Construct, Analyze and Compare Biological Networks with InSyBio BioNets

InSyBio
Intelligent Systems Biology

User Manual

September 2023

Insybio Suite v3.1

www.insybio.com

InSyBio BioNets

BioNets is a biological networks analysis tool for the

- Preprocessing and analysis of gene expression data
- Biomarker discovery through differential expression analysis and/or network comparison
- Preprocessing, meta-analysis and visualization of biological networks

The execution of complex biological processes requires the precise interaction and regulation of thousands of molecules. Systematic approaches to study large numbers of proteins, metabolites, and their modifications have revealed complex molecular networks. Network representation of intracellular biological networks typically considers molecular components within a cell as nodes and their direct or indirect interactions as edges. Network representation enables integration of data from many different studies into a single framework.

Biological networks are significantly different from random networks and often exhibit ubiquitous properties in terms of their structure and organization. The analysis of these networks provides novel insights in understanding basic cellular mechanisms and the mechanisms of disease pathologies.

There exist different categories of biological networks. With InSyBio BioNets users can handle undirected weighted biological networks. The main undirected biological networks are gene co-expression and protein-protein interaction networks. InSyBio BioNets can also analyze directed biological networks but the edges' direction will be ignored.

In gene co-expression networks, each node is a gene and each edge indicates a correlation between the expression profiles of the two-node genes it connects. Gene co-expression networks can be used to find genes with similar functional properties and to uncover biomarkers through the comparison of networks derived from different states (e.g. control vs disease states). Gene co-expression networks can be

constructed by using gene expression experimental techniques, such as microarray or RNA-seq experiments.

Protein-protein Interaction (PPI) networks consist of nodes which represent proteins and edges which represent the physical or functional interactions between the nodes-proteins. PPI networks can be constructed using experimental (e.g. Yeast2hybrid technique [1], Tandem Affinity Purification [2] and so on) or computational techniques. The edges of the PPI networks can have weights representing the strength or the confidence score of the interaction. PPI networks can be used to functionally characterize proteins, predict protein complexes and protein biomarkers when comparing PPI networks from different cellular states to locate significantly altered regions of the network.

BioNets provides a set of tools for the construction, preprocessing, meta-analysis and visualization of biological networks. Moreover, additional tools for parsing gene expression files, creating gene co-expression files and uncovering differential expression biomarkers have been added to InSyBio BioNets to enable the construction and analysis of gene co-expression networks. Additional tools are also offered to ease InSyBio BioNets users. Specifically, the users can perform time consuming analysis steps and analyze large biological networks and large gene expression files using our user friendly job scheduling mechanism. Regarding the uncovered biomarkers, users have access to informative biomarker reports which include links to other publicly available databases for these biomarkers, links to our PPI interaction repository (InSyBio Interact) and information about the prior knowledge on associating these biomarkers to diseases.

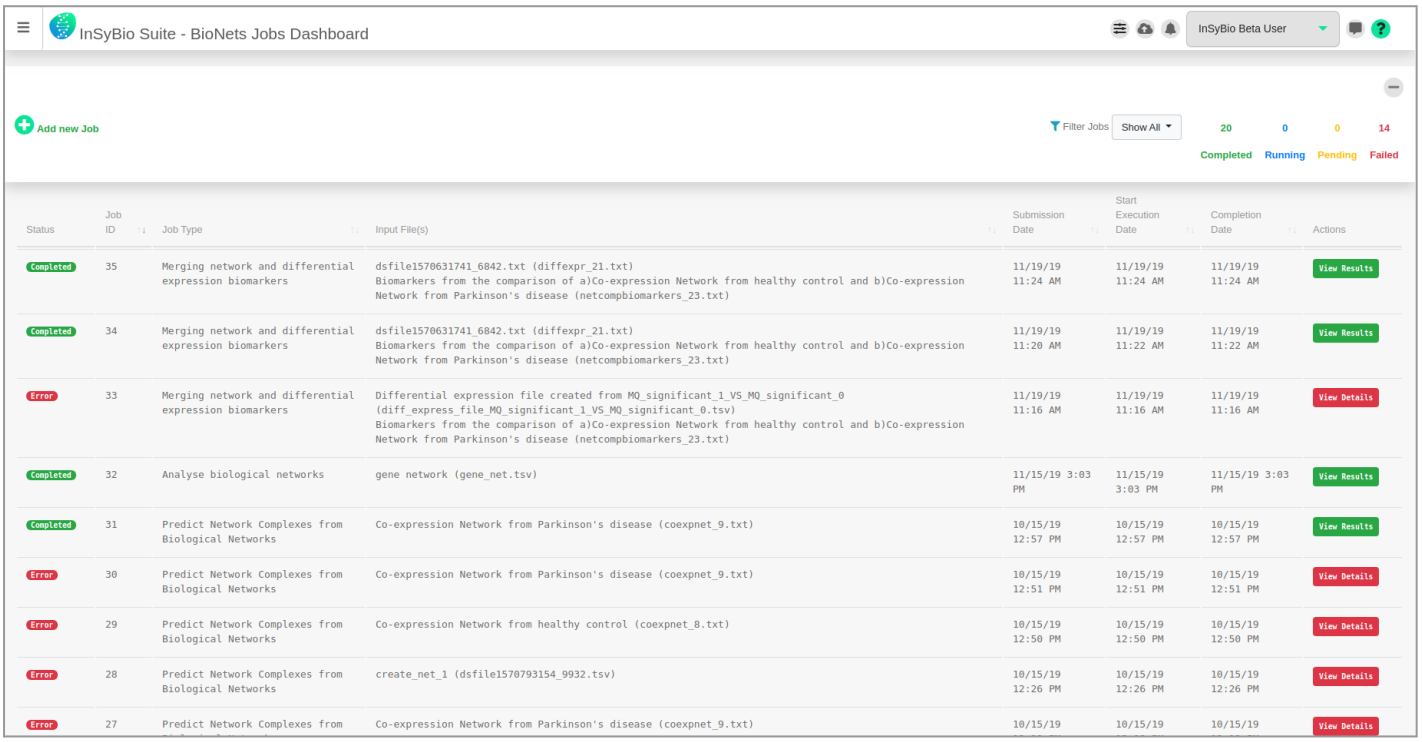
With InSyBio BioNets you can:

1. Parse .soft files
2. Parse gene expression data files
3. Construct gene co-expression networks from gene expression data
4. Uncover biomarkers using differential expression analysis
5. Uncover biomarkers using network comparison analysis
6. Merge differential expression biomarkers with network comparison biomarkers

7. Access informative biomarker reports on the extracted biomarkers
8. Analyze biological networks to control their quality, find significant nodes and edges, conduct shortest path analysis and so on.
9. Visualize biological networks
10. Predict clusters (e.g. protein complexes) from a biological network (e.g. Protein-Protein Interaction graph)
11. Combine networks
12. Analyze large biological networks and large gene expression files using our user friendly job scheduling mechanism

BioNets Jobs Dashboard

The analysis of biological networks and of gene expression data includes some time intensive steps. To deal with this complexity, InSyBio BioNets offers a simple and user friendly Jobs Dashboard. Users can start as many jobs as they need and monitor the progress of their jobs using BioNets Jobs Dashboard. Metadata concerning the execution of the analysis task (start date, duration, job type, and so on) are presented in the Dashboard. When a job is completed, the results can be accessed through the View Results link.



Gene Expression Data Parsing, Preprocessing and Analysis

BioNets offers a set of tools for handling, preprocessing and analyzing gene expression data in order to construct gene co-expression networks, and predict biomarkers.

SOFT Files Parsing

InSyBio BioNets supports the universally accepted format for gene expression data named SOFT. Simple Omnibus Format in Text (SOFT) is the format supported by Gene Expression Omnibus database [13] (http://www.ncbi.nlm.nih.gov/geo/) and it is the prevalent format for gene expression experiments. It is a simple line-based, plain text format, meaning that SOFT files may be readily generated from common spreadsheet and database applications. A single SOFT file can hold both data tables and accompanying descriptive information for multiple, concatenated Platforms, Samples, and/or Series records.

