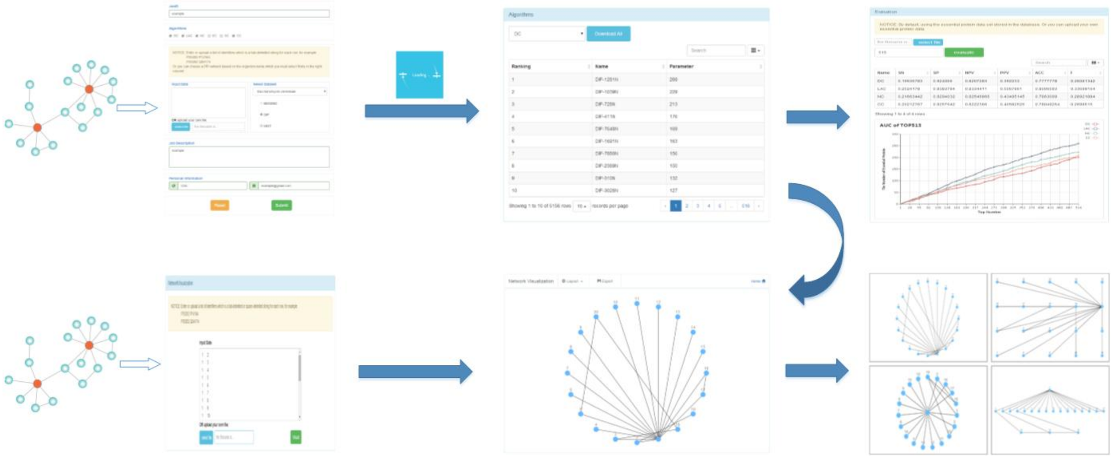
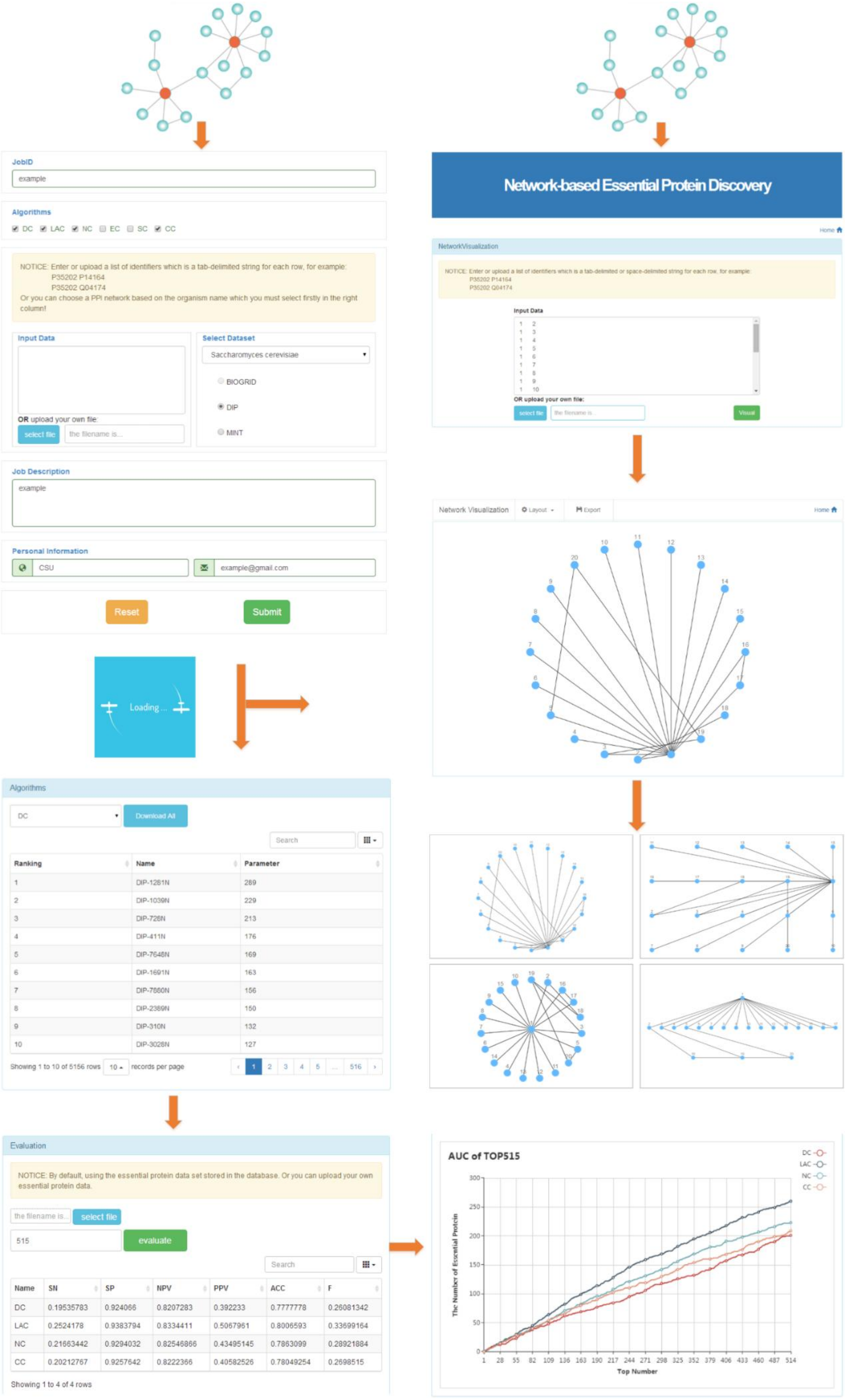
NetEPD: an integrated information platform for network-based essential protein discovery

