window controls that allow you to collapse, minimize, maximize, and close the popup, respectively.

4. Data Tabs

These are different ways to view the metadata. The Table tab shows the metadata test attribute, description, and p-value for that set of the metadata (Figure 1). The list tab shows the sample id, and description for this group (Figure 2). The Graph tab allows you to select

an attribute and see a comparison of data between samples inside the selected group and outside this group (Figure 3). The Info tab shows some information about this group and a button to download the metadata file.

Metadata	Description	p-value
tumor_residual_disease	tumor residual disease	8.57e-4
PlatinumFreeInterval	This is negative if the patient had received treatment after the date of progression/recurrence or if the date of last platinum treatment is after the date of last follow-up.	1.23e-3
Normal_Control	Normal Control	4.83e-3
daystotumorprogorrecur	days to tumor progression or recurrence	7.61e-3
ProgorRecur	progression or recurrence	7.77e-3
TSS	Tissue source site	1.55e-2
tissue_source_site	tissue source site	1.55e-2
vital_status	vital status	2.39e-2
race	race	1.51e-1
age_at_initial_pathologic_diagnosis	age at initial pathologic diagnosis	1.93e-1

Showing 1 to 10 of 30 entries

Previous 1 2 3 Next

Figure 1

Show 10 entries Search:

Name	Description
PID_00167	C
PID_00166	C
PID_00165	C
PID_00164	C
PID_00157	C
PID_00154	C
PID_00150	C
PID_00147	C
PID_00145	C
PID_00140	C

Showing 1 to 10 of 54 entries

Previous 1 2 3 4 5 6 Next

Figure 2

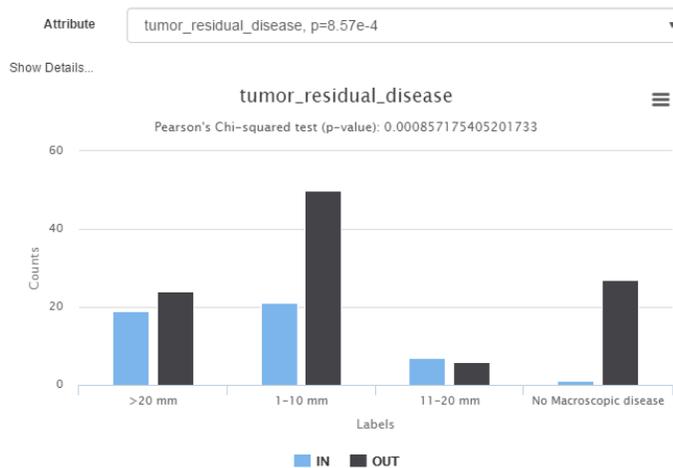
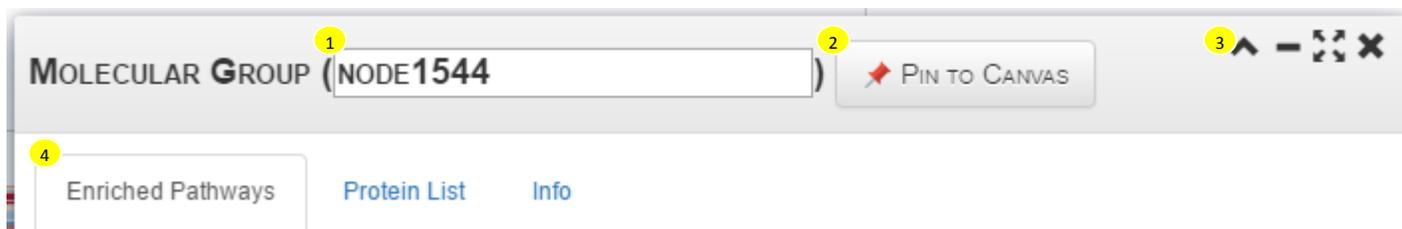


Figure 3

Enrichment tests



1. Name
2. Pin to canvas button
3. Window controls
4. Data Tabs

1. Name

This is the name of that blue circle we call a node. You can change this allowing you to make the nodes have a better readability.