InSyBio BioNets soft parsing includes the following preprocessing steps:

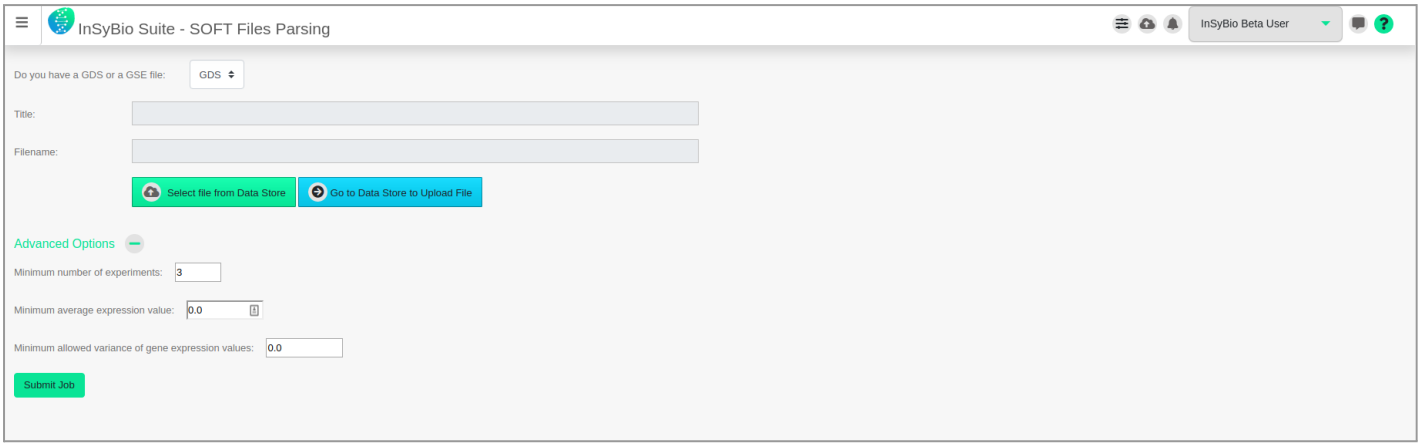
- logarithmic normalization (if the expression values are not normalized),
- missing values estimation with the knn-impute method,
- filtering using minimum average expression values and minimum expression values variance filters.

The different experimental states (conditions) defined in the SOFT file will be automatically recognized and a gene expression tab delimited file will be constructed for each state.

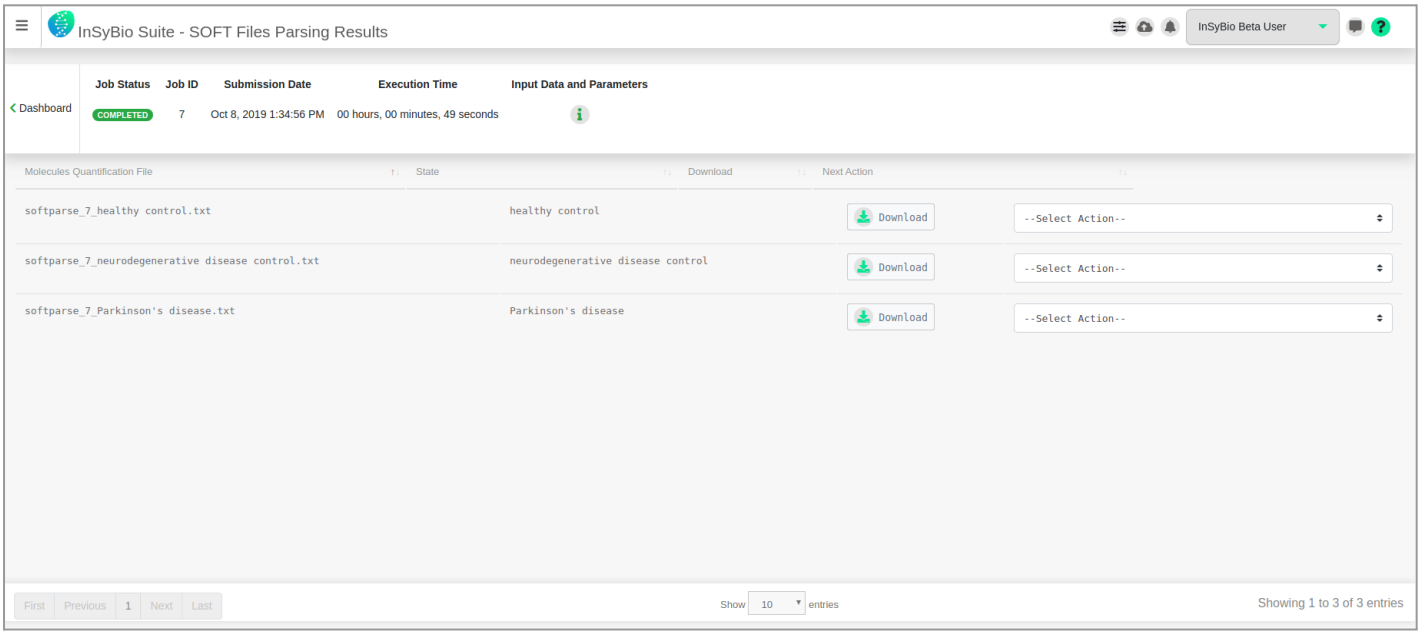
Users are enabled either to use default parameter values or to tune the following parameters:

- Minimum number of experiments in every single experimental state (default value 3)

- Minimum average expression value (default value 0.0)
- Minimum allowed variance of gene expression values (default value 0.0)



When a SOFT file is parsed, the users can further analyze the extracted tab delimited gene expression files. They can use them either to create a co-expression network or to extract biomarkers using differential expression analysis.



Upload gene expression files

Users can generate gene expression files using their own methods and directly upload them to InSyBio DataStore. The uploaded files should be tab delimited files

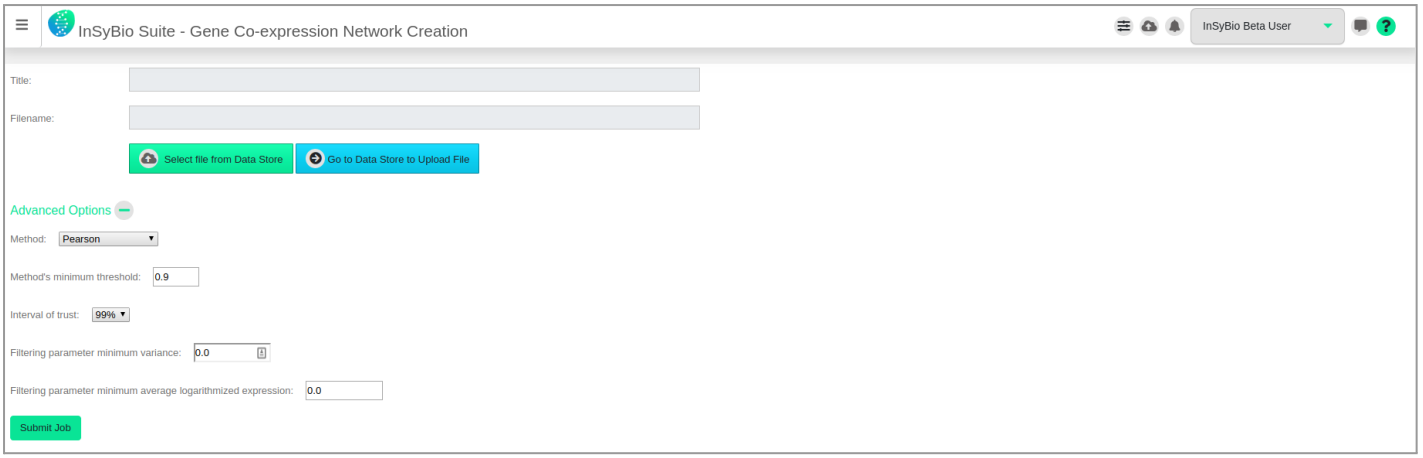
with the first column reporting the gene's symbol or name and the other columns being the expression values in different experiments, conditions and/or samples.