Essential proteins are important participants in various life activities and play a vital role in the survival and reproduction of living body. In biological research field, essential proteins have been discovered through experimental methods, such as single gene knockouts, RNA interference and conditional knockouts. With the increase of protein-protein interaction (PPI) data gradually by high-throughput technologies, which can be modeled as a PPI network for proteomics research, the network-based essential protein discovery measures have been given more and more attention. And we can identify the essential protein through analyzing PPI network.

NetEPD based on the web service, which requires no configuration or installation, is very convenient to analyze the protein function and the essentiality of protein via web browser. It provides several common used PPI datasets which have been preprocessed before storing in the database, and supplies two ways for users to submit own PPI data. The integration of computational methods for detecting essential protein is a critical part of NetEPD. There are a variety of centrality measures and performance measures implemented in this platform to analyze the PPI network. In addition, the network visualization technique facilitates the researchers easily comprehending and analyzing the raw complex network data by translating the target data into computer graphics. This platform can be helpful for effective and efficient researching on PPI network. The workflow of NetEPD is shown in the below figure.





JobID

example

Algorithms

☑ DC ☑ LAC ☑ NC □ EC □ SC ☑ CC

NOTICE: Enter or upload a list of identifiers which is a tab-delimited string for each row, for example:
P35202 P14164
P35202 Q04174
Or you can choose a PPI network based on the organism name which you must select firstly in the right column!

Input Data

OR upload your own file:

select file | the filename is...

Select Dataset

Saccharomyces cerevisiae

○ BIOGRID

● DIP

○ MINT

Job Description

example

Personal Information

CSU

example@gmail.com

Reset | Submit

Loading ...