2. Pin to canvas button

The pin to canvas button is used to pin this statistical test to the canvas. It will use the name as the label on the pin.

3. Window controls

These are basic window controls that allow you to collapse, minimize, maximize, and close the popup, respectively.

4. Data Tabs

These are different ways to view the protein data. The Enriched pathways tab shows a list of pathway names, an adjusted p-value, and a count of the number of proteins in that pathway (Figure 4). Clicking the red pin button pins that pathway to the canvas. Clicking on a Kegg pathway name will open the pathway view. The protein list tab shows a list of proteins in that group, with gene, chromosome, and location for each of them (Figure 5). Clicking on the red pin button pins that protein to the canvas. The Info tab shows some information about this group and a button to download the raw data file.

Pathway Name	adj. p-value	count
REACTOME_INTERFERON_SIGNALING Genes involved in Interferon Signaling	7.73e-19	36
REACTOME_INTERFERON_GAMMA_SIGNALING Genes involved in Interferon gamma signaling	2.27e-18	24
REACTOME_IMMUNE_SYSTEM Genes involved in Immune System	3.46e-15	68
REACTOME_CYTOKINE_SIGNALING_IN_IMMUNE_SYSTEM Genes involved in Cytokine Signaling in Immune system	3.46e-15	38
REACTOME_INTERFERON_ALPHA_BETA_SIGNALING Genes involved in Interferon alpha/beta signaling	6.02e-14	17
KEGG_ANTIGEN_PROCESSING_AND_PRESENTATION Antigen processing and presentation	2.28e-12	20

Showing 1 to 10 of 593 entries

Previous **1** 2 3 4 5 ... 60 Next

Figure 4

show 10 entries Search:

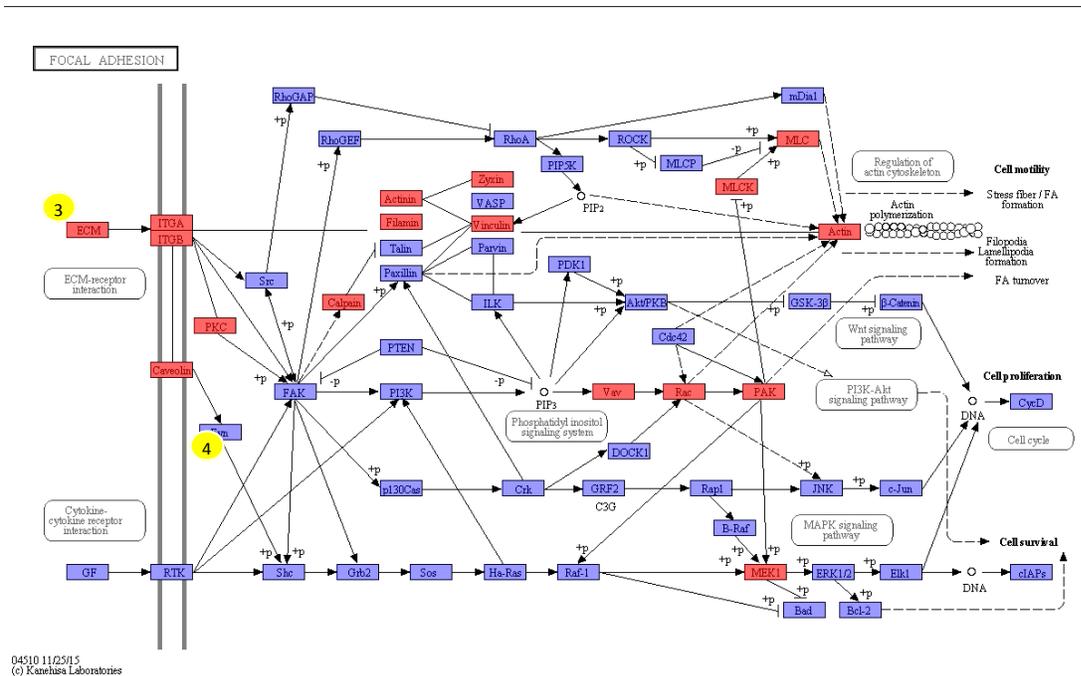
Protein	Gene	Chromosome	Location
NP_981961	CD163	12	7623409 - 7656489
NP_976319	RNH1	11	494512 - 507300
NP_945146	PYHIN1	1	158900586 - 158946844
NP_870986	PKLR	1	155259630 - 155271225
NP_861970	NF2	22	29999545 - 30094587
NP_839943	IQGAP3	1	156495197 - 156542396
NP_808592	ASAH1	8	17913934 - 17942494
NP_776049	WARS	14	100800125 - 100843142

