

Pathway Association Enrichment Analysis



Documentation

Supported and Powered by



مؤسسة مستشفى سرطان
الأطفال - مصر
Children's Cancer Hospital
Foundation - Egypt



جمعية أصدقاء المبادرة
القومية ضد السرطان
Association of Friends of the
National Cancer-free Initiative



Egypt
Cancer Network
Canada



Egypt
Cancer Network
USA 1

Description

PAEA (Pathway Association Enrichment Analysis), is a new method for metabolomics pathway enrichment analysis using a novel association algorithm to increase the initial input list with significant associates metabolites (compounds), to give a bigger picture on possible activated pathways. In this version KEGG (Kyoto Encyclopedia of Genes and Genomes) is used as background database, for enrichment and association analysis. All the species available on KEGG can be used by the user.

Availability and Installation

PAEA is available in two forms on R;

1. Offline version as a Shiny app.
2. Online version as a Shiny app.

Contribution Guidelines

For bugs and suggestions, the most effective way is by raising an issue on the [GitHub](#) issue tracker. [GitHub](#) allows you to classify your issues so that we know if it is a bug report, feature request or feedback to the authors.

Email: Proteomicslab2017@gmail.com

Email: ali.anwar@57357.org

Email: aliali.mostafa99@gmail.com

Email: MohmedSoudy2009@gmail.com

If you find PAEA useful, please cite our paper:



Installation Steps

1. For the Offline version of PAEA Shiny app.

For installation procedures please follow the link to the [GitHub](#) page for better explanation.

2. For the Online version of PAEA Shiny app.

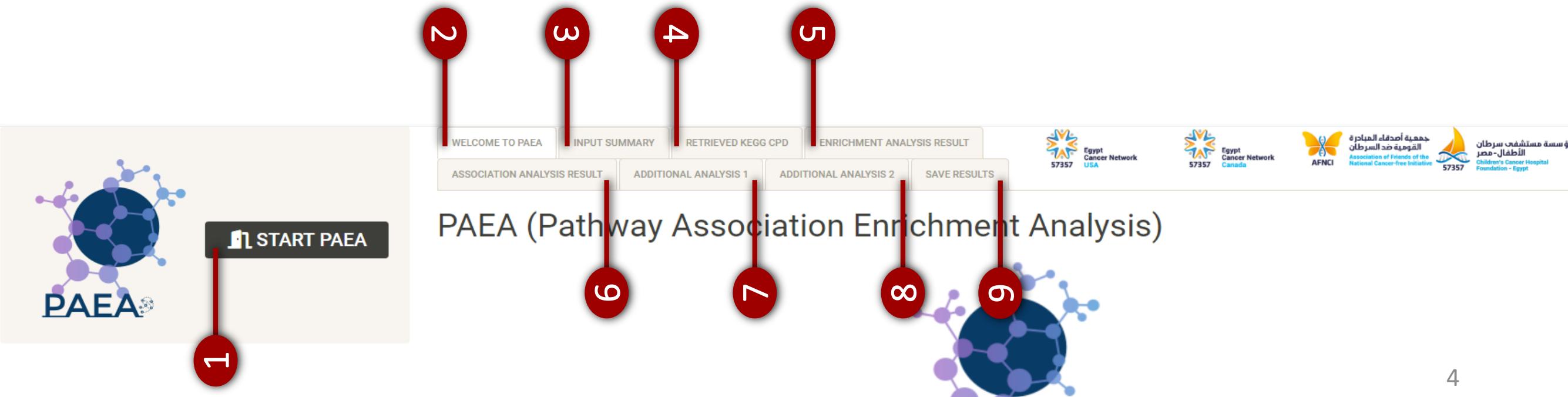
For the online version of PAEA shiny app without any prerequisites or installation steps, please follow the link to the app. <https://proteomicslab57357.shinyapps.io/PAEA/>



PAEA Welcome Page

The PAEA welcome page contains 9 objects;

1. By pressing the “START PAEA” button, it will take the user automatically to the main tabs.
2. This tab is referring to the welcome page, it contains general information about PAEA and its GitHub page. As well as, how someone can cite PAEA.
3. The “INPUT SUMMARY” tab. This tab will show the pathways of specific organisms chose by the user.
4. The “RETRIEVED KEGG CPD” tab. This tab will show all the CPDs of all pathways presented in the “INPUT SUMMARY” tab.
5. The “ENRICHMENT ANALYSIS RESULT” tab. This tab will show the first enrichment analysis result (the normal enrichment analysis) represented in a plot.
6. The “ASSOCIATION ANALYSIS RESULT” tab. This tab will show the association enrichment analysis result (the new method) represented in a plot.
7. The “ADDITIONAL ANALYSIS 1” tab. This tab will contain additional analysis for comparison of the two outputs from the enrichment analysis and from the association enrichment analysis. As well as, related diseases and other related pathways for only the significant pathways resulted from the association enrichment analysis.
8. The “ADDITIONAL ANALYSIS 2” tab. This tab will contain additional network analysis, for only the significant pathways resulted from the association enrichment analysis.
9. The “SAVE RESULTS” tab. This tab is the last tab where the user can choose where to save the results and control some plotting parameters as (width, height, and dpi).



Sidebar Panel 1

- Sidebar panel 1 is the main sidebar in PAEA, where the user can enter his input (CPDs) and controlling the parameters of the association enrichment.

- The input and the parameters in this sidebar panel should be entered and selected carefully, as PAEA will run based on it.

1. In this text input window, the user has to enter three- letters KEGG code corresponding to an organism. The PAEA will retrieve the pathways for the corresponding organism and use it as a background database for the enrichment and the association analysis (e.g. “hsa” is corresponding to human).
2. The “OPEN SPECIES IDS GUIDE” button directs the user to an online pdf file containing all the KEGG species IDs.
3. The first file the user has to upload the KEGG CPDs file (a CSV file contains all the KEGG pathways with all the CPDs). This file is already supported by PAEA and can be easily downloaded from the [GitHub page](#).
4. The second file must contain one column of the user CPDs IDs. The column should have no header and a CSV or TXT file could be used ([example](#)).
5. After uploading the necessary files, the user can choose one of the available methods for adjusting the p-value (for more dealies about the available method, please follow this [link](#)).
6. The first slider control the alpha of the adjusted p-value (for the enrichment analysis, the first and the second one), with 0.05 as a default.
7. The second slider control the alpha of the p-value (for the association analysis), with 0.01 as a default.
8. The third slider to put a cutoff for the minimum number of CPDs founded in a KEGG pathway and the user input to be considered in the enrichment analysis.

The screenshot shows the PAEA Sidebar Panel 1 with several numbered callouts pointing to specific controls:

- 1. A red circle points to a text input field labeled "Enter the Species KEGG ID:" containing "hsa".
- 2. A red circle points to a button labeled "OPEN SPECIES IDS GUIDE".
- 3. A red circle points to a "BROWSE..." button and a text field showing "No file selected" for "Enter the Species KEGG CPD".
- 4. A red circle points to a "BROWSE..." button and a text field showing "No file selected" for "Choose The CPD File".
- 5. A red circle points to a dropdown menu set to "none" under "Adjusted p-value method:". A small "▼" indicates it's a dropdown.
- 6. A red circle points to a horizontal slider for "Adjusted p-value (enrichment analysis)" currently set at 0.05.
- 7. A red circle points to a horizontal slider for "p-value (association analysis)" currently set at 0.01.
- 8. A red circle points to a horizontal slider for "Only use metabolite sets containing at least" currently set at 2.
- A red arrow points from the top right towards the sidebar panel.

At the bottom of the sidebar are two buttons: "ENRICHMENT ANALYSIS" and "ASSOCIATION ANALYSIS", followed by "SAVE THE RESULT" and "RESTART THE SESSION".

Sidebar Panel 1

1. The “ENRICHMENT ANALYSIS” button, will run the first enrichment analysis on the original user CPDs input.
2. The “ASSOCIATION ANALYSIS” button, will run the association analysis and the enrichment analysis on the new list of CPDs resulted from the association scan.
3. The “SAVE THE RESULT” button, will take the user to the “SAVE RESULT”, to choose where to save the results.
4. The “RESTART THE SESSION” button, will restart the running session. The PAEA will restart and all running and current data will be deleted.

The screenshot shows the PAEA software interface with the following components:

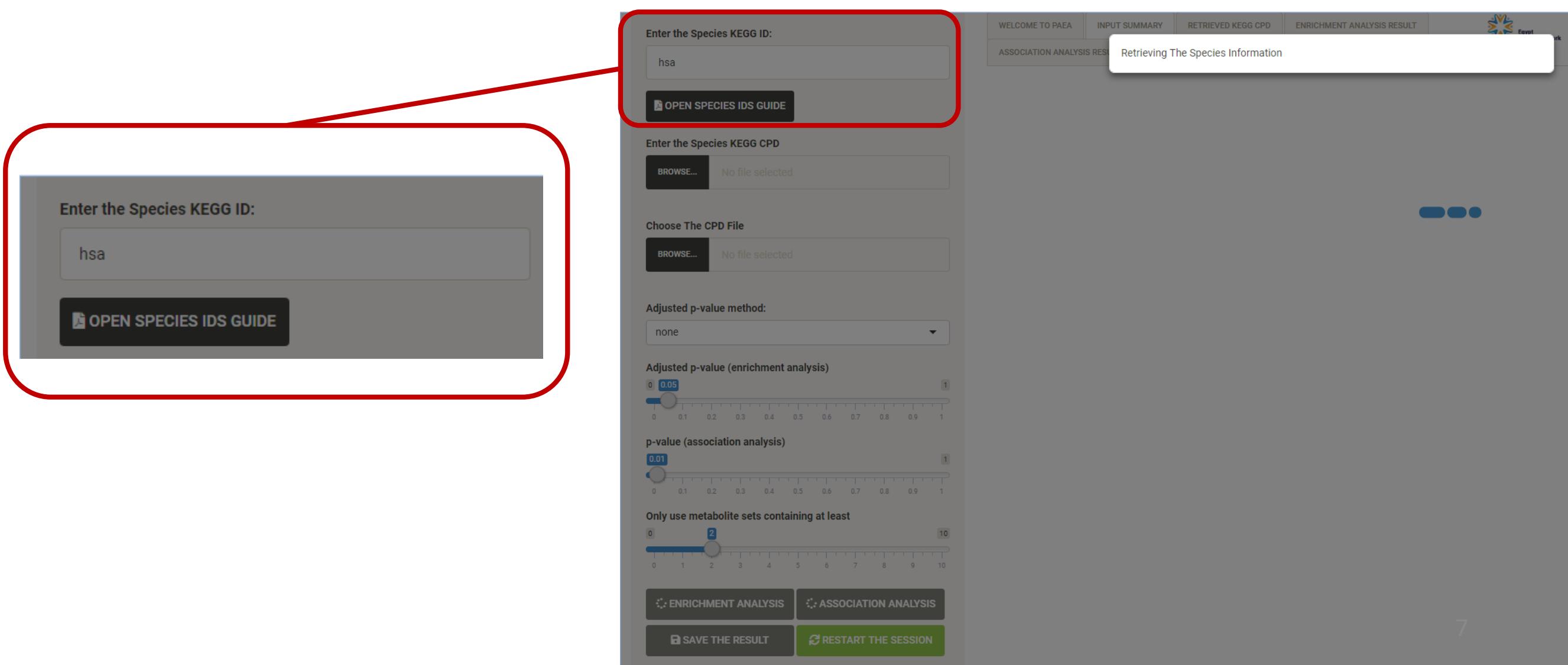
- Enter the Species KEGG ID:** A text input field containing "hsa".
- OPEN SPECIES IDS GUIDE**: A button.
- Enter the Species KEGG CPD**: A section with a "BROWSE..." button and a text input field showing "No file selected".
- Choose The CPD File**: A section with a "BROWSE..." button and a text input field showing "No file selected".
- Adjusted p-value method:** A dropdown menu set to "none".
- Adjusted p-value (enrichment analysis)**: A slider set to 0.05.
- p-value (association analysis)**: A slider set to 0.01.
- Only metabolite sets containing at least**: A slider set to 2.
- ENRICHMENT ANALYSIS**: A grey button.
- ASSOCIATION ANALYSIS**: A grey button.
- SAVE THE RESULT**: A grey button.
- RESTART THE SESSION**: A green button.

Four red circles with numbers 1 through 4 point to the following buttons:

- 1 points to the ENRICHMENT ANALYSIS button.
- 2 points to the ASSOCIATION ANALYSIS button.
- 3 points to the SAVE THE RESULT button.
- 4 points to the RESTART THE SESSION button.

Step 1. Retrieving The Species Information

- In the first step, the PAEA should retrieve the specific species KEGG pathways information, using the KEGG species ID (as shown in the figure below).
- When the user enters a three- letters ID the PAEA will start automatically to download the information.
- This step can take between 5 seconds to 30 seconds maximum, if the internet connection is stable.