Algorithms

DC | Download All

Search

Ranking	Name	Parameter
1	DIP-1281N	289
2	DIP-1039N	229
3	DIP-728N	213
4	DIP-411N	176
5	DIP-7648N	169
6	DIP-1691N	163
7	DIP-7880N	156
8	DIP-2389N	150
9	DIP-310N	132
10	DIP-3028N	127

Showing 1 to 10 of 5156 rows | 10 | records per page

‹ 1 2 3 4 5 ... 516 ›

Evaluation

NOTICE: By default, using the essential protein data set stored in the database. Or you can upload your own essential protein data.

the filename is... | select file

515 | evaluate

Search

Name	SN	SP	NPV	PPV	ACC	F
DC	0.19535783	0.924066	0.8207283	0.392233	0.7777778	0.26081342
LAC	0.2524178	0.9383794	0.8334411	0.5067961	0.8006593	0.33699164
NC	0.21663442	0.9294032	0.82546866	0.43495145	0.7863099	0.28921884
CC	0.20212767	0.9257642	0.8222366	0.40582526	0.78049254	0.2698515

Showing 1 to 4 of 4 rows

Network-based Essential Protein Discovery

Home

NetworkVisualization

NOTICE: Enter or upload a list of identifiers which is a tab-delimited or space-delimited string for each row, for example:
P35202 P14164
P35202 Q04174

Input Data

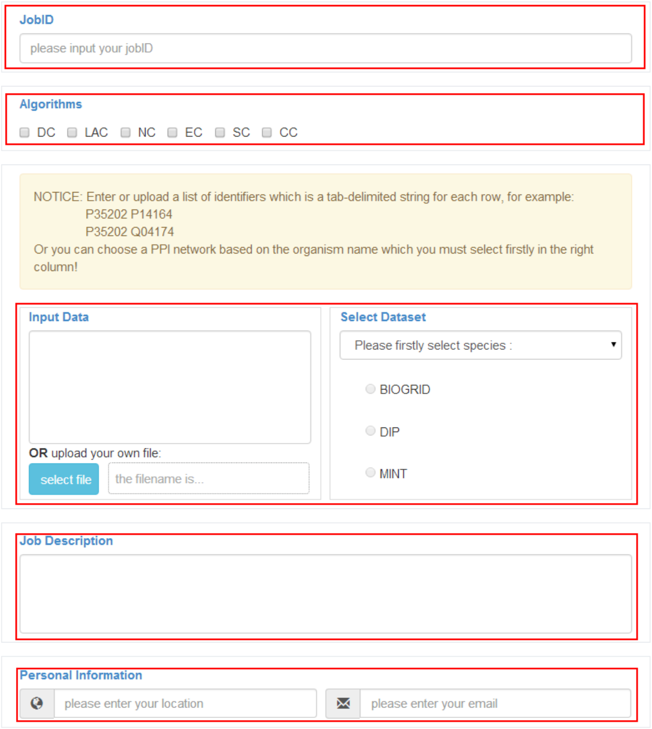
```
1	2
1	3
1	4
1	5
1	6
1	7
1	8
1	9
1	10
```

OR upload your own file:

select file | the filename is... | Visual

1. Create the job

When users start to use NetEPD, they should fill out the form as shown in the picture below, including: (1)JobID (the user's personal identification), (2)Algorithms (user's selection centrality measures for discovering essential protein), (3)PPI data, (4)JobDescription (the detailed description of this Job), (5)Location and Email (Only used for notifying users about their submitted job).



JobID

please input your jobID

Algorithms

□ DC □ LAC □ NC □ EC □ SC □ CC

NOTICE: Enter or upload a list of identifiers which is a tab-delimited string for each row, for example:
P35202 P14164
P35202 Q04174
Or you can choose a PPI network based on the organism name which you must select firstly in the right column!

Input Data

OR upload your own file:

select file | the filename is...