Showing 1 to 10 of 250 entries

Previous **1** 2 3 4 5 ... 25 Next

Figure 5

Pathway View



1. Pathway name
2. Pathway View
3. Gene
4. Pathway

1. Pathway name

This is the pathway name, clicking on the information icon displays a description of the pathway.

2. Pathway View

2.1. The pathway view

button changes when in this view clicking it will open a new drop down menu.

2.2. View in heat map,

clicking on this opens a new tab on the browser taking you to

t: ADBio User Study Dataset... 1

- 2 View in Heatmap
- 3 View on Kegg
- 4 Pathway Description
- 5 Pin to Canvas
- 6 Pathway Legend
- 7 Pathway Chooser

the heat map.

2.3. View on Kegg, clicking this opens a new tab on the browser taking you to Kegg.com to view the pathway there.

2.4. Pathway description, clicking this opens a popup with the description of the pathway in the viewer.

2.5. Pin to canvas, clicking this will pin this pathway to the canvas.

2.6. Pathway legend, clicking this will open a popup with a legend of all the elements of a pathway.

2.7. Pathway chooser, clicking this opens a popup with a list of pathways for this project.

Clicking on a Kegg pathway will open a new tab on the browser loading that pathway in a new viewer.

3. Gene

The small rectangles are gene product. Red means the dataset includes proteins that are associated with this gene. Clicking this will open a popup with a graph that shows the abundance of the proteins in that gene for each group (Figure 6). Clicking the name in the popup will take you to Kegg.com viewing that protein. Clicking the pin button will pin this protein or group of proteins to the canvas. Blue means the dataset doesn't have any proteins associated with that gene. Clicking this will open a popup with a list of proteins for that gene. Clicking on their name takes you to Kegg.com to view that protein. Clicking the pin button will pin that protein to the canvas.

4. Pathway

The large rectangle is another pathway. Clicking this will open a new tab on the browser loading a new pathway viewer with that pathway.