The screenshot shows the PAEA (Pathway Analysis and Enrichment Analysis) interface. A red box highlights the 'Enter the Species KEGG ID:' input field, which contains 'hsa'. Below this field is a button labeled 'OPEN SPECIES IDS GUIDE'. Another red box highlights the 'Enter the Species KEGG CPD' section, which includes a 'BROWSE...' button and a message 'No file selected'. The interface also features sections for 'Adjusted p-value method:', 'Adjusted p-value (enrichment analysis)', 'p-value (association analysis)', and 'Only use metabolite sets containing at least'. At the bottom are buttons for 'ENRICHMENT ANALYSIS', 'ASSOCIATION ANALYSIS', 'SAVE THE RESULT', and 'RESTART THE SESSION'. The top navigation bar includes tabs for 'WELCOME TO PAEA', 'INPUT SUMMARY', 'RETRIEVED KEGG CPD', 'ENRICHMENT ANALYSIS RESULT', and 'ASSOCIATION ANALYSIS RESULT'. A status message 'Retrieving The Species Information' is displayed in the top right. The number '7' is located in the bottom right corner of the image.

Step 1. Retrieving The Species Information

- After retrieving the information, it will appear to the user in table format, as shown in the figure below.

The table consists of 4 columns;

- IDs_org:** represent the organism-specific pathway KEGG ID.
- Name:** the name of each pathway.
- IDs_map:** represent the general pathway KEGG ID.
- Org_acc_ids:** show the organism KEGG ID.

Enter the Species KEGG ID:
hsa

OPEN SPECIES IDS GUIDE

Enter the Species KEGG CPD
BROWSE... No file selected

Choose The CPD File
BROWSE... No file selected

Adjusted p-value method:
none

Adjusted p-value (enrichment analysis)
0 0.05 1

p-value (association analysis)
0.01 1

Only use metabolite sets containing at least
0 2 10

ENRICHMENT ANALYSIS **ASSOCIATION ANALYSIS**

SAVE THE RESULT **RESTART THE SESSION**

WELCOME TO PAEA INPUT SUMMARY RETRIEVED KEGG CPD ENRICHMENT ANALYSIS RESULT
ASSOCIATION ANALYSIS RESULT ADDITIONAL ANALYSIS 1 ADDITIONAL ANALYSIS 2 SAVE RESULTS

57357 Egypt Cancer Network USA 57357 Egypt Cancer Network Canada 57357 AFNCI 57357 جمعية أصدقاء المريض القوية ضد السرطان Association of Friends of the National Cancer-free Initiative 57357 مفاسدة مستشفى سرطان الأطفال - مصر National Cancer Hospital Foundation - Egypt

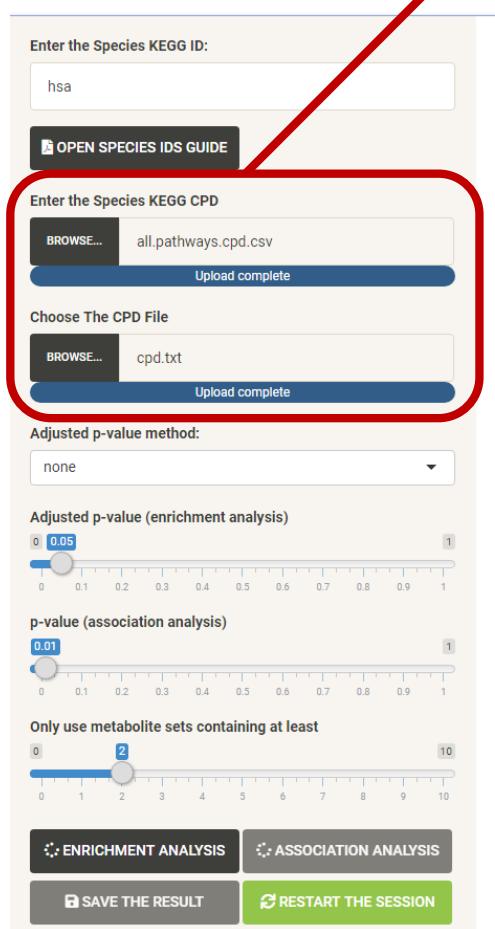
Show 25 entries Search:

IDs_org	Name	IDs_map	org_acc_ids
hsa00010	Glycolysis / Gluconeogenesis	map00010	hsa
hsa00020	Citrate cycle (TCA cycle)	map00020	hsa
hsa00030	Pentose phosphate pathway	map00030	hsa
hsa00040	Pentose and glucuronate interconversions	map00040	hsa
hsa00051	Fructose and mannose metabolism	map00051	hsa
hsa00052	Galactose metabolism	map00052	hsa
hsa00053	Ascorbate and aldarate metabolism	map00053	hsa
hsa00061	Fatty acid biosynthesis	map00061	hsa
hsa00062	Fatty acid elongation	map00062	hsa
hsa00071	Fatty acid degradation	map00071	hsa
hsa00072	Synthesis and degradation of ketone bodies	map00072	hsa
hsa00100	Steroid biosynthesis	map00100	hsa
hsa00120	Primary bile acid biosynthesis	map00120	hsa
hsa00130	Ubiquinone and other terpenoid-quinone biosynthesis	map00130	hsa
hsa00140	Steroid hormone biosynthesis	map00140	hsa
hsa00190	Oxidative phosphorylation	map00190	hsa
hsa00220	Arginine biosynthesis	map00220	hsa
hsa00230	Purine metabolism	map00230	hsa
hsa00232	Caffeine metabolism	map00232	hsa
hsa00240	Pyrimidine metabolism	map00240	hsa

Step 2. Uploading The KEGG CPDs Information & User CPDs

The second step is to upload two files;

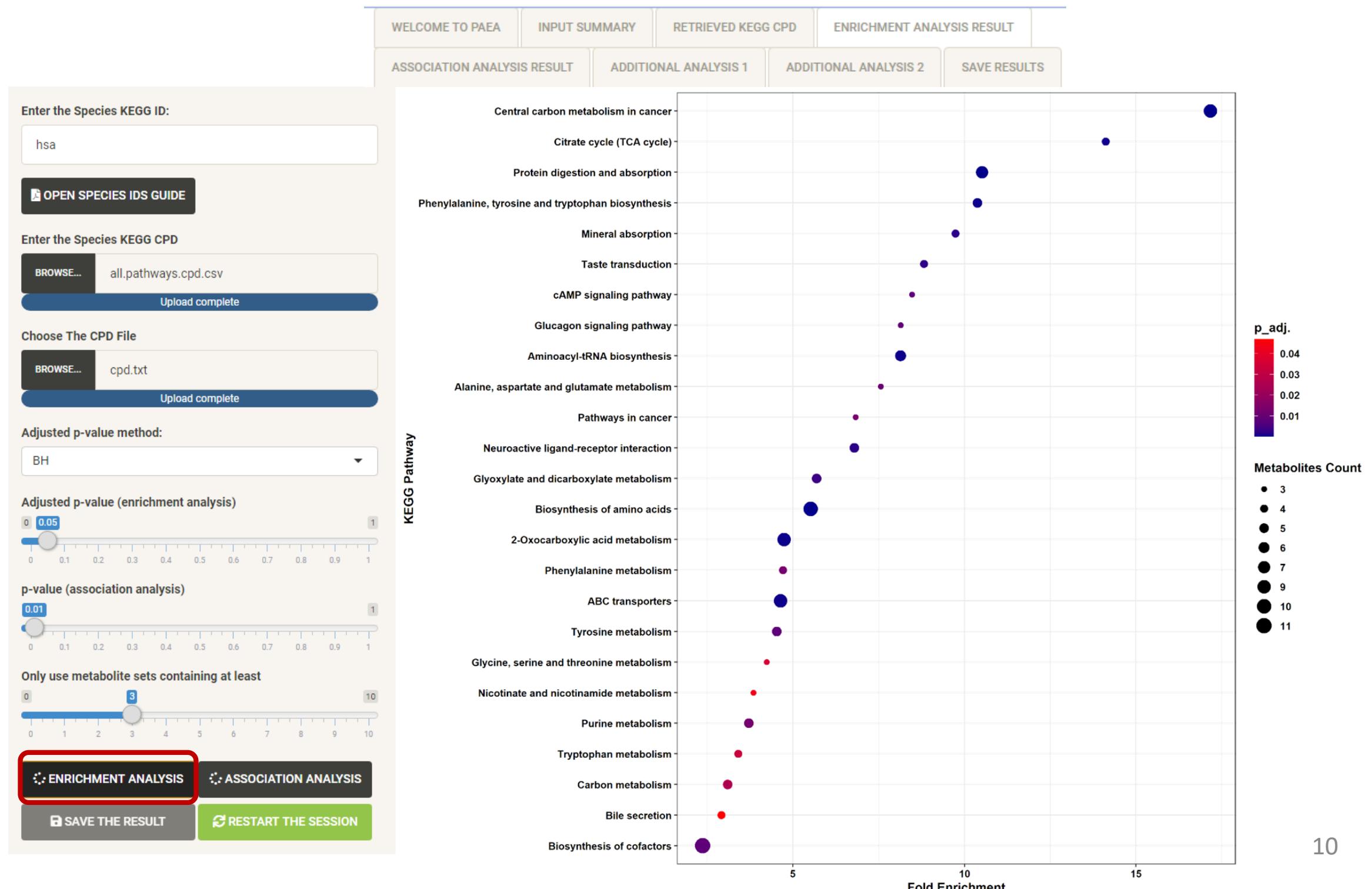
- The first file contains all the pathways and all the CPDs take action in it from the KEGG database, to be used as a background database search for PAEA. This file can be downloaded from the PAEA [GitHub](#) page.
- The second file must contain one column of the user CPDs IDs. The column should have no header and a CSV or TXT file could be used ([example](#)).



The screenshot shows the PAEA software interface. At the top, there is a navigation bar with links for "WELCOME TO PAEA", "INPUT SUMMARY", "RETRIEVED KEGG CPD", and "ENRICHMENT ANALYSIS RESULT". Below the navigation bar, there are tabs for "ASSOCIATION ANALYSIS RESULT", "ADDITIONAL ANALYSIS 1", "ADDITIONAL ANALYSIS 2", and "SAVE RESULTS". A search bar is located at the top right. The main area has a header "Enter the Species KEGG CPD" with a "BROWSE..." button and a file input field containing "all.pathways.cpd.csv". Below this is a blue button labeled "Upload complete". Another section titled "Choose The CPD File" also has a "BROWSE..." button and a file input field containing "cpd.txt", followed by a blue "Upload complete" button. On the left side, there are several adjustable sliders for "Adjusted p-value method", "Adjusted p-value (enrichment analysis)", "p-value (association analysis)", and "Only use metabolite sets containing at least". At the bottom, there are four buttons: "ENRICHMENT ANALYSIS", "ASSOCIATION ANALYSIS", "SAVE THE RESULT", and "RESTART THE SESSION". A red box highlights the "Enter the Species KEGG CPD" section. The right side of the interface displays a list of pathway IDs (map00010, map00020, map00030, map00040, map00051, map00052, map00053, map00061, map00062, map00071, map00072) each followed by a long list of metabolite IDs.