Select Dataset

Please firstly select species :

○ BIOGRID

○ DIP

○ MINT

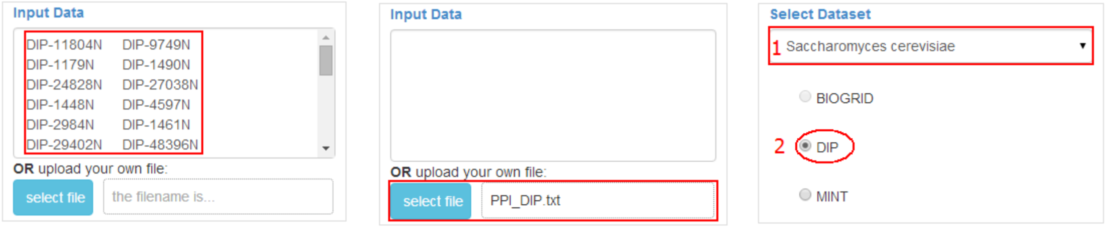
Job Description

Personal Information

please enter your location

please enter your email

The NetEPD can get PPI data in tab-delimited format, either a pasted data or a file (size limit < 2 MB). When users upload their own PPI data, they can paste the tab-delimited data directly into the text area or submit the data from a local file by clicking the select button. In addition, the integration of common used datasets (BioGrid, DIP and MINT) are available to users, and have been preprocessed to avoid the impact of redundant data. To use these dataset, users should select the organism firstly, and then choose which dataset you want to do PPI data analysis.



Input Data

```
DIP-11804N    DIP-9749N
DIP-1179N     DIP-1490N
DIP-24828N    DIP-27038N
DIP-1448N     DIP-4597N
DIP-2984N     DIP-1461N
DIP-29402N    DIP-48396N
```

OR upload your own file:

select file | the filename is...

Input Data

OR upload your own file:

select file | PPI_DIP.txt

Select Dataset

1 Saccharomyces cerevisiae

○ BIOGRID

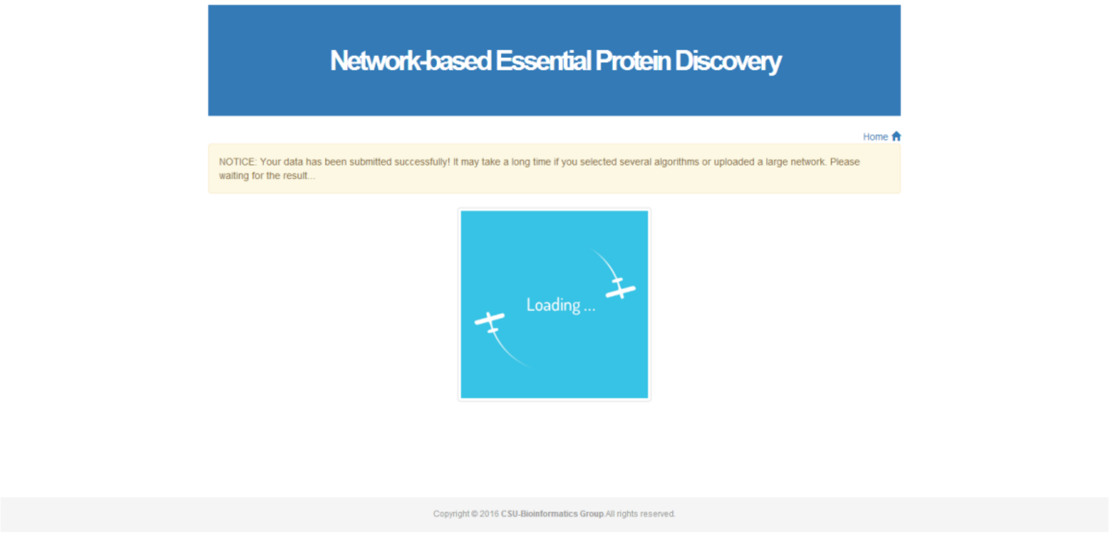
2 ● DIP

○ MINT

2. Wait for the result

If there are no errors reported after you click the submit button. Congratulations! You have submitted your job to NetEPD server successfully; Otherwise, please check if the input information are qualified, and correct the errors according to the required format and submit the job once again.

Next we are now redirecting to the waiting page from the home page as it is shown below.