Gene co-expression network creation

Gene expression files, either uploaded directly by the users or generated through soft files parsing, can be used to generate weighted gene co-expression networks. Experienced users can tune the parameters of the algorithms used for this step and select the most suitable algorithm for them. Three options are offered:

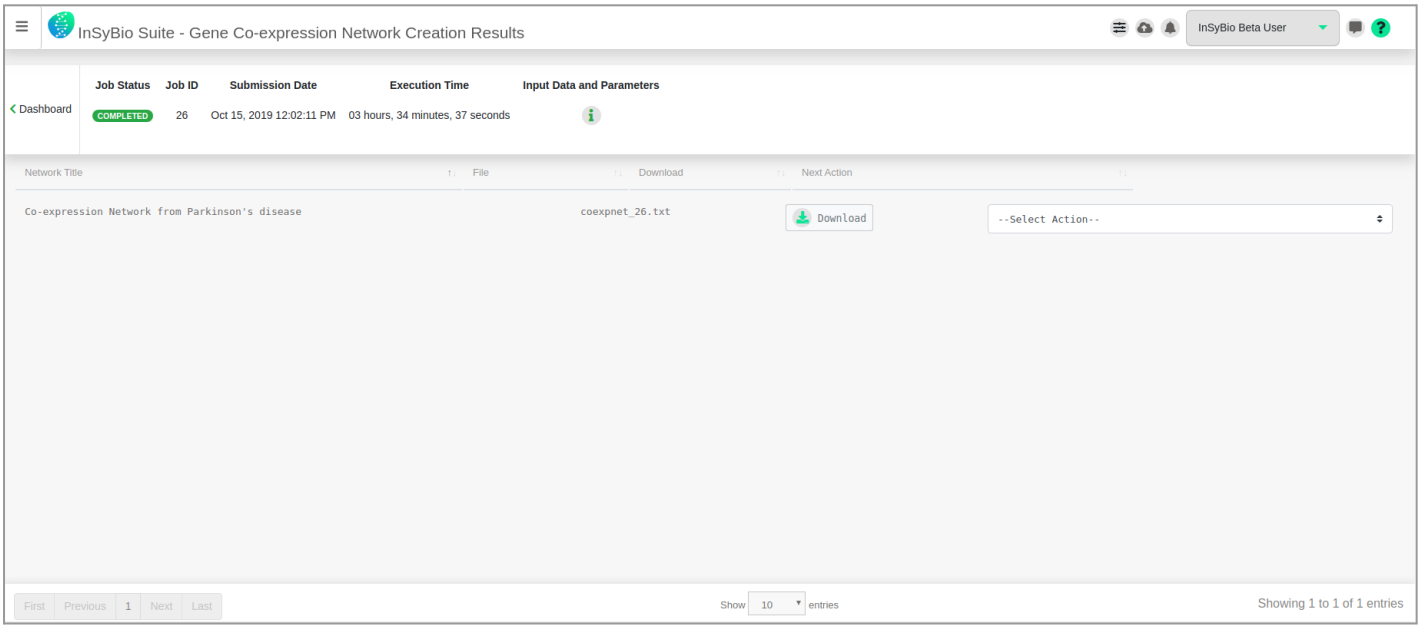
- **Pearson Correlation [14]:** This method adds an edge to a network if the Pearson correlation of the nodes adjacent to the edge exceeds a threshold.
- **Mutual information [15]:** This method adds an edge to a network if the mutual information among the expression profiles of the two nodes of the edge exceeds a threshold.
- **Spearman Correlation [23]:** This method adds an edge to a network if the Spearman correlation of the nodes adjacent to the edge exceeds a threshold.

The thresholds for adding edges are dynamically generated to alleviate problems occurring by using the same threshold for all nodes. In particular, for a single node Pearson correlations or Mutual Information or Spearman correlations between this node and all other nodes are calculated. Assuming that the Pearson correlation/Mutual Information/Spearman correlation values between a single node and all other nodes follow a normal distribution, then the threshold for adding edges is selected to be in a predefined a confidence interval (90%, 95% or 99%). The confidence interval is predefined at 99% but the users can change this value in order to get denser or sparser networks. In order to force nodes to have a minimum number of edges users can also specify a minimum value for the threshold of adding an edge in the network. Experienced users can further filter nodes from the network by altering the minimum expression variance threshold and the minimum average of the logarithmized expression values threshold.



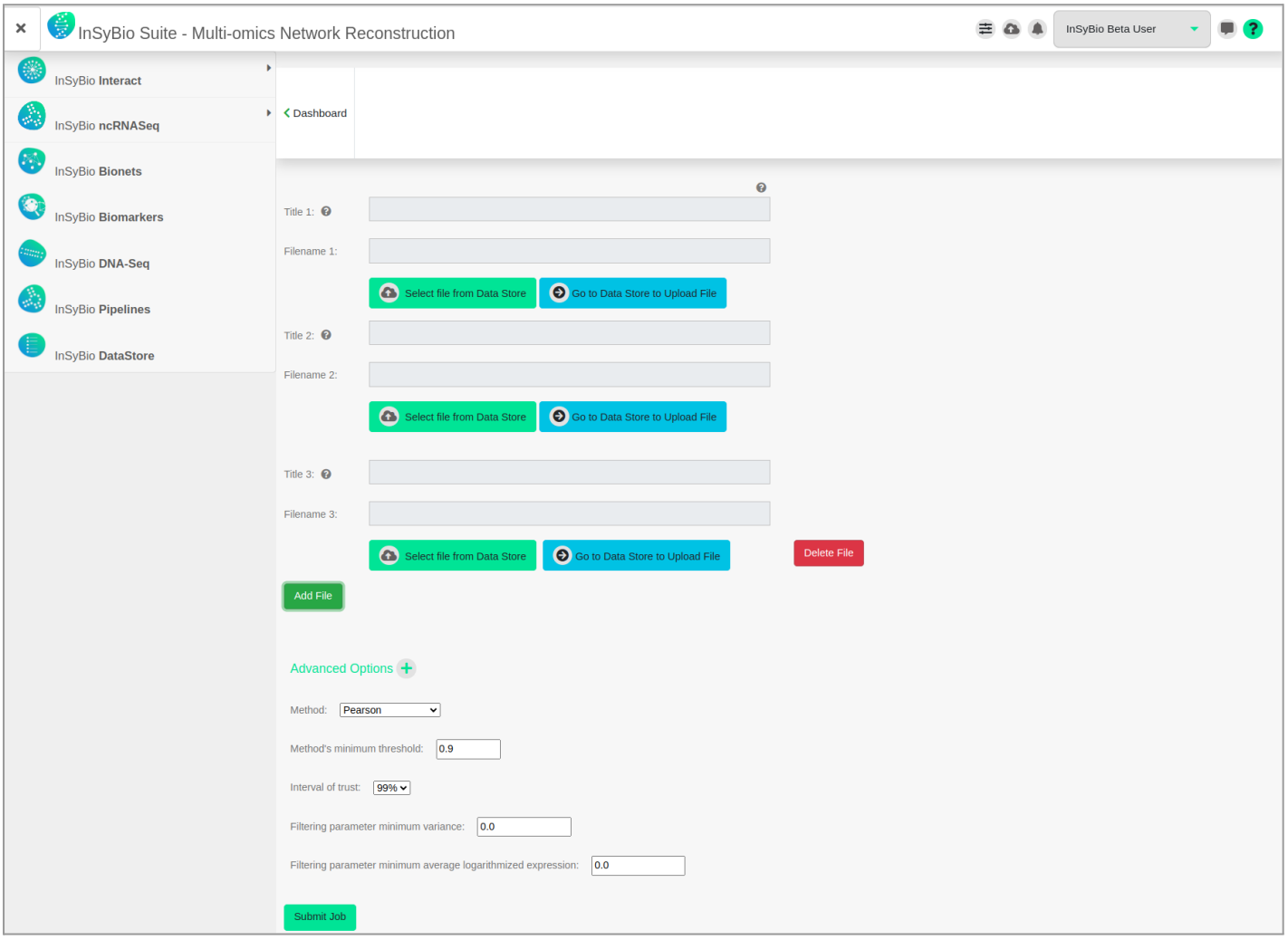
View Results

After starting an analysis job you can go to “BioNets Jobs Dashboard”, where you can view the status of your current and previous BioNets jobs. After the completion of the analysis you can select the “View Results” at the Actions column and view the created network.