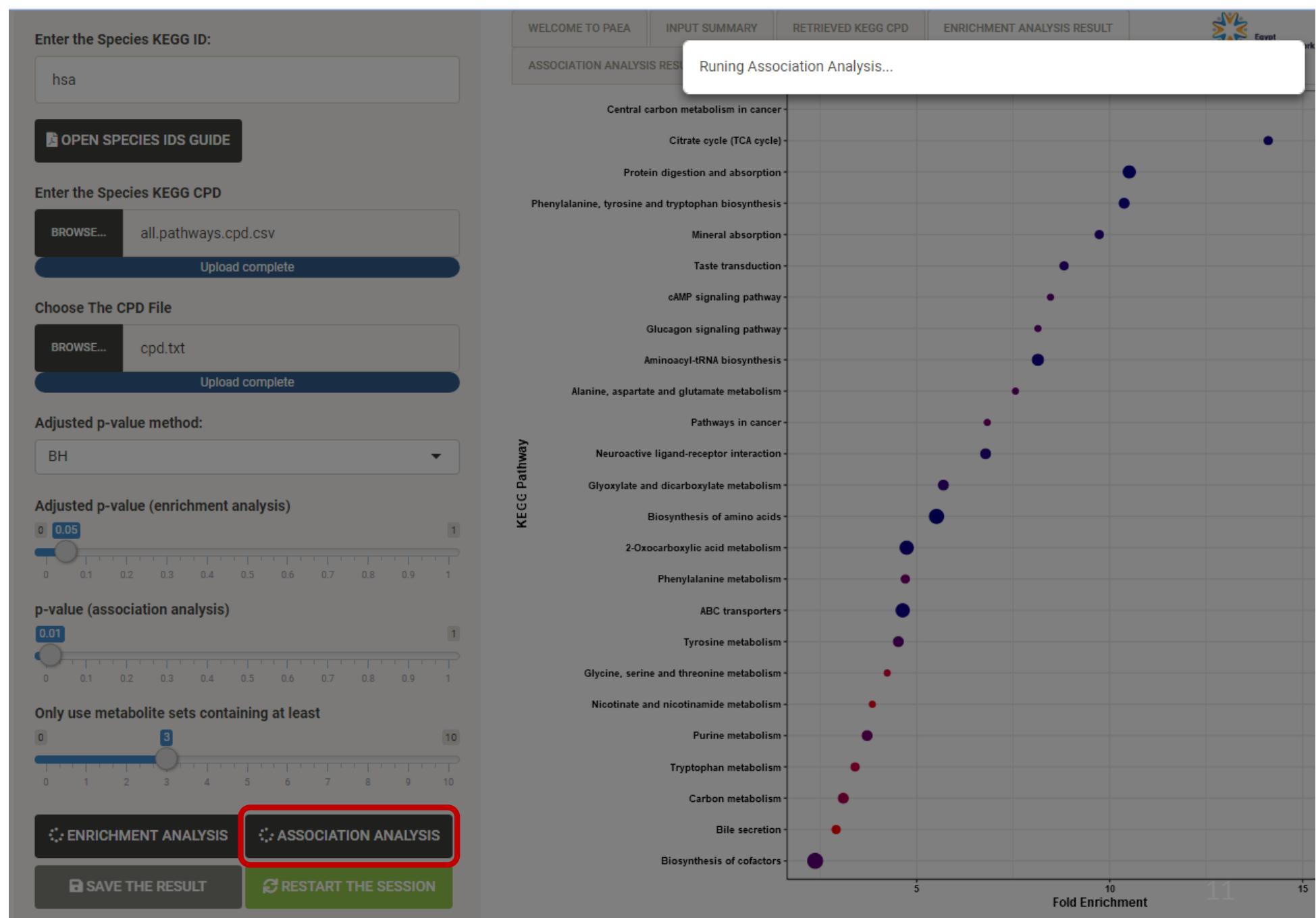
Step 3. The first Enrichment Analysis

- After choosing the parameter for PAEA as described [previously](#). The user should press the “ENRICHMENT ANALYSIS” button to start the first enrichment analysis. This process can take between 10 seconds to 30 seconds.
- The result will appear as the enrichment analysis plot, as shown in the figure below.
- The x-axis represent the “Fold Enrichment” for each pathway, the y-axis show the name of the pathways, the size of the circles reflects the number of the CPDs in each pathway (described by the black legend), and the color reflects the adjusted p-value.



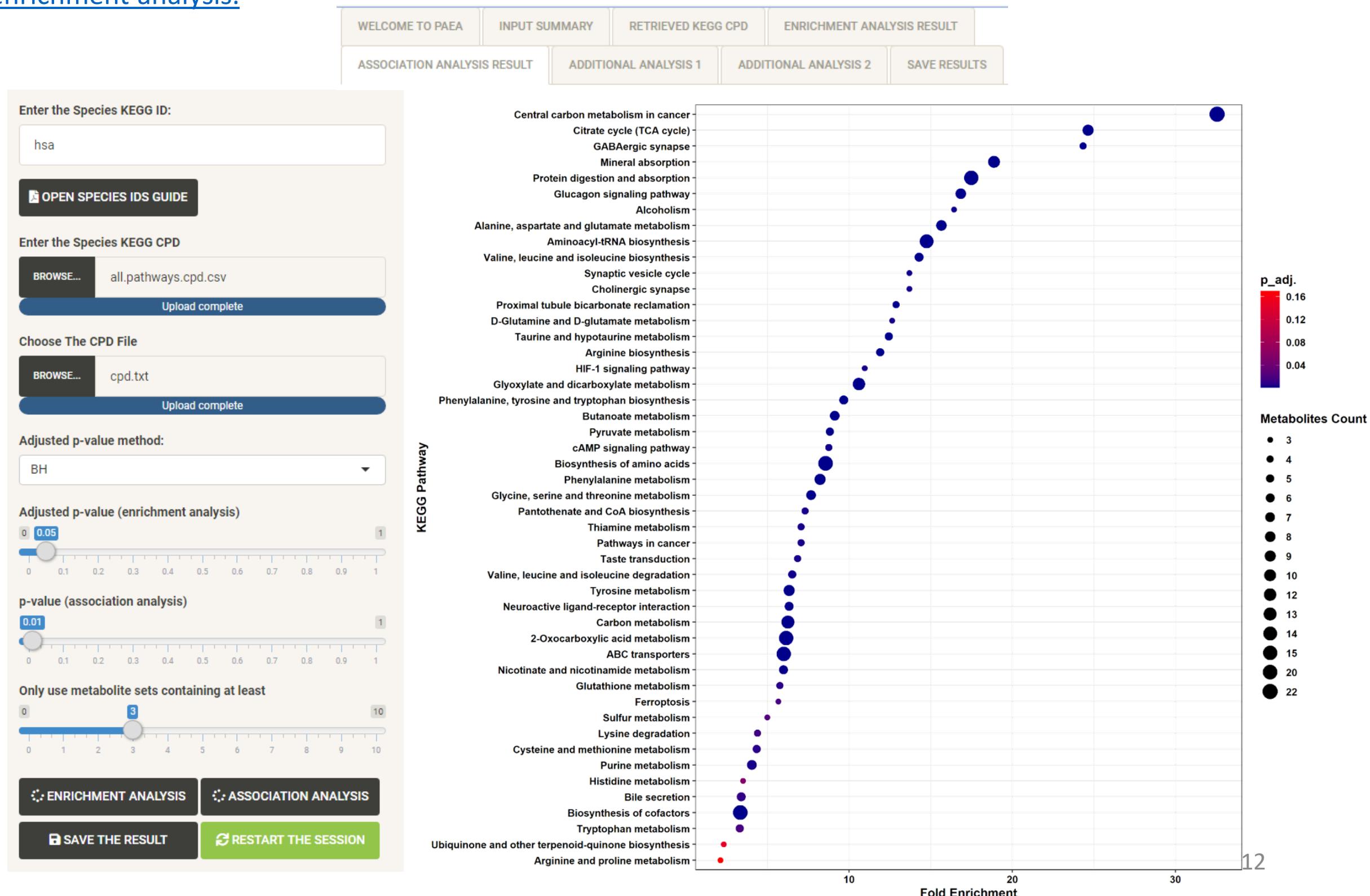
Step 4. The Association analysis & The Second Enrichment Analysis

- By pressing the “ASSOCIATION ANALYSIS” button, the association enrichment analysis will run automatically without any other requirements.
- This step can take some time depending on the user CPDs input and the organism (used by the user) pathways’ size (the background database).



Step 4. The Association analysis & The Second Enrichment Analysis

- Automatically PAEA will take the user to the tab of the association enrichment analysis result.
- As shown in the figure below, this plot can be described as same as the [plot of the first enrichment analysis](#).



Sidebar Panel 2

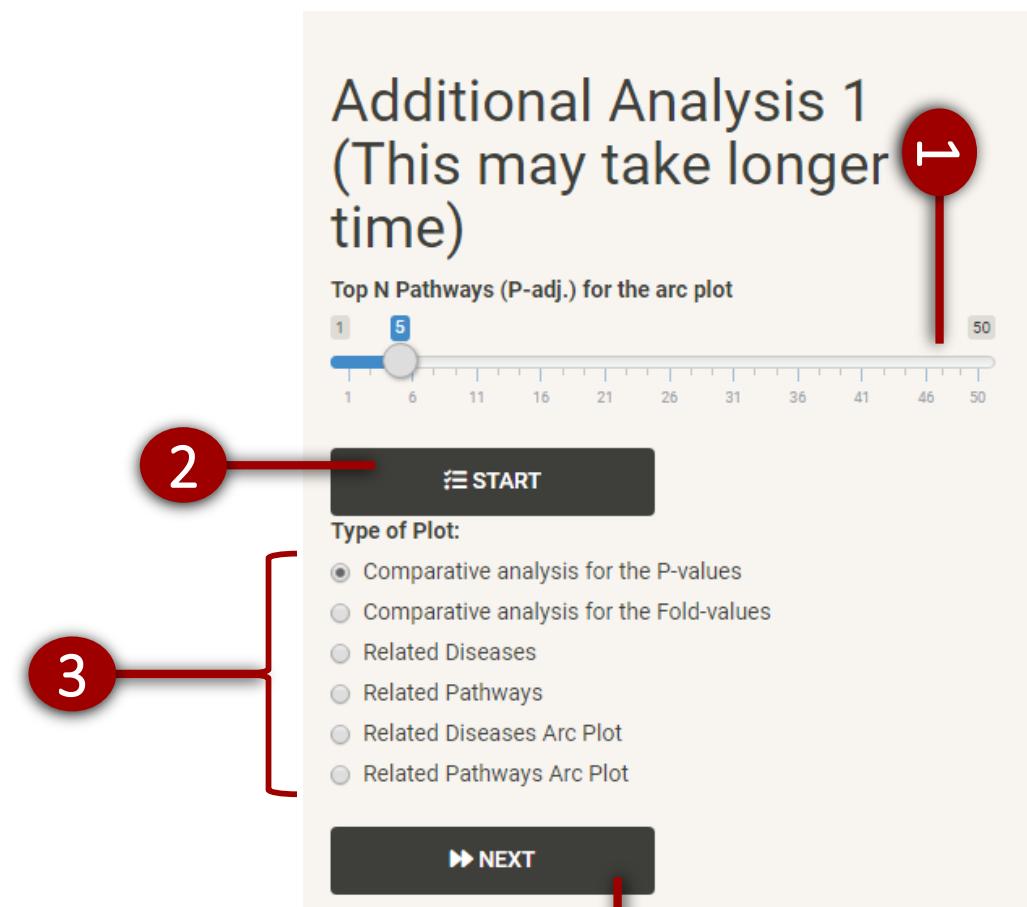
- Sidebar Panel 2 contains the controllers for the additional analysis 1 implemented in PAEA.
- Note; all the additional analysis conducted by the PAEA is only for the pathways and CPDs resulted from the association enrichment analysis.

1. The slider controls the number of the top N pathways (resulted from the association enrichment analysis) that will be considered in the additional analysis (for the arc plot only).

2. The “START” button will start the additional analysis. The running time will depend mainly on two factors (**the internet connection speed and the number selected in the slider**).

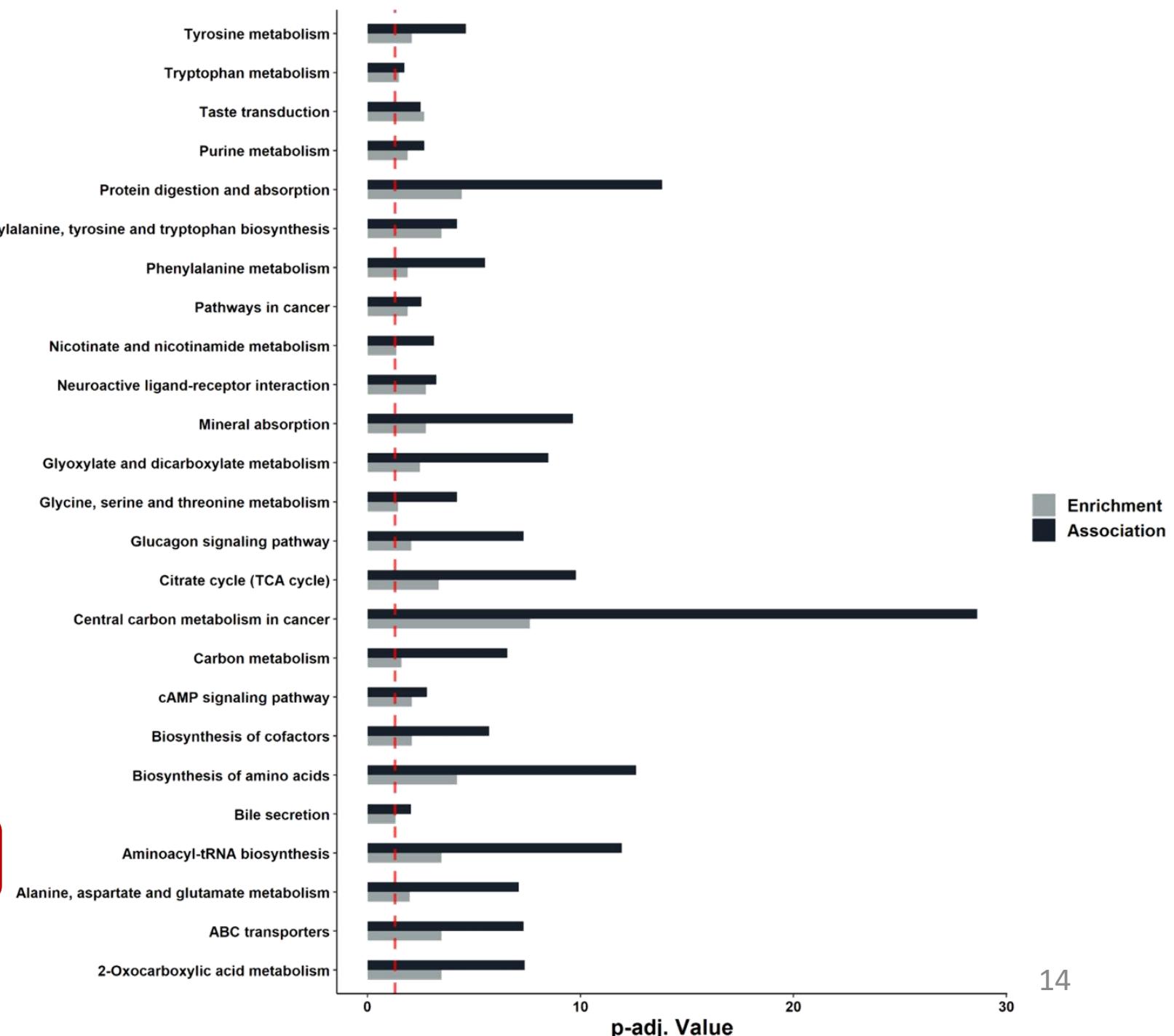
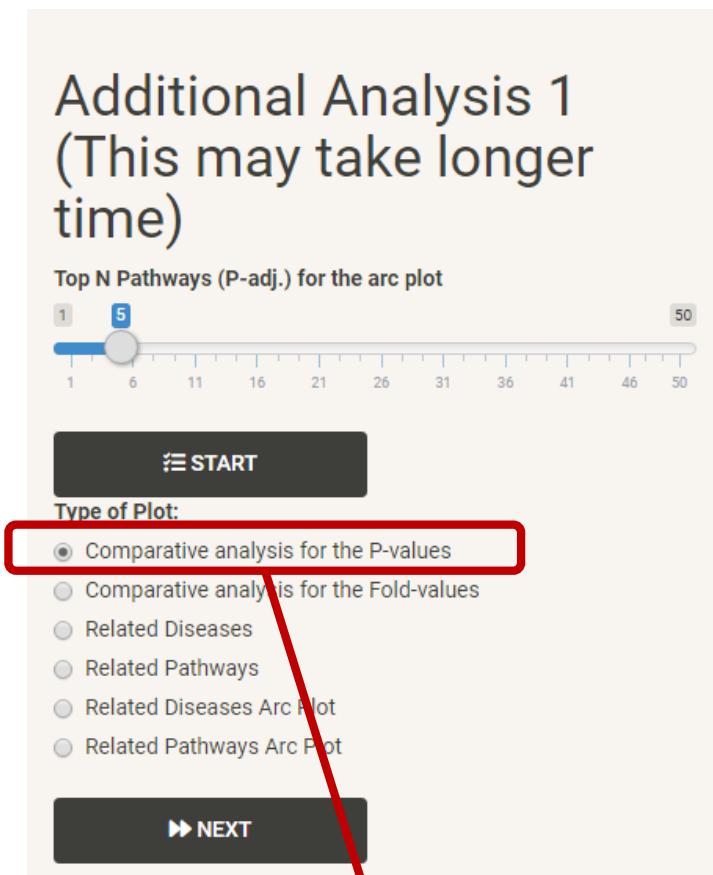
3. The “radio buttons”, will change the analysis and the plot type presented in the tab.