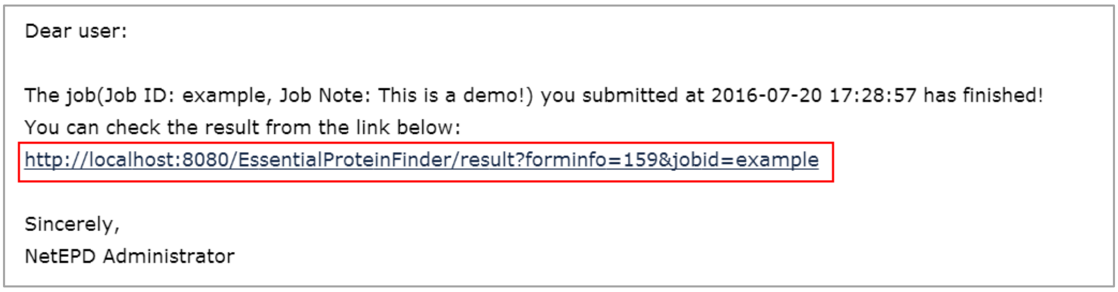


Network-based Essential Protein Discovery

NOTICE: Your data has been submitted successfully! It may take a long time if you selected several algorithms or uploaded a large network. Please waiting for the result...

Copyright © 2016 CSU-Bioinformatics Group.All rights reserved.

If the PPI data is small, it may jump directly to the result page in seconds or minutes; If not (such as thousands of proteins in PPI data), users may therefore have to wait for a long time depending on the size of the PPI data, before showing the computational result. The experimental results showed that, EC, SC and BC are able to take a long time to calculate each protein node's score in PPI network. Then users can close this waiting page after submitting the job successfully. As seen in the below picture, the notification will be sent to user's email address provided at the form in home page, when the job is completed. Users can check the result through clicking the hyperlink.



Dear user:

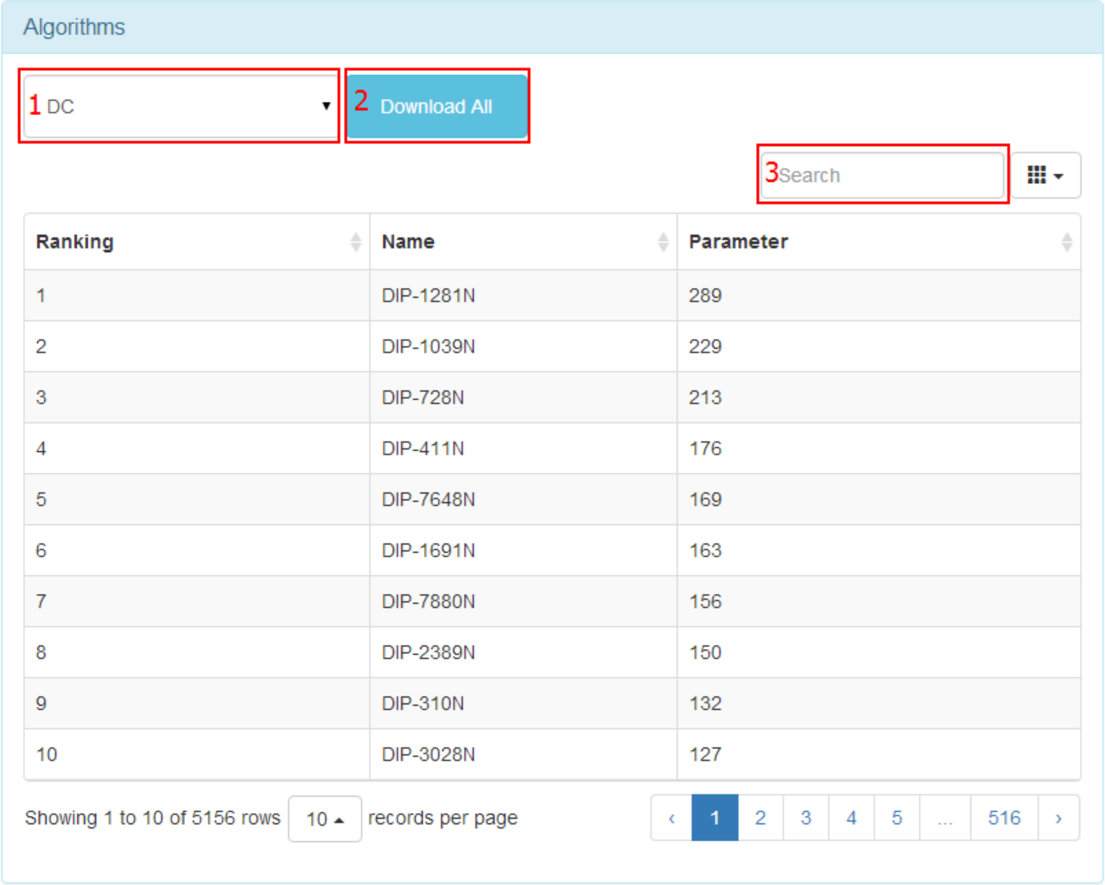
The job(Job ID: example, Job Note: This is a demo!) you submitted at 2016-07-20 17:28:57 has finished!
You can check the result from the link below:
http://localhost:8080/EssentialProteinFinder/result?forminfo=159&jobid=example