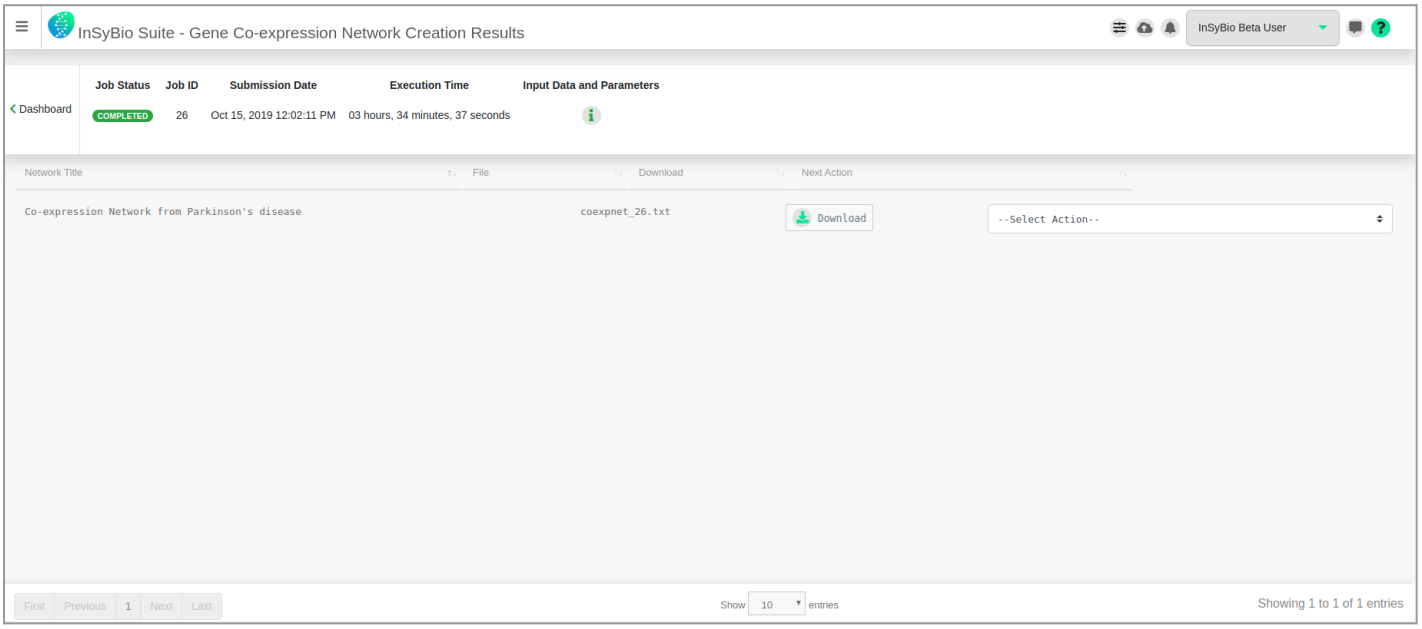
Multi-omics Network Reconstruction

Multi-omics network reconstruction combines two or more gene expression files, either uploaded directly by the users or generated through soft files parsing, in order to generate weighted gene co-expression networks. The job has the same prerequisites as the “Gene co-expression network creation” job.



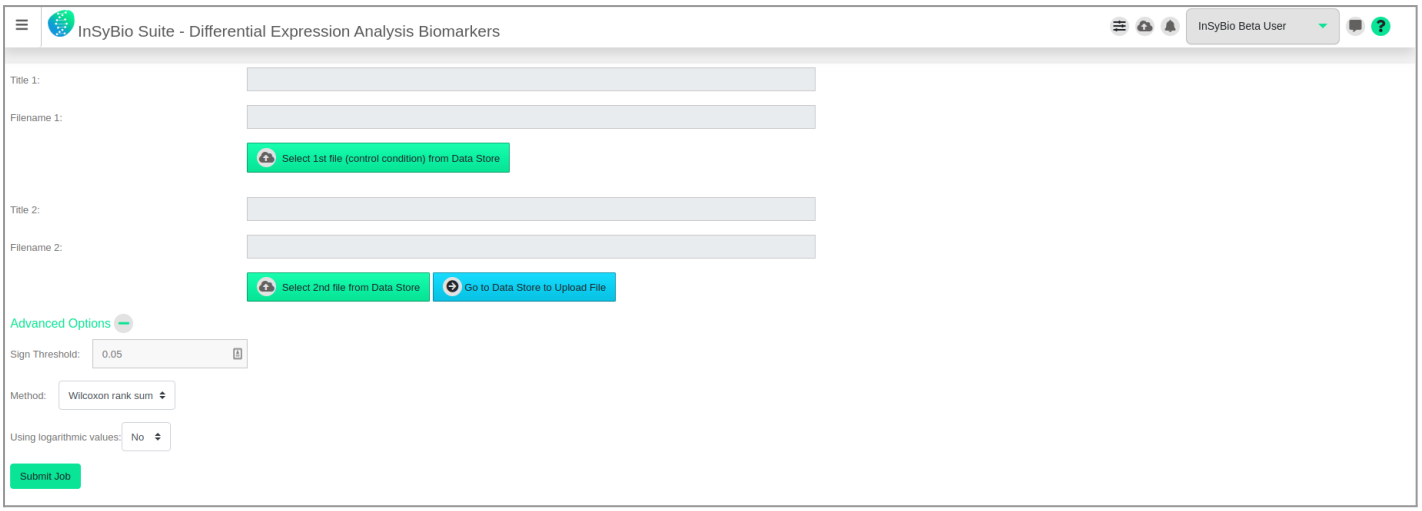
View Results

After starting an analysis job you can go to “BioNets Jobs Dashboard”, where you can view the status of your current and previous BioNets jobs. After the completion of the analysis you can select the “View Results” at the Actions column and view the created network. The results direct the user to the “Gene Co-expression Network Creation Results” view.



Differential Expression Analysis Biomarkers

Another action the users can perform if they have two gene expression files is the differential expression analysis to uncover biomarkers. Experienced users can select among T-Test [16] and Wilcoxon Rank Sum [17] statistical methods, stating their p-value thresholds. Default method and threshold proposed by InSyBio are Wilcoxon Rank Sum with p-value threshold equal to 0.05. Bonferoni corrections [18] are by default applied in the computation of a p-value to reduce the number of false positive predictions. Also the users can choose to use logarithmic values.