4. The “NEXT” button, will take the user to the additional analysis 2 tab.



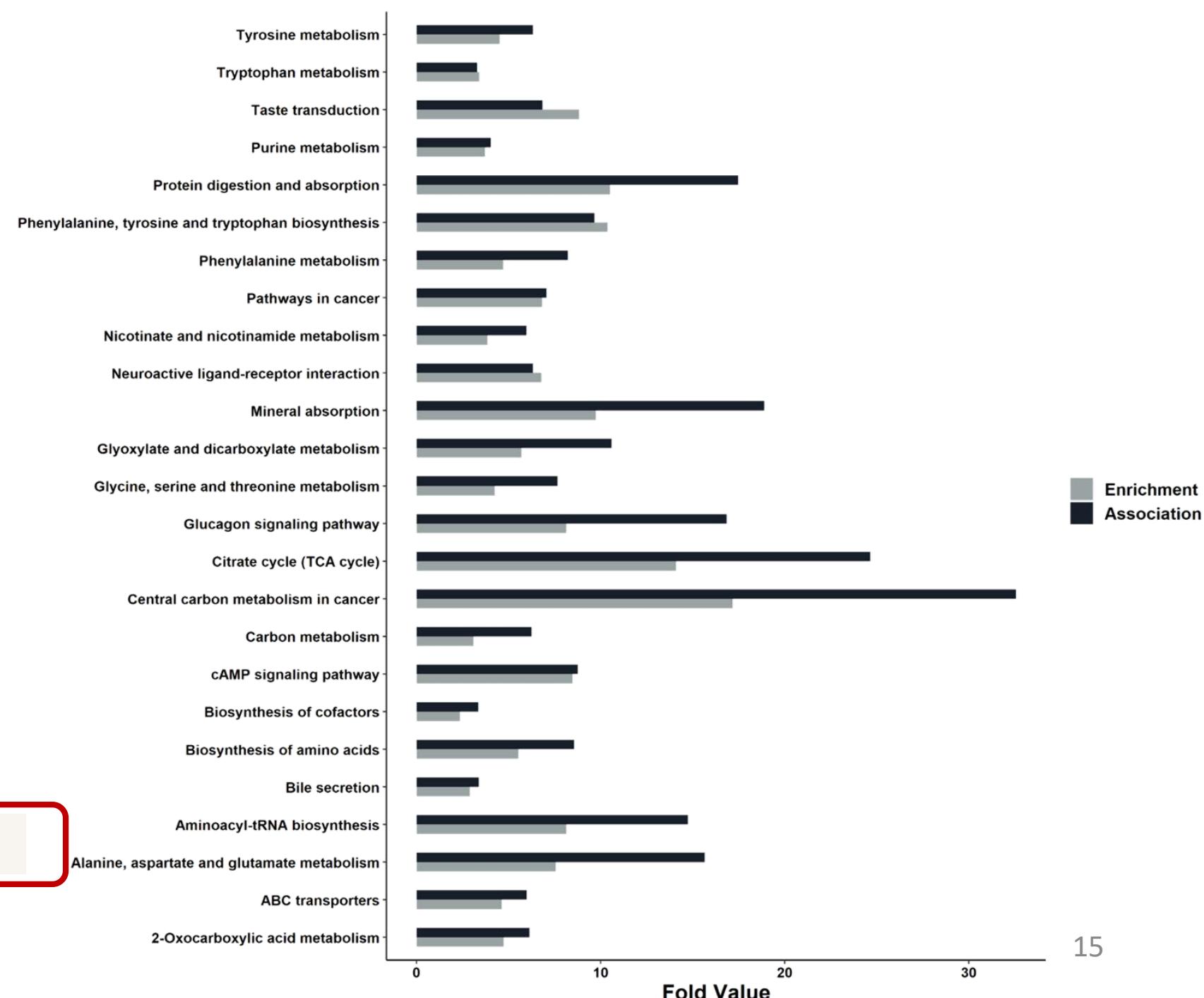
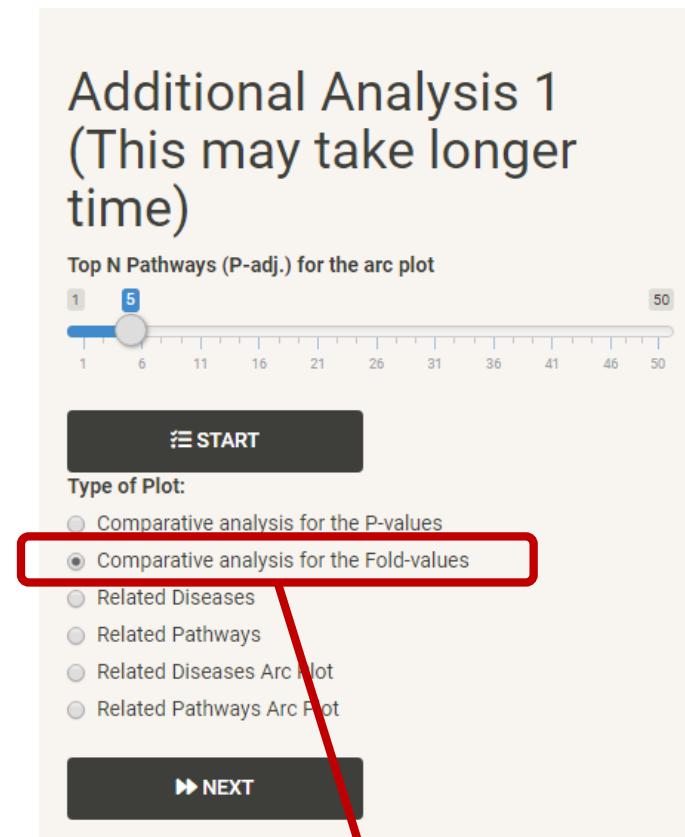
Step 5. Additional Analysis 1

- The first plot in the additional analysis 1; shows the adjusted p-value comparison between the normal enrichment analysis (using only the user CPDs) and the association enrichment analysis, of the shared pathways.
- The black bars show the adjusted p-value of the pathways resulted from the association analysis. And the lighter bars show the adjusted p-value of the pathways resulted from the normal enrichment analysis
- The dashed red line represents the cutoff selected by the user as $-\log_{10}$ (adjusted p-value).



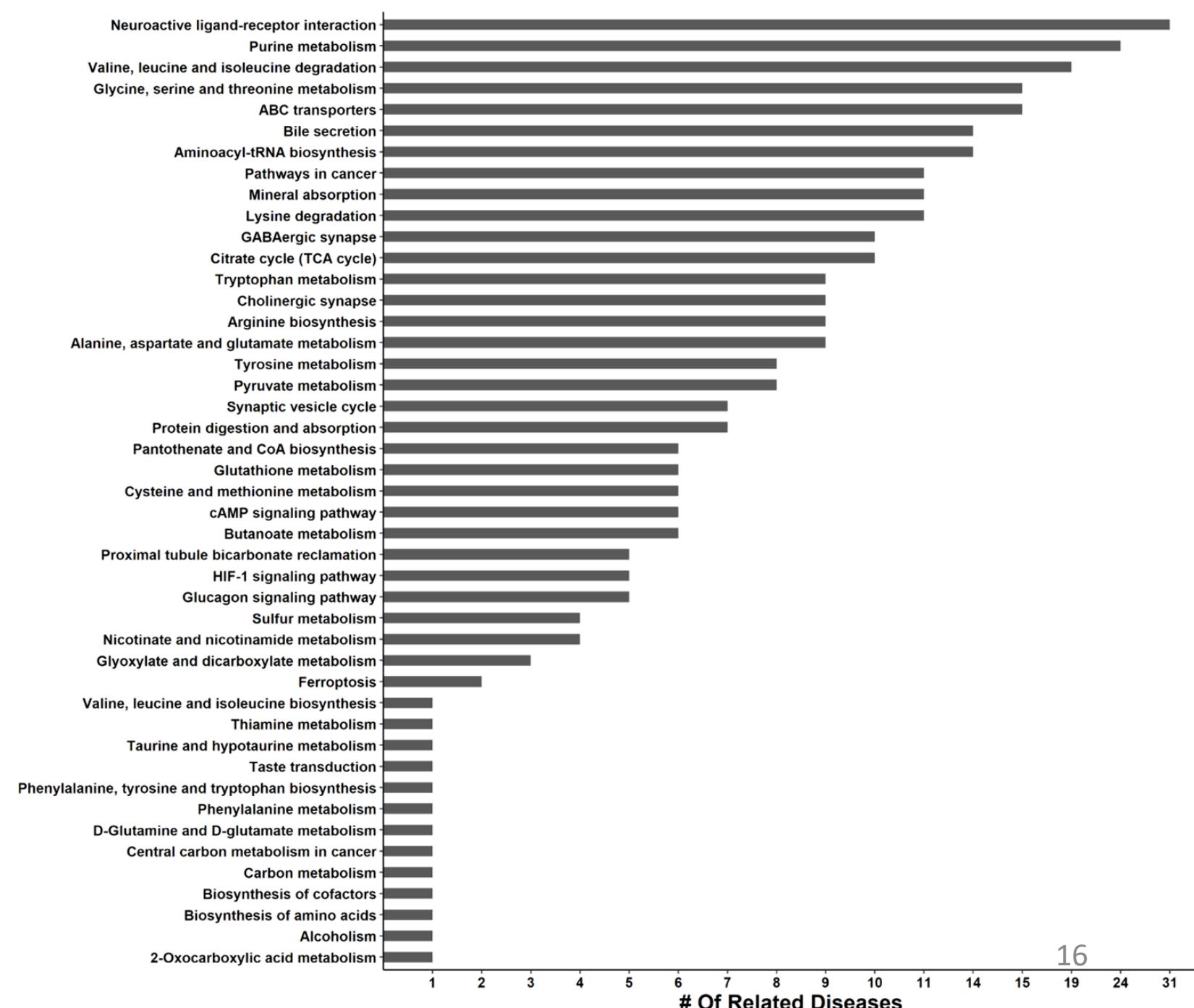
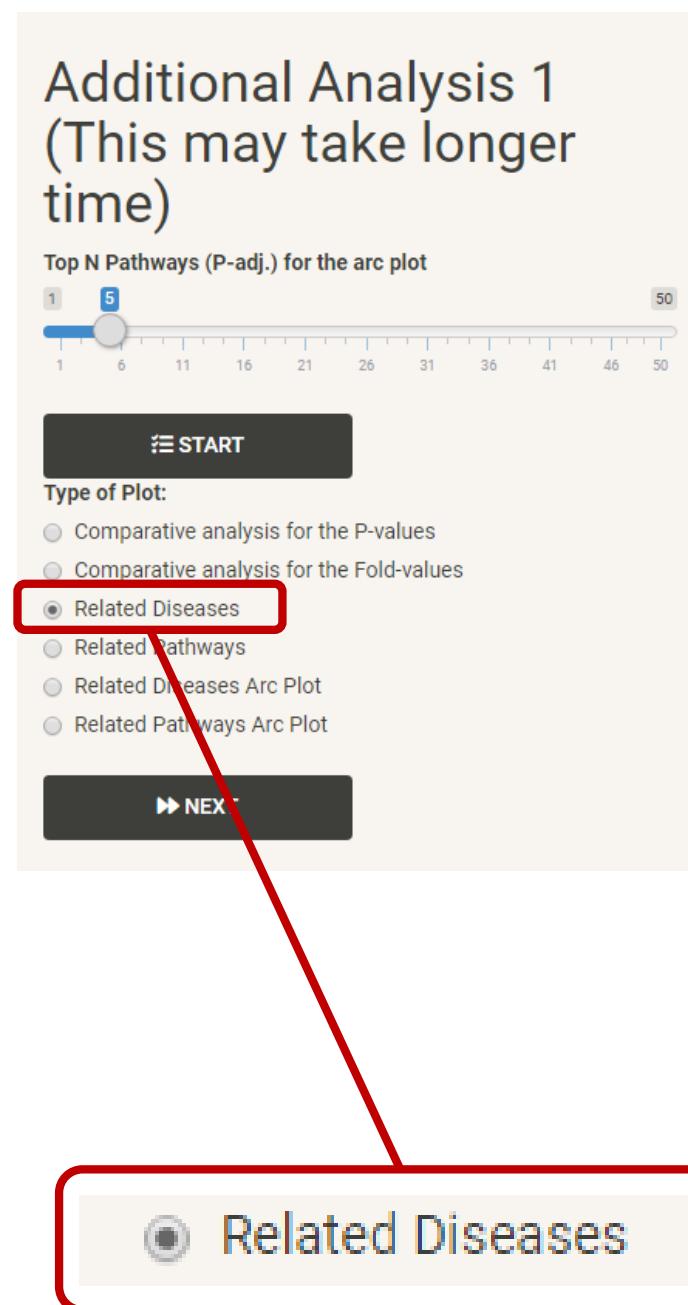
Step 5. Additional Analysis 1