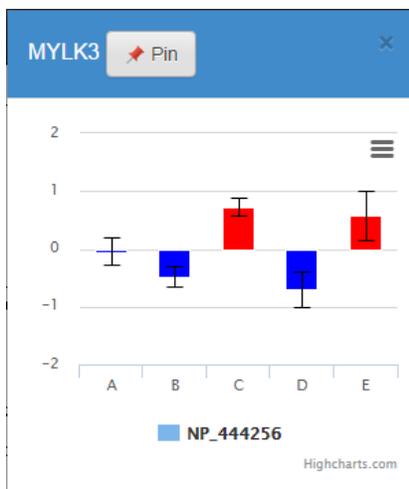
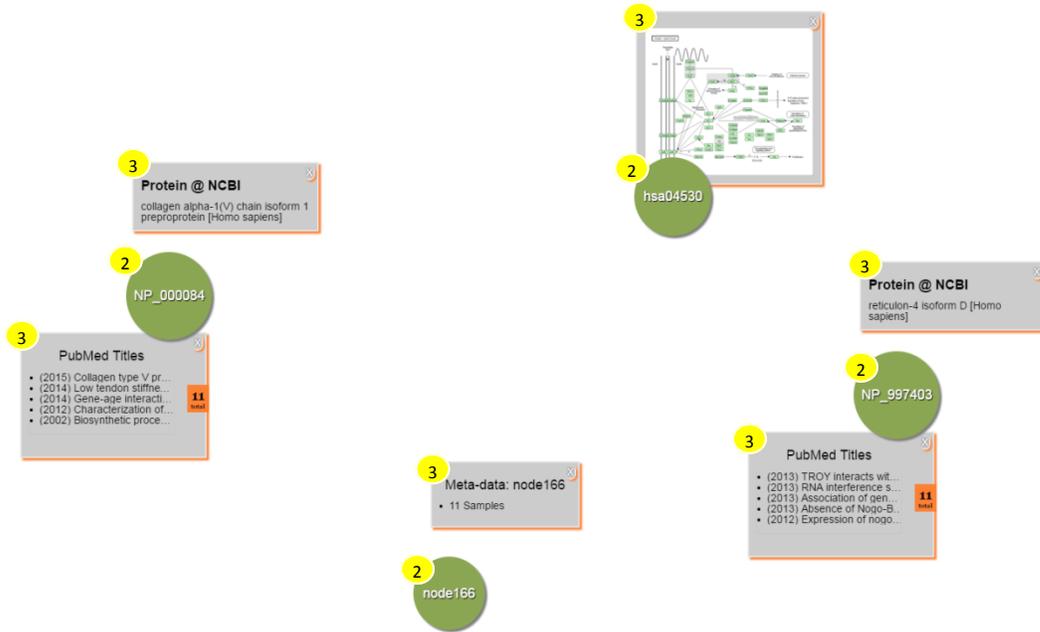


Figure 6

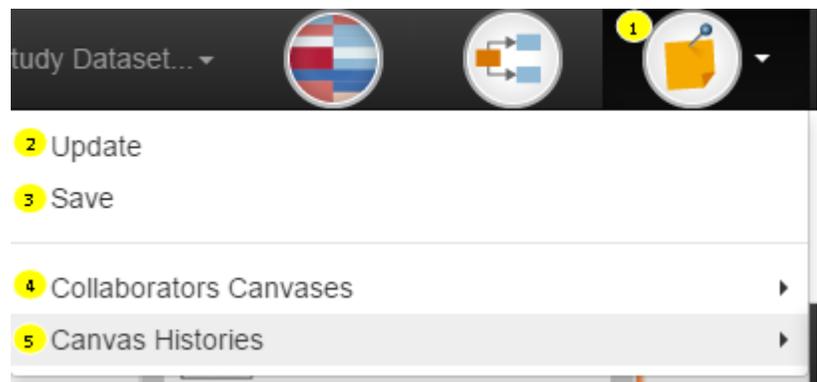
Canvas View



1. Canvas View
2. Pin
3. Recommender Card

1. Canvas View

- 1.1. Clicking the Canvas view button opens this dropdown menu.
- 1.2. This menu button updates the canvas to what is on Github.com
- 1.3. This menu button saves the canvas to Github.com
- 1.4. This menu button is used to load other collaborator canvases. Clicking will open another dropdown menu with a list of collaborators. Then clicking on their name will load their canvas. From there you can view, or while holding 'C' then clicking on a pin it will copy that pin to your canvas.
- 1.5. Canvas histories



This is the Canvas histories, a list of all the times that your canvas was saved.

1.5.1. This is the commit message sent when the canvas was saved and a time from now that it was committed. Hovering over the time will display a date and time the event took place. Clicking this area will load that canvas just like the collaborator canvas. Viewing that previous canvas allowing you to use the 'C' + click to copy that pin to your current canvas.