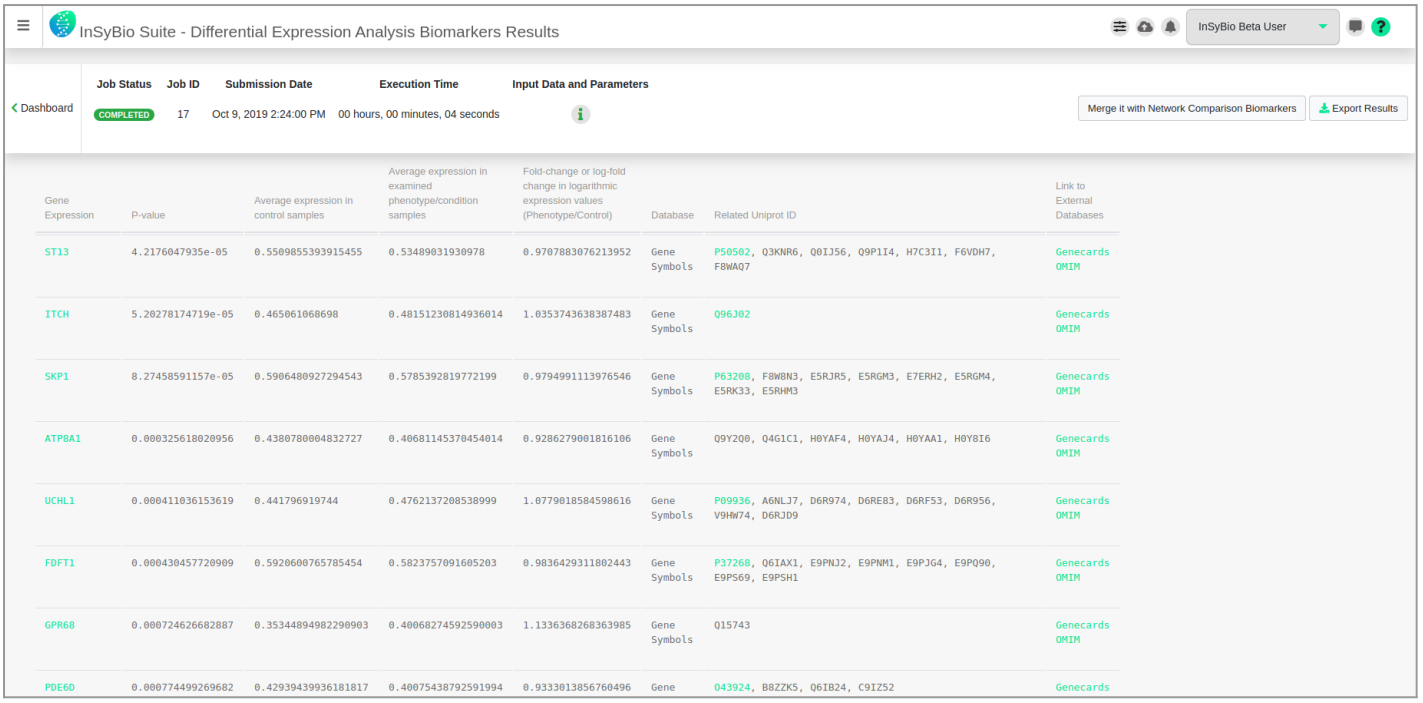


View Results

After starting an analysis job you can go to “BioNets Jobs Dashboard”, where you can view the status of your current and previous BioNets jobs. After the completion of the analysis you can select the “View Results” at the Actions column and view the differential expression analysis biomarkers.

The results are presented in a table with Gene, P-value, Average expression in control samples, Average expression in examined phenotype/condition samples, Fold-change or log-fold change in logarithmic expression values (Phenotype/Control), Database, Related Uniprot ID, Link to External Databases columns. Clicking a Gene Expression field the user can view diseases associated with that gene. Clicking a Related Uniprot ID field the user can view the related protein in our InSyBio Interact tool. Clicking a Link to External Databases the user can view the gene in external databases.

The users are enabled to download detailed reports about their extracted biomarkers.



Gene Expression	P-value	Average expression in control samples	Average expression in examined phenotype/condition samples	Fold-change or log-fold change in logarithmic expression values (Phenotype/Control)	Database	Related Uniprot ID	Link to External Databases
ST13	4.2176047935e-05	0.5509855393915455	0.53489031930978	0.9707883076213952	Gene Symbols	P50502, Q3KNR6, Q0IJ56, Q9P1I4, H7C3I1, F6VDH7, F8WAQ7	Genecards OMIM
ITCH	5.20278174719e-05	0.465061068698	0.48151230814936014	1.0353743638387483	Gene Symbols	Q96J02	Genecards OMIM
SKP1	8.27458591157e-05	0.5906480927294543	0.5785392819772199	0.9794991113976546	Gene Symbols	P63208, F8W8N3, E5RJR5, E5RGM3, E7ERH2, E5RGM4, E5RK33, E5RHM3	Genecards OMIM
ATP8A1	0.000325618020956	0.4380780004832727	0.40681145370454014	0.9286279001816106	Gene Symbols	Q9Y2Q0, Q4G1C1, H0YAF4, H0YAJ4, H0YAA1, H0Y8I6	Genecards OMIM
UCHL1	0.000411036153619	0.441796919744	0.4762137208538999	1.0779018584598616	Gene Symbols	P09936, A6NLJ7, D6R974, D6RE83, D6RF53, D6R956, V9HW74, D6RJD9	Genecards OMIM
FDFT1	0.000430457720909	0.5920600765785454	0.5823757091605203	0.9836429311802443	Gene Symbols	P37268, Q6IAX1, E9PNJ2, E9PNM1, E9PJG4, E9PQ90, E9PS69, E9PSH1	Genecards OMIM
GPR68	0.000724626682887	0.35344894982290903	0.40068274592590003	1.1336368268363985	Gene Symbols	Q15743	Genecards OMIM
PDE6D	0.000774499269682	0.42939439936181817	0.40075438792591994	0.9333013856760496	Gene	O43924, B8ZZK5, Q6IB24, C9IZ52	Genecards

Biomarker Discovery

Using InSyBio BioNets, users can uncover biomarkers by applying:

- Differential Expression Analysis on gene expression files
- Biological Networks Comparison
- Differential Expression Biomarkers and Network Based Biomarkers merging

InSyBio BioNets automatically detects the type of symbols used for the network's nodes. Most known symbol types are supported by InSyBio to generate advanced reports including:

- Gene Symbols,
- Uniprot ids,
- Gene ids,
- EMBL ids,
- Refseq ids,
- RefseqNT ids,
- Kegg ids,
- Reactome ids,
- and many more (this list is continuously updated)

When the symbol's type is detected, then the biomarkers report provides information about the significance of the biomarker, links to InSyBio Interact Tool about the proteins being related with this biomarker, links to Genecards [19] and OMIM [20] and information about prior knowledge associating this biomarker with diseases with information mined from DisGeNet database [21].

Differential expression biomarkers are measured with a single p-value and network based biomarkers with a confidence score. Users have the option to combine the two experiments by predicting combined biomarkers and there BioNets uses a combined confidence score. The combined biomarkers have significantly different expression profiles on the two examined conditions while their role in the network is significantly altered.

Differential Expression Analysis Biomarkers

Differential expression analysis predicts differentially expressed biomarkers. See above.

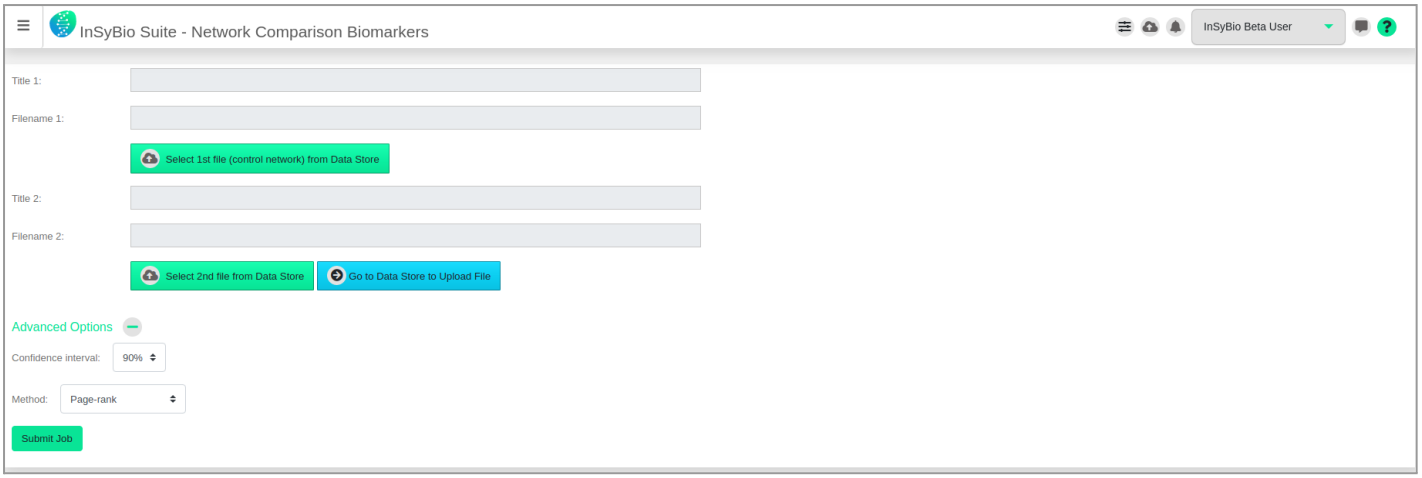
Network Comparison Biomarkers

It is widely accepted lately that differential expression biomarkers are large in numbers, contain a large number of false positives and mainly depict the outcome of disease mechanism and not its cause. For this reason, the current trend in biomarker discovery is to detect biomarkers by comparing biological networks. Biological networks are slightly altered in different biological conditions and changes on them are associated with the causes of disease mechanisms with high probability.