- The second plot same as the [previous](#) one, except in the point of the comparison here the fold enrichment value is being compared between the normal enrichment analysis (using only the user CPDs) and the association enrichment analysis, of the shared pathways.



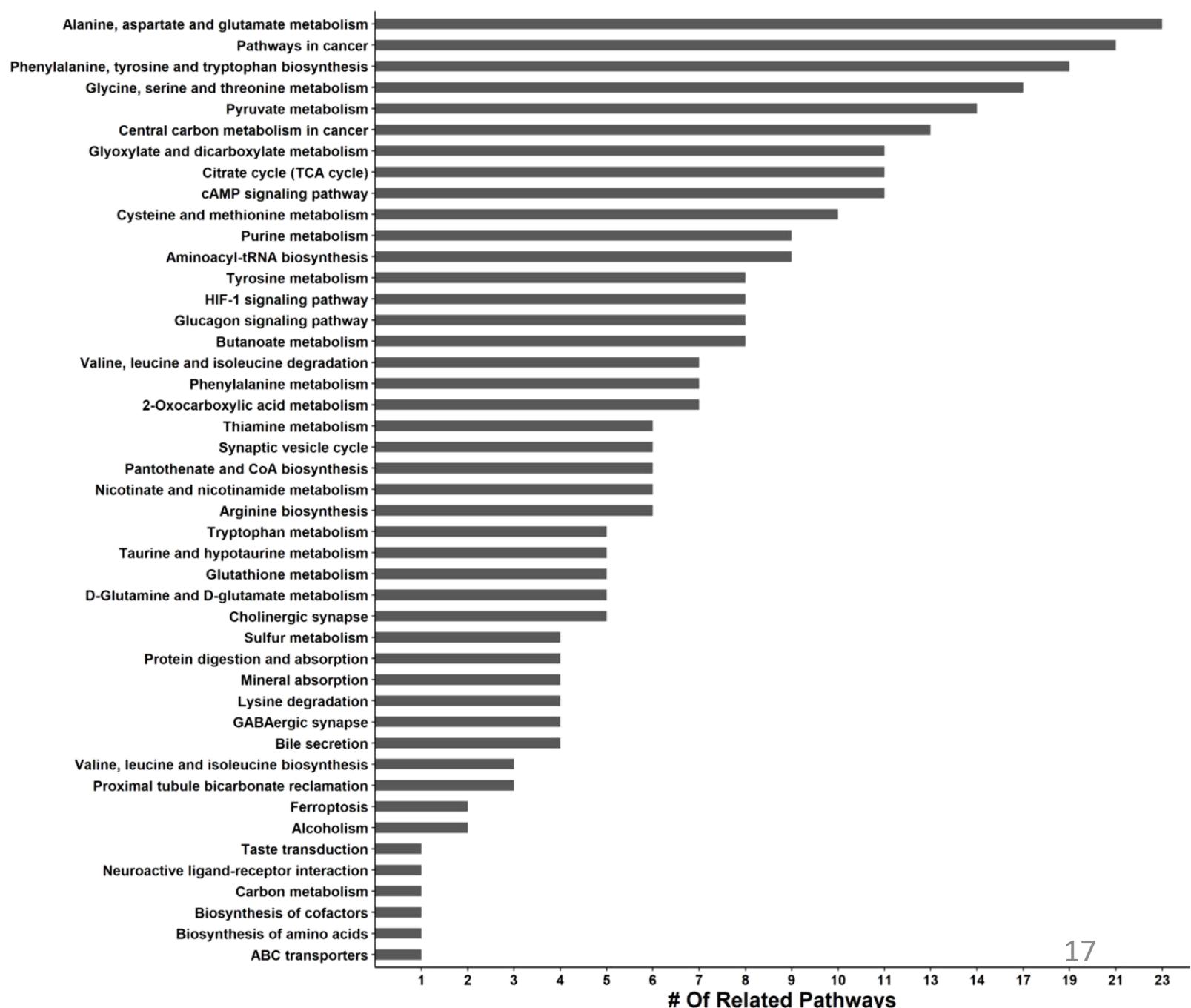
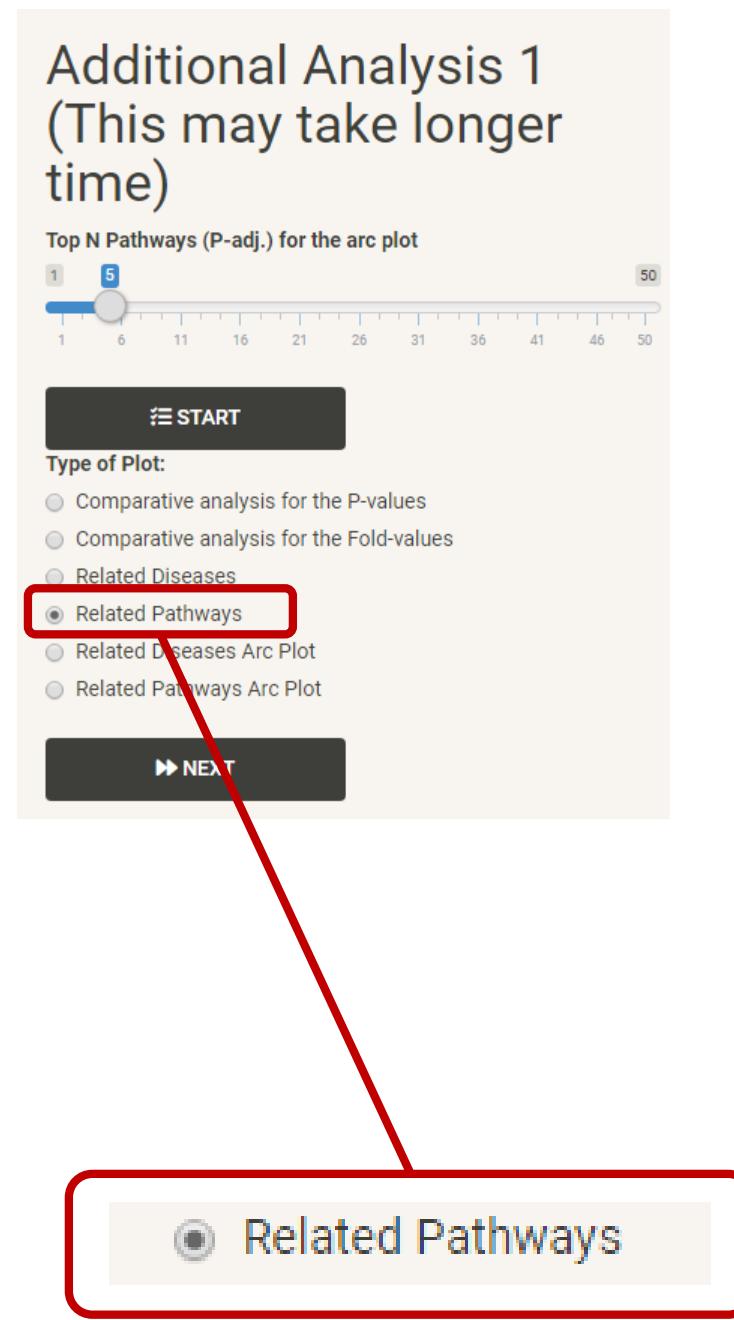
Step 5. Additional Analysis 1

- The third plot shows the count of related diseases for each significant pathway, resulted only from the association enrichment analysis. The data retrieved also from the KEGG database.



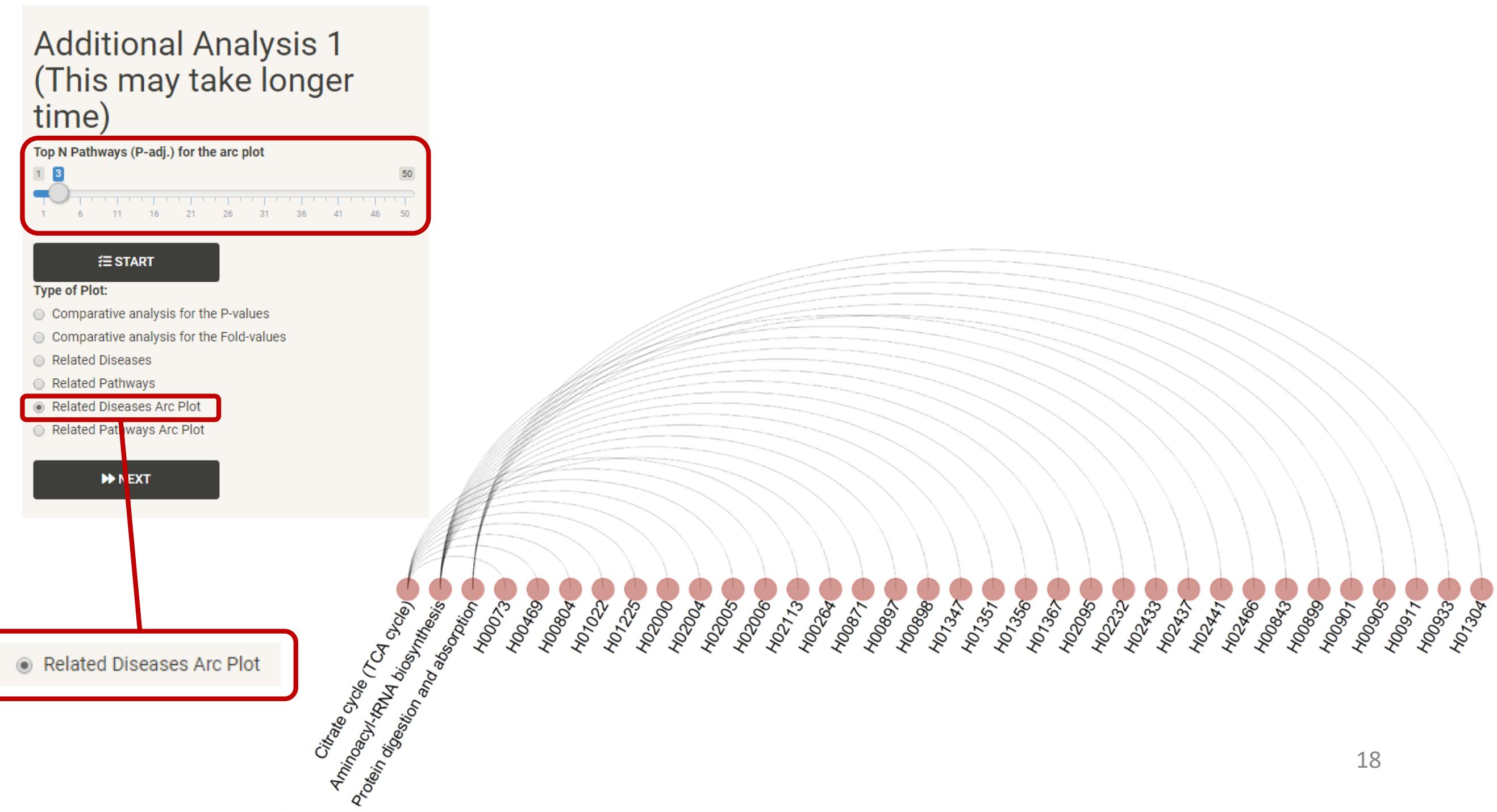
Step 5. Additional Analysis 1

- The 4th plot shows the count of related pathways for each significant pathway, resulted only from the association enrichment analysis. The data retrieved also from the KEGG database.