Sincerely,
NetEPD Administrator

3. Check the result

To better analyze PPI data, the result in NetEPD consists of the contents of three respects: the computational result based on user's selection algorithms, the result of performance measures and the network visualization.

At the module of the computational result as shown below, we rank proteins in decreasing order by each centrality measures. Users can use pull-down menu to change the algorithm for checking the score of each protein, and down all of the results in the form of EXCEL. In addition, user can search the result with specific protein's name.

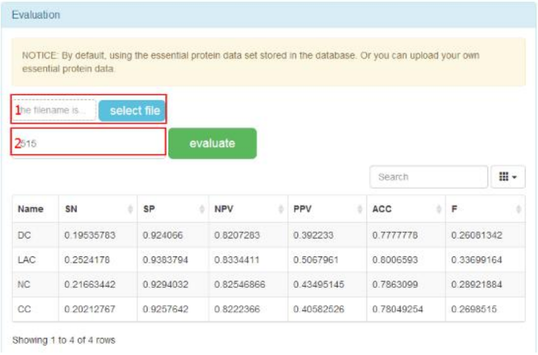


DC

Ranking	Name	Parameter
1	DIP-1281N	289
2	DIP-1039N	229
3	DIP-728N	213
4	DIP-411N	176
5	DIP-7648N	169
6	DIP-1691N	163
7	DIP-7880N	156
8	DIP-2389N	150
9	DIP-310N	132
10	DIP-3028N	127

Showing 1 to 10 of 5156 rows 10 records per page

Before checking the result of performance measures, users should provide their own dataset of essential proteins from the local file, in which each row represents an essential protein. On the other hand, there is no need for users to upload the file if they use the PPI data provided by NetEPD, because this platform has stored the information of essential protein collected and sorted out from DIP. But to determine top ranked proteins as predicted essential proteins (a candidate set), the top number is required as the size of candidate set. And the four performance measures are used to evaluate the result of identifying essential protein by each centrality measure. The algorithm with the highest value is the best. At the same time, we provide for users with line chart to intuitively show the performance of each centrality measure.



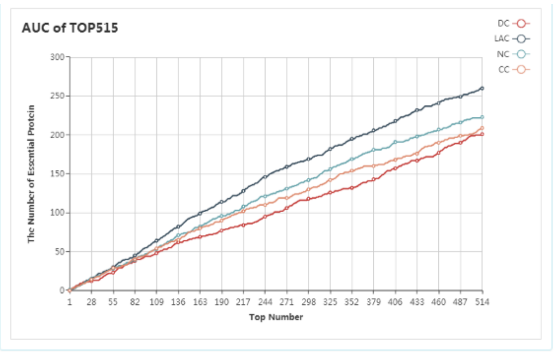
NOTICE: By default, using the essential protein data set stored in the database. Or you can upload your own essential protein data.

the filename is...