When having two biological networks of different conditions, users can use them to predict network biomarkers with an InSyBio's novel methodology. In particular, a certain network metric is selected and InSyBio BioNets attempts to detect network's nodes with significantly altered values for this network metric. Thus, our approach finds nodes whose role in the network has significantly changed among the different conditions. Experienced users can select a specific network metric among the following ones:

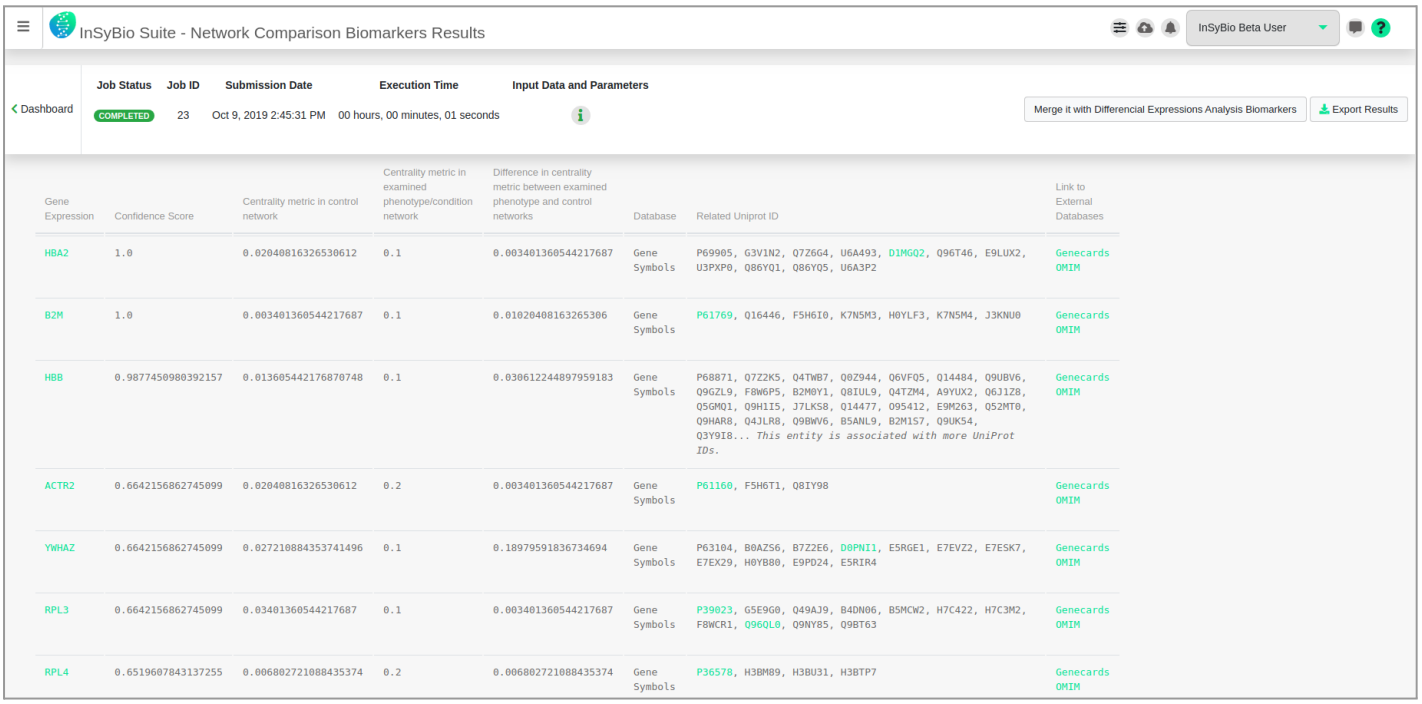
- Degree Centrality [9]
- Clustering Coefficient [8]
- Pagerank method [6, 7]

Pagerank method is the default one. This method triggers random walkers starting from each node. Significant nodes are collecting more information from the diffused quantities of the random walkers over time. Experienced users can also select the confidence interval for tuning the threshold of assigning a node as biomarker. Higher confidence interval values lead to the extraction of more compact sets of biomarkers.



View Results

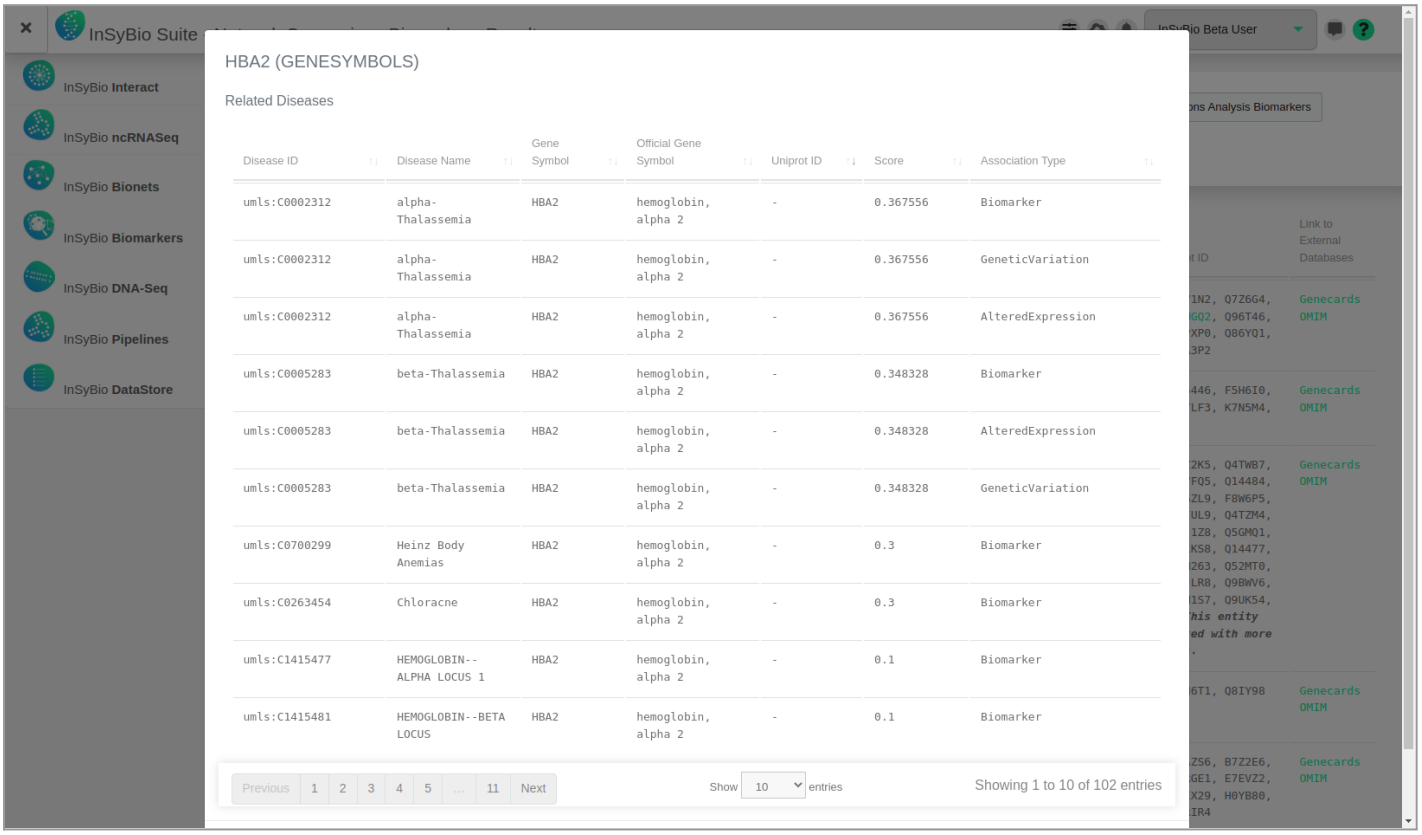
After starting an analysis job you can go to “BioNets Jobs Dashboard”, where you can view the status of your current and previous BioNets jobs. After the completion of the analysis you can select the “View Results” at the Actions column and view the network comparison biomarkers.