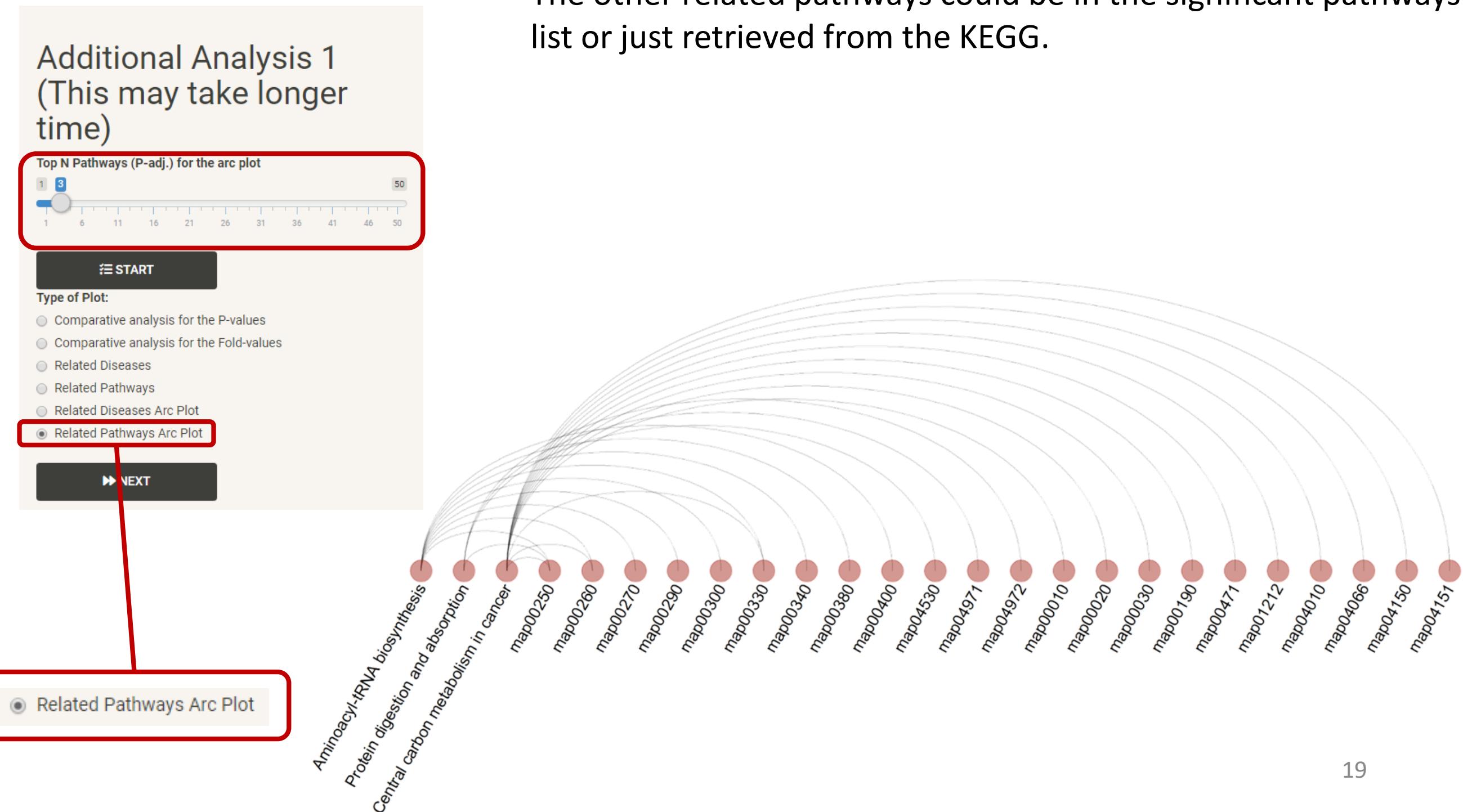
Step 5. Additional Analysis 1

- The 5th plot shows the related diseases for the top number of significant pathways (selected by the user). As shown in the figure, the top 3 significant pathways represented by their names, and their diseases related represented by KEGG diseases IDs.



Step 5. Additional Analysis 1

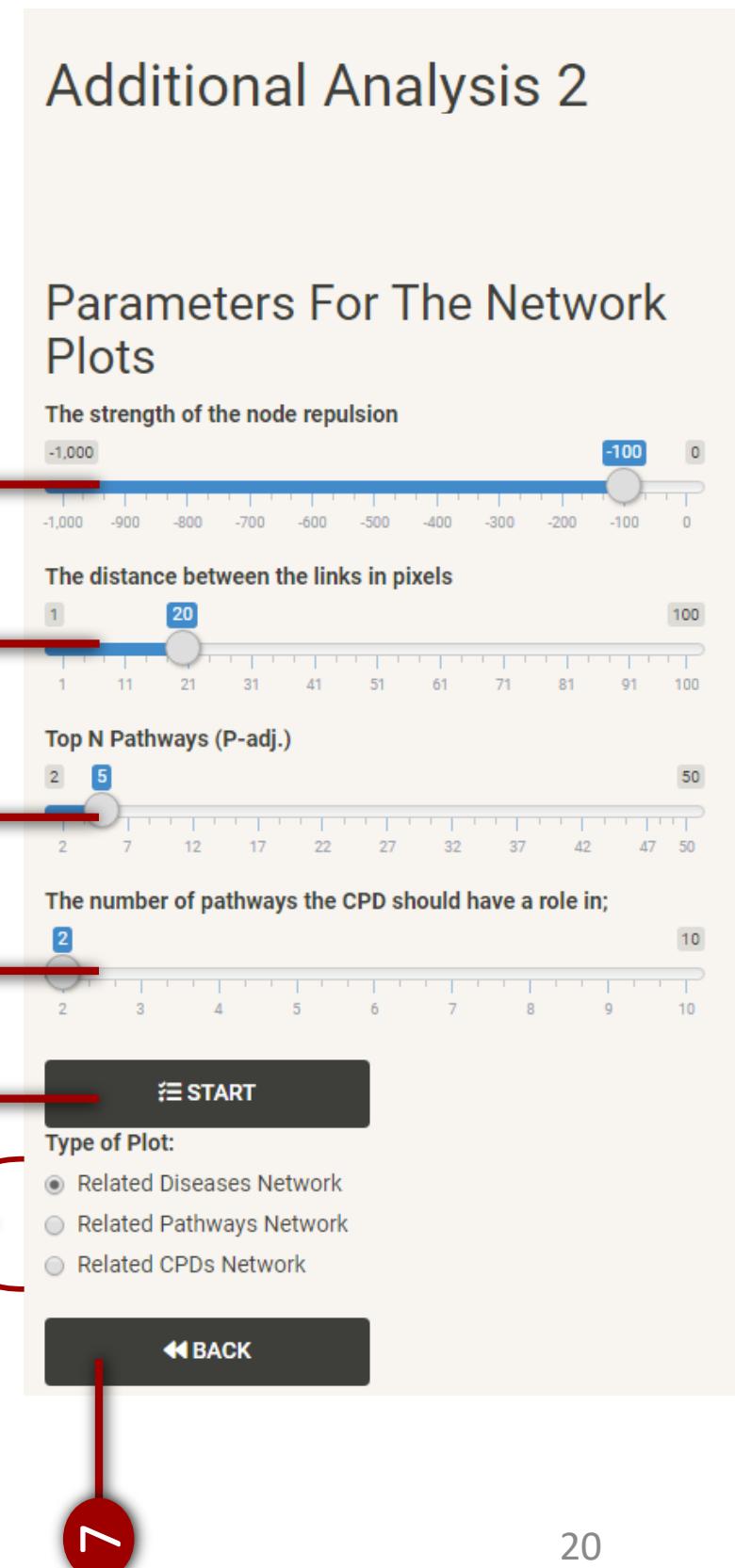
- The last plot shows the related pathways for the top number of significant pathways (selected by the user). As shown in the figure, the top 3 significant pathways represented by their names, and their related pathways are represented by KEGG pathway IDs.
- The other related pathways could be in the significant pathways list or just retrieved from the KEGG.



Sidebar Panel 3

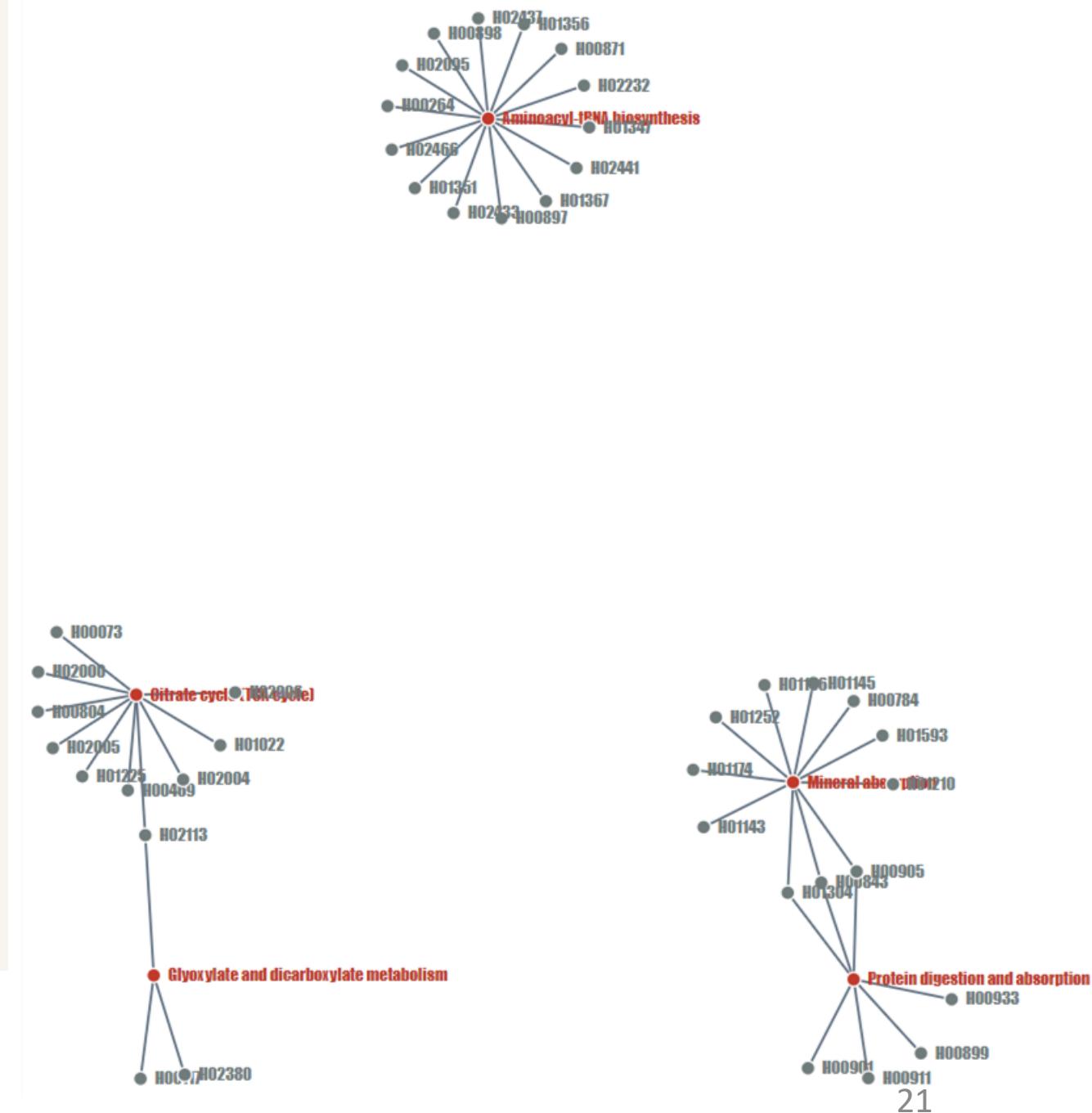
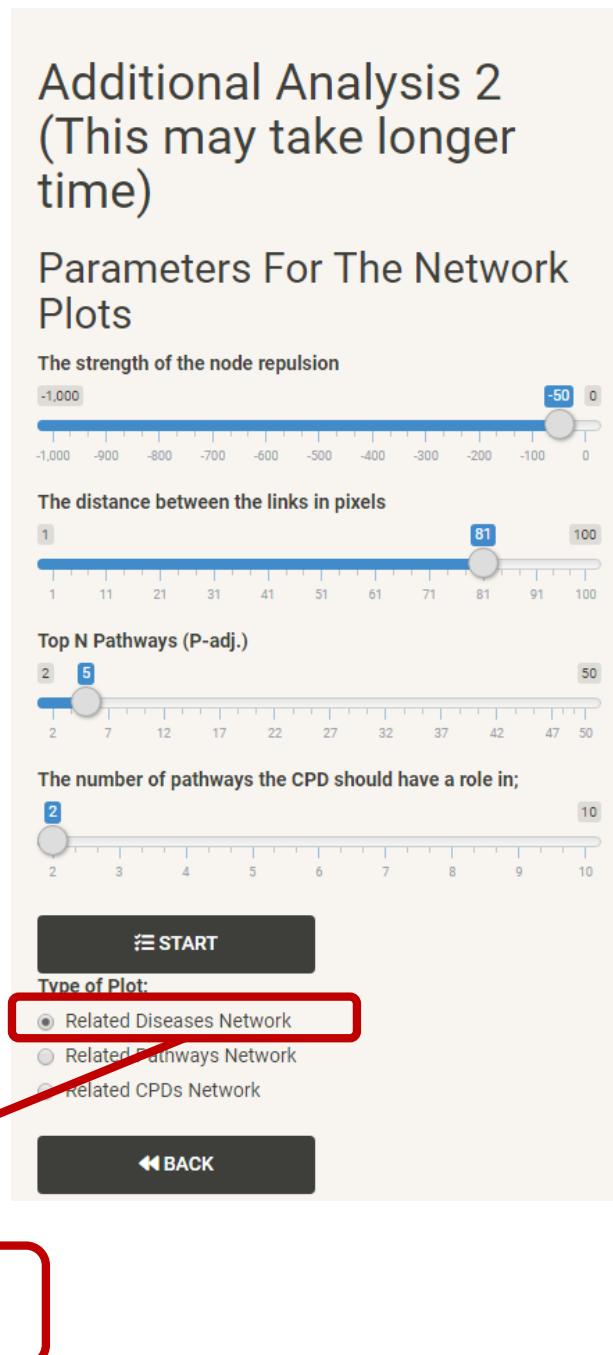
- Sidebar Panel 3 contains the controllers for the additional network analysis 2 implemented in PAEA.