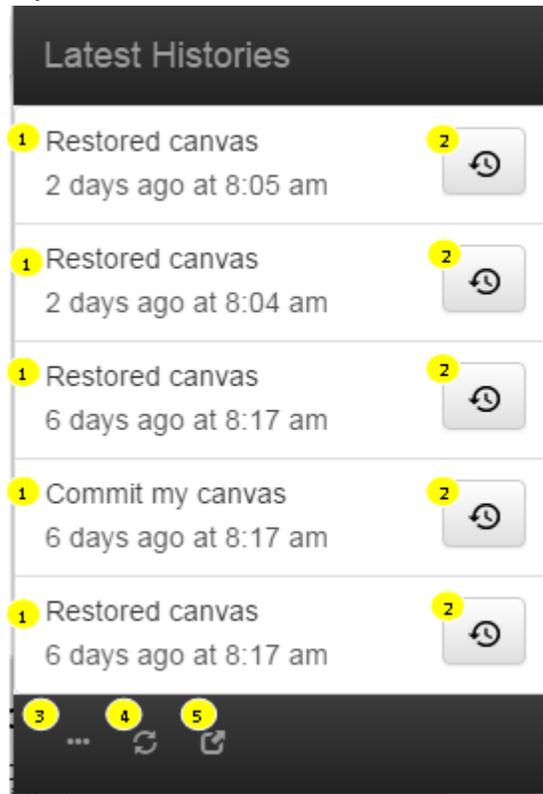
1.5.2. The restore to previous button. Clicking this button will make that previous canvas your current and remove what was there. Committing this restored canvas to Github.com. You never lose the history so you can always restore the previous canvas before you pressed the restore button by finding the correct canvas history and clicking the restore button.

1.5.3. This is the load more button. Clicking will add 15 more entries to the canvas histories.

1.5.4. This is a refresh button. Clicking this will reload the canvas histories.

1.5.5. This is the go to Github.com button. Clicking this button will load Github.com showing you all the commit messages that we used to fill the canvas histories.

#Note while viewing the collaborator canvas or a previous canvas the restore to previous canvas button is disabled.



2. Pin

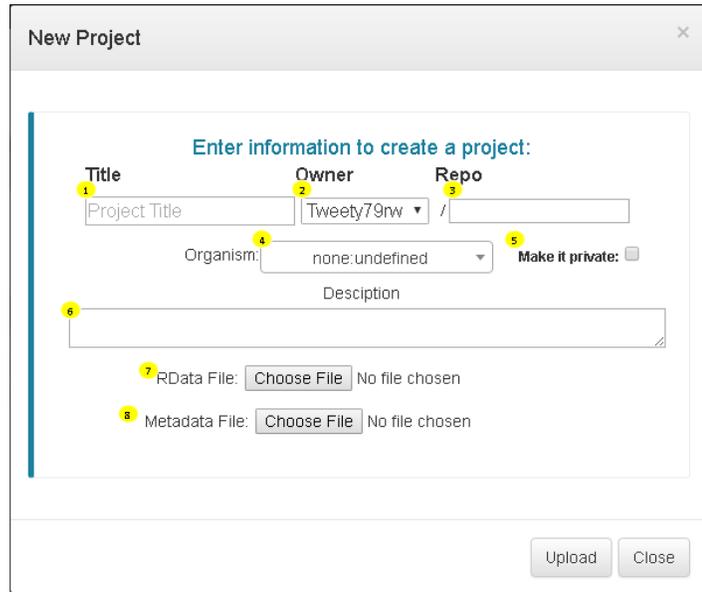
This is called a pin. When you click pin to canvas from one of the other views this is what gets put on the canvas. It can be moved around by clicking and dragging it to a new location. There are other controls for the Pin, holding delete and clicking on the pin will delete the pin from the canvas.

3. Recommender Card

This is a recommender card. The canvas automatically loads these based on the pin on the canvas. The 2 recommender cards for the pin on the left are a list of PubMed titles, and some details about the pin. Different recommenders have different functionality, but all of them have a button in the lower right corner that allows them to be flipped over showing more information and download links to the raw data, or to more information.