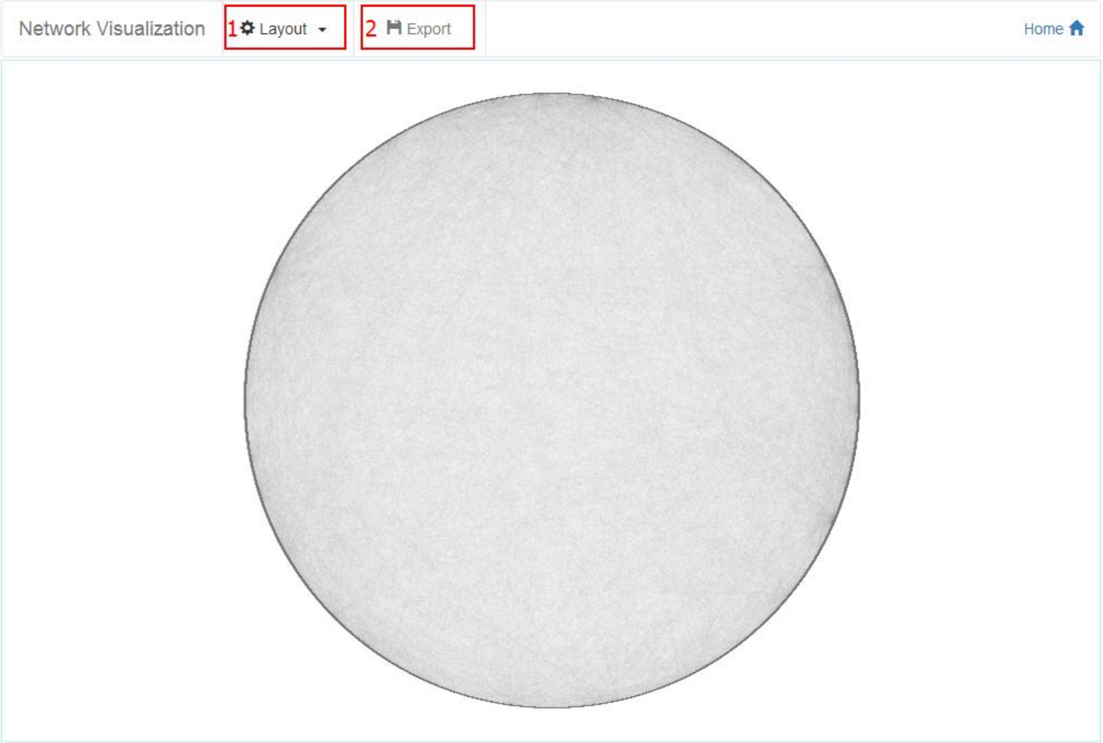
515

Name	SN	SP	NPV	PPV	ACC	F
DC	0.19535783	0.924066	0.8207283	0.392233	0.7777778	0.26081342
LAC	0.2524178	0.9383794	0.8334411	0.5067961	0.8006593	0.33699164
NC	0.21663442	0.9294032	0.82546866	0.43495145	0.7863099	0.28921884
CC	0.20212767	0.9257642	0.8222366	0.40582526	0.78049254	0.2698515