1. The first slider controls the strength of the repulsion between the network nodes for the “Related Diseases Network” and the “Related Pathways Network” analysis (default value -100).
2. The second slider controls the distance between the links of the network for the “Related Diseases Network” and the “Related Pathways Network” (default value 20).
3. The third slider controls the number of pathways that will be considered in the network analysis (the top N pathways sorted by the adjusted p-value), for all plots (default value 5).
4. The last slider set a cutoff for the minimum number of pathways a CPD should be involved in, for the “Related CPDs Network” analysis (default value 2).
5. The “START” button will run the analysis and present the plots in the tab.
6. The “radio buttons”, will change the analysis and the plot type presented in the tab.
7. The “BACK” button, change the tab to the additional analysis 1 tab again.



Step 6. Additional Analysis 2

- The first interactive network plot in the additional analysis 2, is the related diseases for the top number of significant pathways (selected by the user) connected by the network mode.
- Light red nodes represent the pathways and the gray nodes represent the diseases connected or related by thus pathways.



Step 6. Additional Analysis 2

- The second interactive network plot, the related pathways for the top number of significant pathways (selected by the user) connected by the network mode.
- Light red nodes represent the pathways and the gray nodes represent the other pathways connected or related by thus pathways.
- The other related pathways could be in the significant pathways list or just retrieved from the KEGG.

Additional Analysis 2
(This may take longer time)

Parameters For The Network Plots

The strength of the node repulsion
[-1,000] [0]

The distance between the links in pixels
[1] [100]

Top N Pathways (P-adj.)
[2] [50]

The number of pathways the CPD should have a role in;
[2] [10]

START

Type of Plot:
 Related Diseases Network
 Related Pathways Network
 Related CPD Network

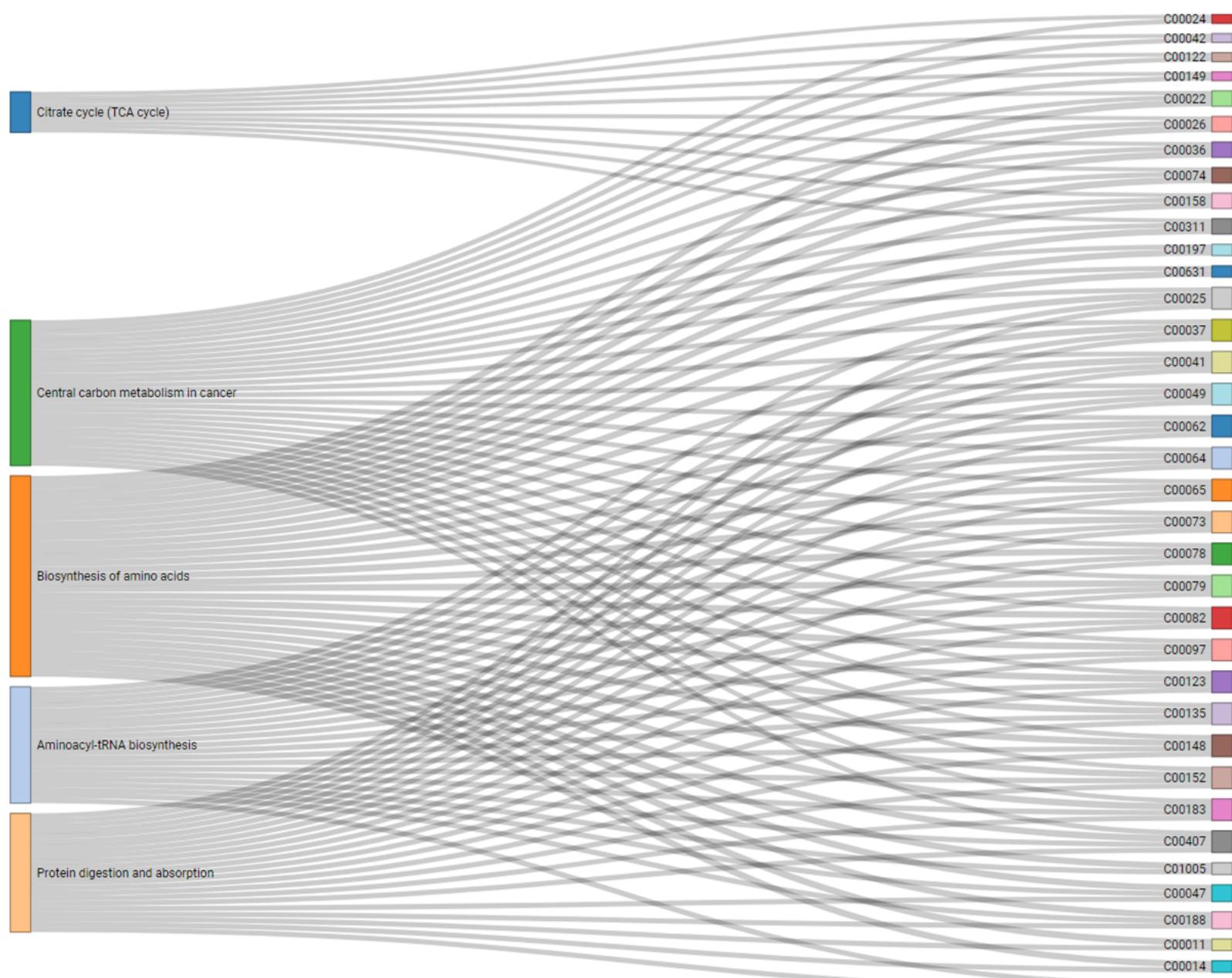
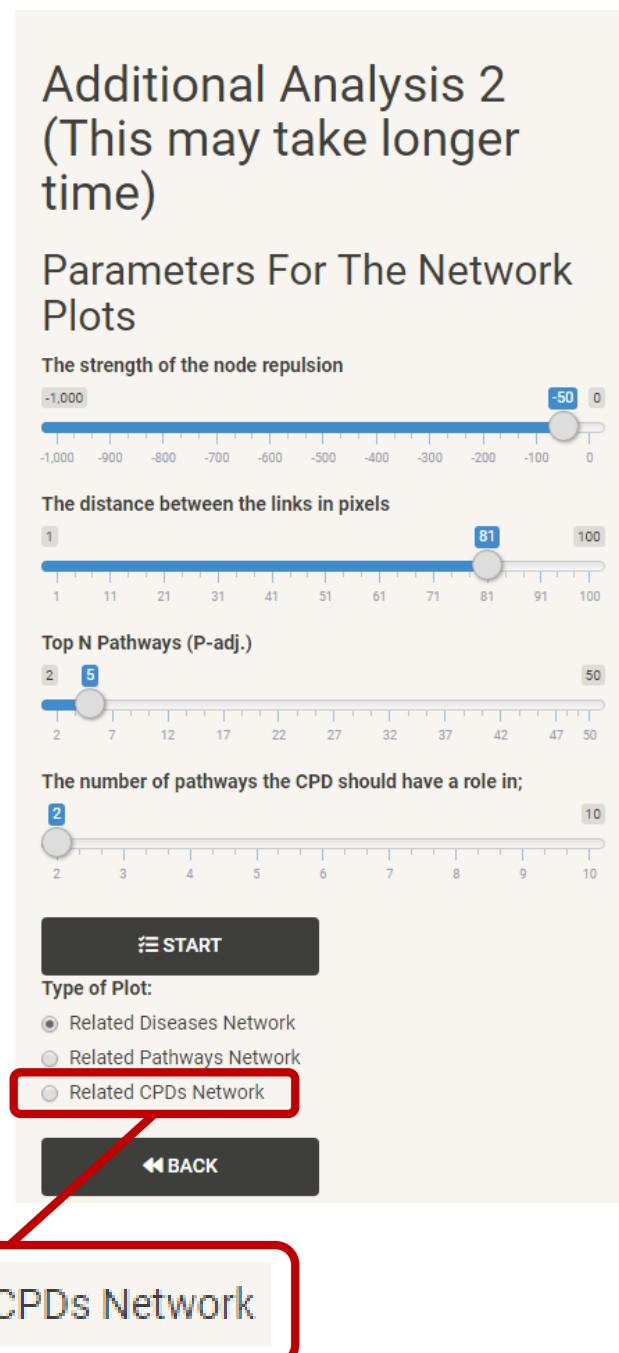
BACK

Related Pathways Network



Step 6. Additional Analysis 2

- In the last interactive Sankey plot, the CPDs for the top number of significant pathways (selected by the user) are connected by the Sankey mode to the pathways they involve in.
- The pathways are represented at the right side of the plot (names of the pathways), and the CPDs are represented at the left side of the plot using the CPD KEGG ID.



Parameters For The Network Plots

The strength of the node repulsion

The distance between the links in pixels

Top N Pathways (P-adj.)

The number of pathways the CPD should have a role in;

START

Type of Plot:

- Related Diseases Network
- Related Pathways Network
- Related CPDs Network

BACK

- The two sliders can control the parameters of the network analysis.
- By changing one or two of them the three plots will change according to the selected parameters, [as described before](#).
- The second slider is only for the Sankey plot.

Parameters For The Network Plots

The strength of the node repulsion

The distance between the links in pixels

Top N Pathways (P-adj.)