Create New Project

1. Title, chose a title for the project.
2. Owner, this dropdown is a folder type structure that Github.com uses to organize files saved on their website. If you have admin permissions for an organization you can use select it by clicking the dropdown menu and selecting it.
3. Repo, this is the name of the folder inside the owner folder where all the files for the project will be stored on Github.com.
4. Organism, this is used to classify the dataset you are uploading.
5. The private checkbox makes the repository private. You can only make a private repository if you have private repositories available on Github.com.
6. Description chose a description for the project.
7. Choose Rdata File, clicking the button will open a file browser so you can select the Rdata file for the project.
8. Choose metadata file, clicking the button will open a file browser so you can select the metadata file for the project.



The screenshot shows a 'New Project' dialog box with the following fields and controls:

- Title:** A text input field with a yellow callout '1'.
- Owner:** A dropdown menu showing 'Tweety79nw' with a yellow callout '2'.
- Repo:** A text input field with a yellow callout '3'.
- Organism:** A dropdown menu showing 'none:undefined' with a yellow callout '4'.
- Make it private:** A checkbox with a yellow callout '5'.
- Description:** A large text area with a yellow callout '6'.
- RData File:** A 'Choose File' button with a yellow callout '7' and the text 'No file chosen'.
- Metadata File:** A 'Choose File' button with a yellow callout '8' and the text 'No file chosen'.

At the bottom right, there are 'Upload' and 'Close' buttons.

To create a new project or import an existing project you must fill in the form then click the upload button. The importing of a project assumes you already made a repository on Github.com and uploaded the required files to it. With properly formatted files the rest of the process is automatic. Use the information below to make sure your files are formatted properly.

Data Import for the Active Data Biology

The Active Data Biology is a tool for you (and collaborators) to view and explore your data. Getting your data uploaded is the first step. For each project, there are two primary types of required data: a quantitative data matrix and the meta-data.

3.1. Data matrix

In the Heat map view, we display the data matrix in the center and dendrograms to show row and column hierarchical clustering [1, 2]. Hence, to displaying heat map with your dataset, we need the clustering information as well as the data matrix.

The data matrix file contains the quantitative data to be displayed in the heat map. The underlying data can be proteomics, transcriptomics, metabolomics, or a mix of data types. To import the data, we require that it be formatted as an R object (using an .Rdata file), which is necessary to preserve the clustering that is commonly done to large datasets. Clustering is an essential part of data analysis and can be quite time consuming. We want to preserve and accurately display that effort. See <HERE> for an example R object file. If you need help creating an appropriate file, see <HERE> which starts with a simple spreadsheet and uses the online R resource R-fiddle. You can use [this](#) RScript to convert a csv file of your data to the Rdata file.

How to generate R object

I assume that you have a simple spreadsheet as a CSV file, which contains the expression/abundance level of genes/proteins for each sample across columns as below.

	A	B	C	D	E
1		PA9.79EE.001A01	PA9.79EE.002A02	PA9.79EE.003A03	PA9.79EE.0
2	NP_001387	1.277257855	0.437513838	-0.284448153	0.0018
3	NP_004577	0.124207971	-0.844039525	-0.605290267	-0.661
4	NP_001002032	1.141099288	1.082154503	0.69950289	0.6008
5	NP_001269119	-1.913132336	-1.11390171	0.141604532	-0.0221
6	NP_001070729	-2.900709365	-0.157302225	0.47068991	0.5886
7	KIAA0701	2.748201889	0.663846765	2.225995505	-0.0979
8	NP_001073308	2.171828951	0.690946864	0.431951666	0.508
9	NP_001248745	2.602658406	2.072816847	1.731488599	1.3294
10	NP_001012768	-1.520338046	-1.659507695	-1.167639288	-0.5609
11	NP_001229343	-0.696836141	0.779523689	0.203489511	0.3569
12	NP_001127836	-2.073977387	-0.454482165	-0.046639682	0.2301
13	NP_001139019	1.629996207	0.788366506	0.212246733	0.1011

Column names should describe the unique identifications for samples. These ID must be identical to the sample ID containing in a meta-data file.