Showing 1 to 4 of 4 rows

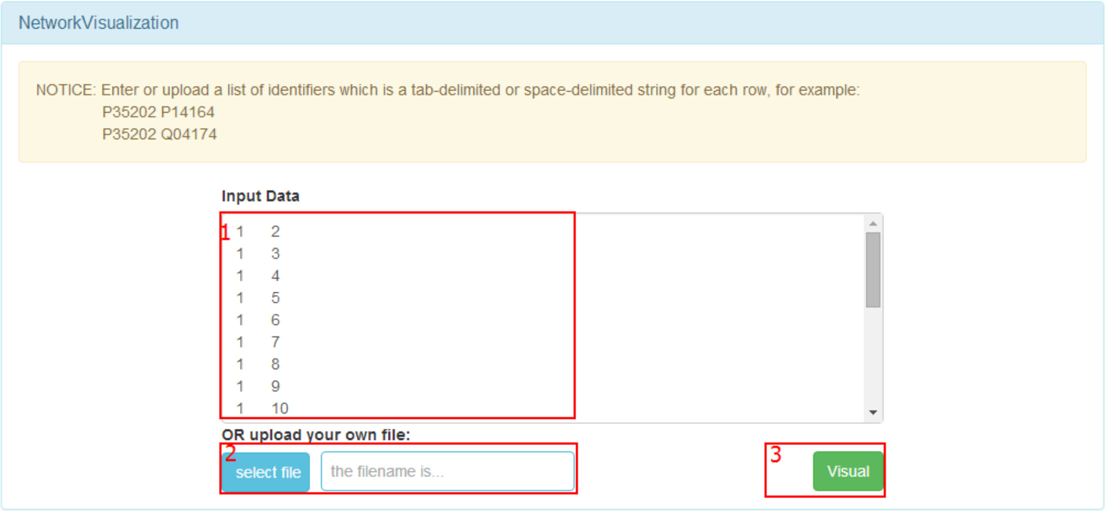


At last, user can view the structure of PPI network by clicking the network button, as shown below. Users not only can clearly view the network's structure (default display in circle layout), but also are able to change the type of layouts (the grid layout, the concentric layout, and the breadth-first layout). Users can interact with the graph to pan with the mouse and zoom in and out with the mouse wheel, and select the nodes and edges to mark them in different color. In addition, user can export the graph as an image of format PNG to save in the local workspace.



4. Network visualization

In order to do the visual analysis of PPI network conveniently, there is a web page for user to upload the PPI data directly as shown below. Users have two ways to submit the PPI data. The detailed requirements are same as what have mentioned above. And clicking the visual button, we can view the PPI network on the web.



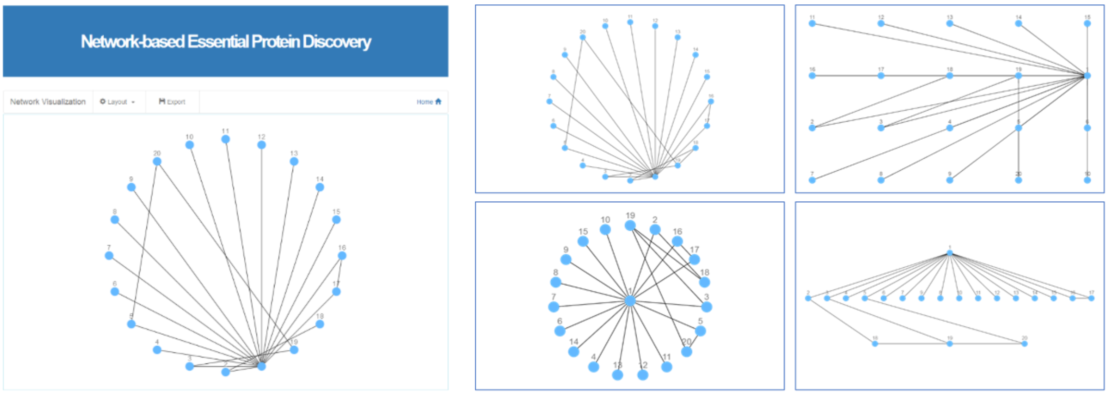
NOTICE: Enter or upload a list of identifiers which is a tab-delimited or space-delimited string for each row, for example:
P35202 P14164
P35202 Q04174

Input Data

```
1	2
1	3
1	4
1	5
1	6
1	7
1	8
1	9
1	10
```

OR upload your own file:

the filename is...


Network-based Essential Protein Discovery
Network Visualization
Layout
Export
Home