The number of pathways the CPD should have a role in;

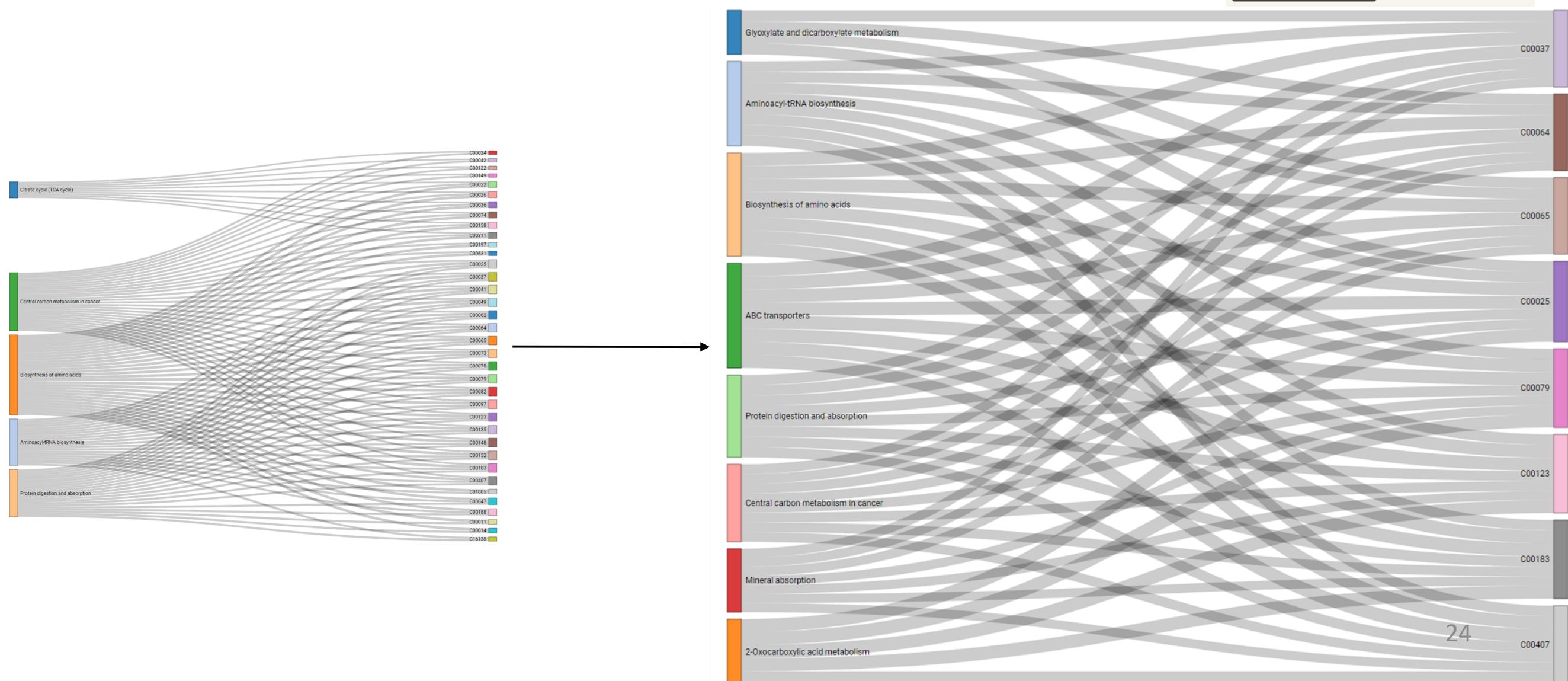
START

Type of Plot:

- Related Diseases Network
- Related Pathways Network
- Related CPDs Network

BACK

Example



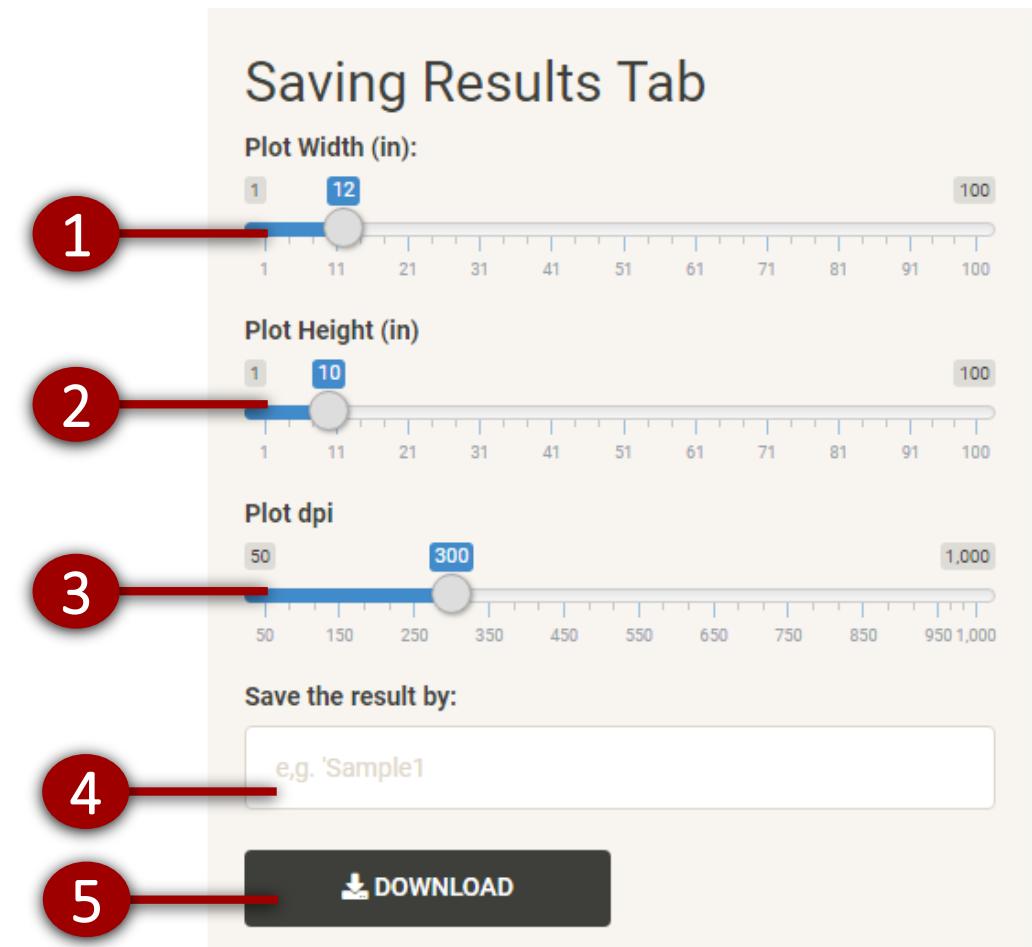
Note: Sometimes some of the top number significant pathways (selected by the user) does not have any related diseases or other pathways recorded in KEGG. So, if the user selects to present the top 10 pathways in the network analysis, instead, PAEA shows only 7, which means that there are 3 pathways in the top 2 that have no related diseases or other related pathways recorded in the KEGG database.



Sidebar Panel 4

- The last sidebar 4 contains the controllers for the parameters of the plots and where to save them.

1. The first slider controls the plot width in inches (in) (default value 12).
2. The second slider controls the plot height in inches (in) (default value 10).
3. The third slider controls the plot dpi (default value 300).
4. The text input can take a name from the user as a keyword to save all the session results by it.
5. The “DOWNLOAD” button, will save all the session result where the user specified.



Step 7. Saving The Results

- The final step is to save the results. By clicking the “SAVE THE RESULT” button, PAEA will automatically ask the user where to save the result. Then changing the tab to the SAVE RESULTS tab.
- In the “SAVE RESULTS” tab the user can control some parameters for the plots to be saved with; width, height, and dpi. As well as, a keyword to save all the results with.

WELCOME TO PAEA INPUT SUMMARY RETRIEVED KEGG CPD ENRICHMENT ANALYSIS RESULT

ASSOCIATION ANALYSIS RESULT ADDITIONAL ANALYSIS 1 ADDITIONAL ANALYSIS 2 SAVE RESULTS

Enter the Species KEGG ID:
hsa

[OPEN SPECIES IDS GUIDE](#)

OR

Enter the Species KEGG CPD
BROWSE... all.pathways.cpd.csv
Upload complete

Choose The CPD File
BROWSE... cpd.txt
Upload complete

Adjusted p-value method:
BH

Adjusted p-value (enrichment analysis)
0.05

p-value (association analysis)
0.01

Only use metabolite sets containing at least
3

ENRICHMENT ANALYSIS **ASSOCIATION ANALYSIS**

SAVE THE RESULT **RESTART THE SESSION**

Saving Results Tab

Plot Width (in): 12

Plot Height (in): 10

Plot dpi: 300

Save the result by: e.g. 'Sample1'

DOWNLOAD

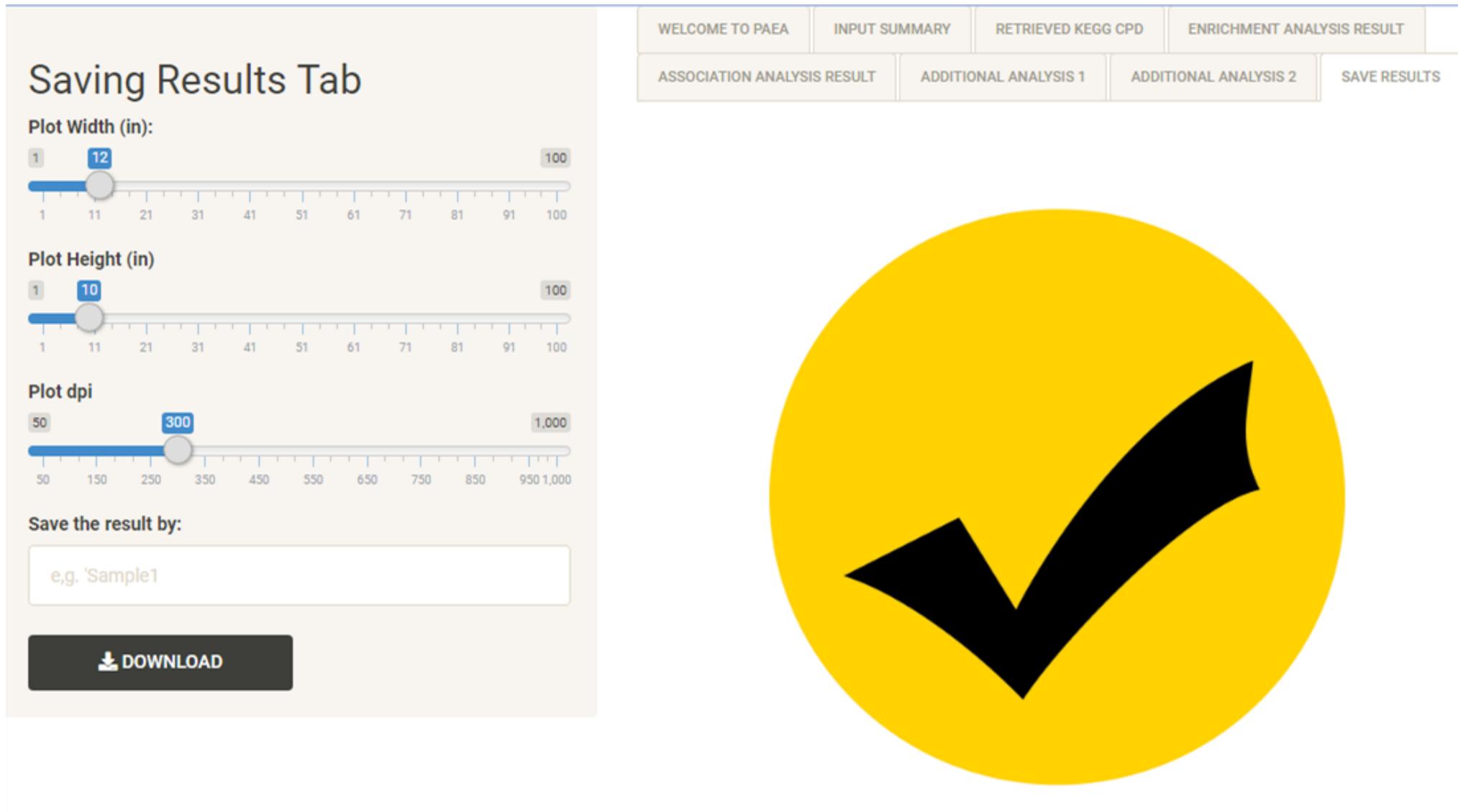
INPUT SUMMARY RETRIEVED KEGG CPD ENRICHMENT ANALYSIS RESULT

SAVING... Saving...

- By clicking the Download button the saving process will start.

Step 7. Saving The Results

- After saving all the results (which may take 30 second or less) a yellow true sign will appear, indicating that everything is okay and all the results saved.



Step 7. Saving The Results

- Example_CDPsSankey_files
- Example_DiseaseNetwork_files
- Example_PathWayNetwork_files
- Example_CDPsSankey.html
- Example_DiseaseNetwork.html
- Example_PathWayNetwork.html
- Example_Associationplot.csv
- Example_Enrichmentplot.csv
- Example_Related Pathways and Diseases....
- Example_Associationplot.tiff
- Example_Comparative analysis for the Fo...
- Example_Comparative analysis for the P-...
- Example_Enrichmentplot.tiff
- Example_Overlap analysis for the CPDs.tiff
- Example_Overlap analysis for the pathways...
- Example_Related Diseases arc.tiff
- Example_Related Diseases.tiff
- Example_Related Pathways arc.tiff
- Example_Related Pathways.tiff

File Name
CDPsSankey.html
DiseaseNetwork.html
PathWayNetwork.html
Associationplot.csv
Enrichmentplot.csv
Related Pathways and Diseases.csv
Associationplot.tiff
Comparative analysis for the Fold-values.tiff
Comparative analysis for the P-values
Enrichmentplot.tiff
Overlap analysis for the CPDs.tiff
Overlap analysis for the pathways.tiff
Related Diseases arc.tiff
Related Diseases.tiff
Related Pathways arc.tiff
Related Pathways.tiff



For easily step by step on how to use PAEA, follow the link to a [tutorial video](#).

Package information

- Version: v1.0.0
- License: GPL-3
- Encoding: UTF-8

Citation



Supported and Powered by



مؤسسة مستشفى سرطان
الأطفال - مصر
Children's Cancer Hospital
Foundation - Egypt



جمعية أصدقاء المبادرة
القومية ضد السرطان
Association of Friends of the
National Cancer-free Initiative



Egypt
Cancer Network
Canada



Egypt
Cancer Network
USA