Row names must be annotated to RefSeq, or Uniprot ID corresponding to genes/proteins. These IDs are associated to pathway analysis and visualization.

Names must be syntactically valid. A syntactically valid name consists of letters, numbers and the dot or underline characters and starts with a letter or the dot not followed by a number. (See [check.names](#) option in 'read.csv' function in R)

Run R-fiddle (<http://www.r-fiddle.org/#/fiddle?id=FK38NvqU>)

This example looks quite straightforward. If you run this code in R-Fiddle, you can see the two plots. The first one describes the initial dendrogram for samples, which is randomly generated. And the second plot shows the different dendrogram after grouping through the sample numbers.

RData: it contains an R object named 'hc.out', which has a list class as below. It should have the same format to the output of a 'heatmap.2' function in R. However, if you are familiar with R,

you can customize the dendrograms containing the designated grouping instead of the current one. In other words, you can replace the rowDendrogram/colDendrogram into a new dendrogram via 'hclust' function with different manners.

```

Values
hc.out                               List of 10
rowInd : int [1:85] 60 17 25 13 31 58 11 5 76 4 ...
colInd : int [1:87] 37 38 33 36 85 84 83 87 82 86 ...
call : language heatmap.2(x = mat_data)
carpet : num [1:87, 1:85] 0.557 -0.333 -3.954 -5.456 2.711 ...
..- attr(*, "dimnames")=List of 2
.. ..$ : chr [1:87] "PA9.79EE.073G01" "PA9.79EE.075G03" "PA9.79EE.063F03" "PA9.79EE.066F06" ...
.. ..$ : chr [1:85] "NP_001019837" "NP_001157016" "NP_036442" "NP_839943" ...
rowDendrogram: ..--[dendrogram w/ 2 branches and 85 members at h = 34.9, midpoint = 53, value = 0.241]
.. |--[dendrogram w/ 2 branches and 80 members at h = 31, midpoint = 23.9, value = -1.14]
.. | |--[dendrogram w/ 2 branches and 7 members at h = 22.7, midpoint = 3.69, value = -0.841]
.. | | |--[dendrogram w/ 2 branches and 4 members at h = 16.4, midpoint = 2.12, value = -0.562]
.. | | | |--[dendrogram w/ 2 branches and 3 members at h = 11.2, midpoint = 1.25, value = -0.627]
.. | | | | |--[dendrogram w/ 2 branches and 2 members at h = 8.1, midpoint = 0.5, value = -0.584]
.. | | | | | |--leaf "NP_001019837" ( value.NP_001019837 = -0.37 )
.. | | | | | | |--leaf "NP_001157016" ( value.NP_001157016 = -0.712 )

```

Editing for clustering

If you want to dynamically change the clustering information, you can also change the dendrogram files uploaded in GitHub. They have a JSON-formatted tree structure as shown in the figure so that it is easy to understand and edit.

```

{
  "name": "node86",
  "children": [
    {
      "name": "node85",
      "children": [
        {
          "name": "node78",
          "children": [
            {
              "name": "PA9.79EE.073G01"
            },
            {
              "name": "node76",
              "children": [
                {
                  "name": "PA9.79EE.075G03"
                },
                {
                  "name": "PA9.79EE.063F03"
                }
              ]
            }
          ]
        }
      ]
    }
  ]
}

```

3.2. Meta-data

The meta data is used to store a wide variety of information about the sample or experiments. This file should be formatted as CSV file. The type of data tracked is variable for each project. The most important information is a sample identifier that matches the samples in the data matrix.

References

1. Wilkinson, Leland, and Michael Friendly. "The history of the cluster heat map." *The American Statistician* 63.2 (2009).
2. Gehlenborg, Nils, Janko Dietzsch, and Kay Nieselt. "A framework for visualization of microarray data and integrated meta information." *Information Visualization* 4.3 (2005): 164-175.