The results are presented in a table with Gene, Confidence Score, Centrality metric in control network, Centrality metric in examined phenotype/condition network, Difference in centrality metric between examined phenotype and control networks,

Database, Related Uniprot ID, Link to External Databases columns. Clicking a Gene Expression field the user can view diseases associated with that gene. Clicking a Related Uniprot ID field the user can view the related protein in our InSyBio Interact tool. Clicking a Link to External Databases the user can view the gene in external databases.

The users are allowed to download detailed reports about their extracted biomarkers.



Merge Biomarkers

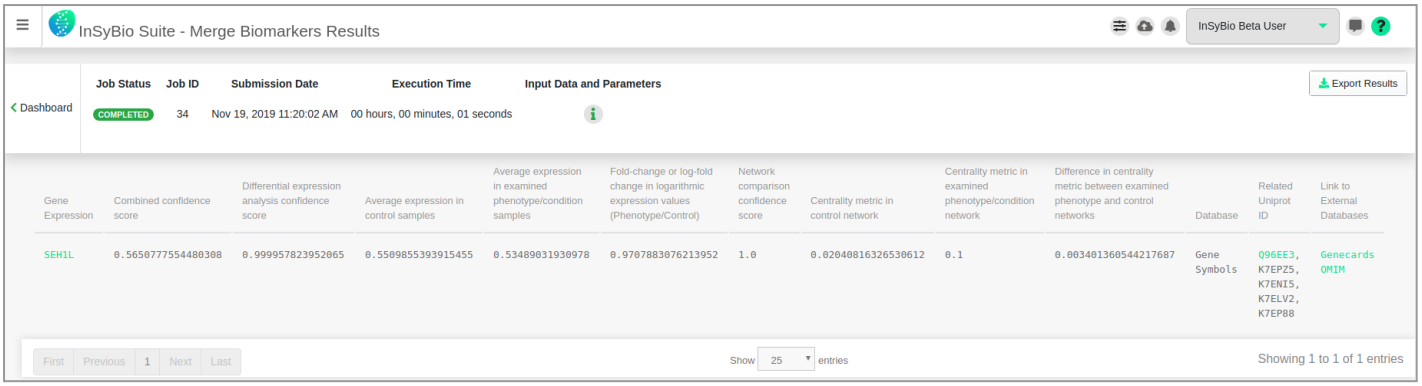
Users can select to merge the biomarker results of differential expression analysis and network comparison analysis. In fact the real biomarkers should have different expression profiles in the examined biological conditions and their role in the biological networks should be significantly altered in order to reassure that they are the real cause of biological variation among the condition and not a result of this variation.

View Results

After starting an analysis job you can go to “BioNets Jobs Dashboard”, where you can view the status of your current and previous BioNets jobs. After the completion of the analysis you can select the “View Results” at the Actions column and view the merged biomarkers.

The results are presented at a table with Gene, Combined confidence score, Differential expression analysis confidence score, Average expression in control samples, Average expression in examined phenotype/condition samples, Fold-change or log-fold change in logarithmic expression values (Phenotype/Control), Network comparison confidence score, Centrality metric in control network, Centrality metric in examined phenotype/condition network, Difference in centrality metric between examined phenotype and control networks, Database, Related Uniprot ID, Link to External Databases columns. Clicking a Gene Expression field the user can view diseases associated with that gene. Clicking a Related Uniprot ID field the user can view the related protein in our InSyBio Interact tool. Clicking a Link to External Databases the user can view the gene in external databases.

The users are allowed to download detailed reports about their extracted biomarkers.

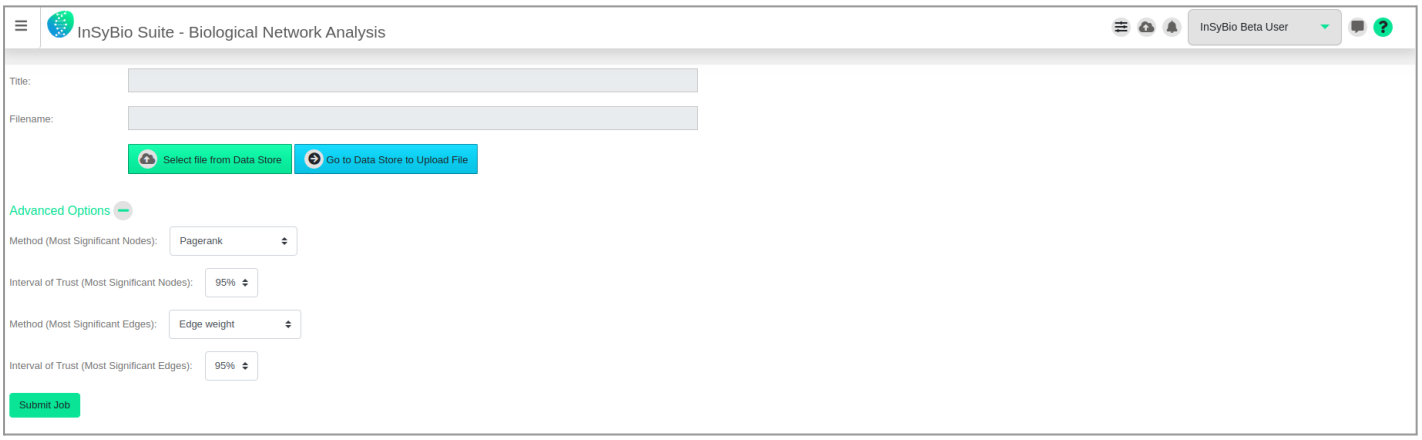

InSyBio Suite - Merge Biomarkers Results
InSyBio Beta User
Dashboard
Export Results
Job Status
Job ID
Submission Date
Execution Time
Input Data and Parameters
COMPLETED
34
Nov 19, 2019 11:20:02 AM
00 hours, 00 minutes, 01 seconds
Gene Expression
Combined confidence score
Differential expression analysis confidence score
Average expression in control samples
Average expression in examined phenotype/condition samples
Fold-change or log-fold change in logarithmic expression values (Phenotype/Control)
Network comparison confidence score
Centrality metric in control network
Centrality metric in examined phenotype/condition network
Difference in centrality metric between examined phenotype and control networks
Database
Related Uniprot ID
Link to External Databases
SEH1L
0.5650777554480308
0.999957823952065
0.5509855393915455
0.53489031930978
0.9707883076213952
1.0
0.02040816326530612
0.1
0.003401360544217687
Gene Symbols
Q96EE3, K7EPZ5, K7ENI5, K7ELV2, K7EP88
Genecards OMIM
First
Previous
1
Next
Last
Show
25
entries
Showing 1 to 1 of 1 entries

Biological Networks Preprocessing and Analysis

Biological Network Analysis

When users create or upload a biological network, they can access a menu of six analytical options described below. In order to analyze a biological network the users should:

- Select a biological network file from the ones in InSyBio DataStore or upload a new biological network file using InSyBio DataStore.
- If they are experienced, they can tune the following parameters:
  - Method for selecting significant nodes (Pagerank (default) [6, 7], Clustering Coefficient [8], degree centrality [9])
  - Confidence interval for locating significant nodes
  - Method for selecting significant edges (Edge weight (default), Inbetweeness centrality [10])
  - Confidence interval for locating significant edges



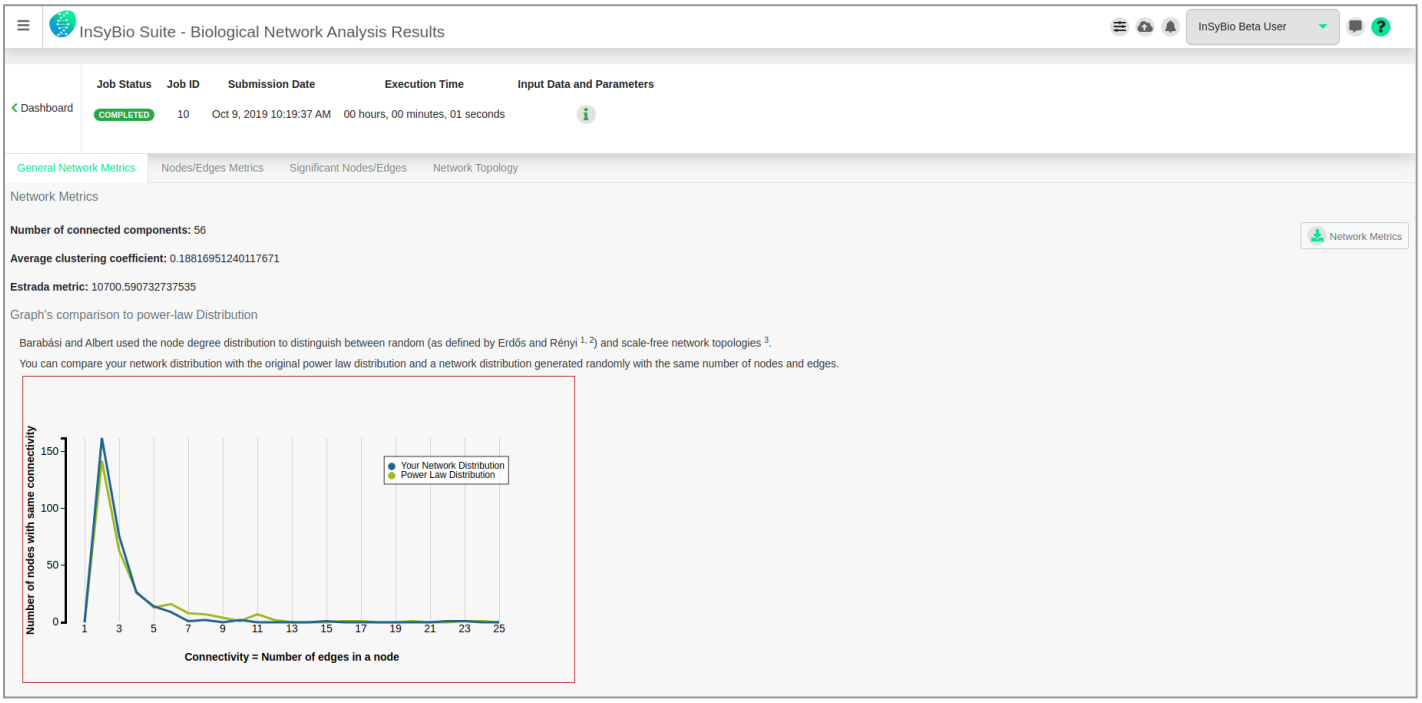
View Results

After starting an analysis job you can go to “BioNets Jobs Dashboard”, where you can view the status of your current and previous BioNets jobs. After the completion of

the analysis you can select the “View Results” at the Actions column and view the Network Analysis details.

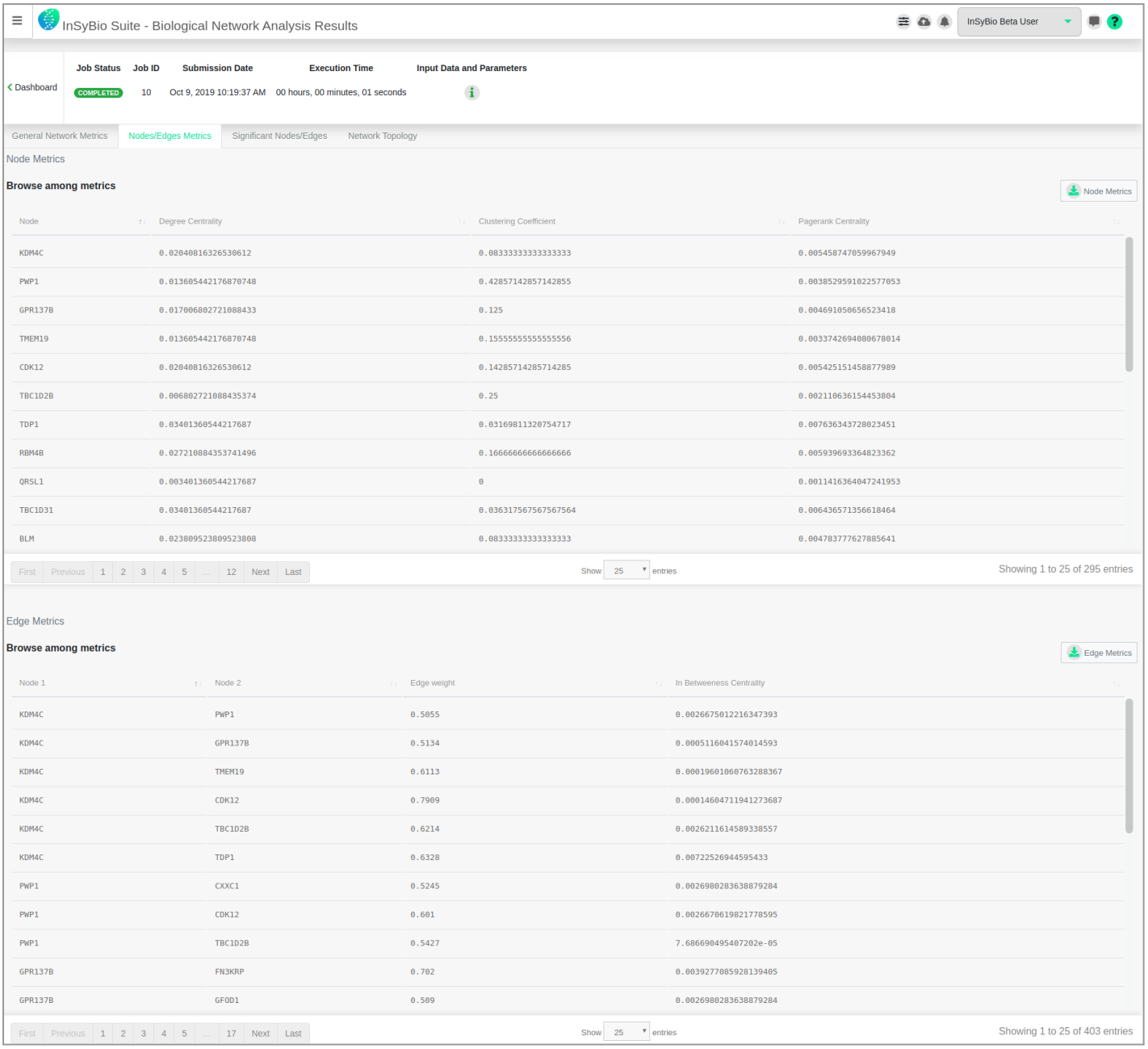
General Network Analysis

In the General Network Analysis tab users can view the most significant network metrics (clustering coefficient [8], Estrada index [11] and so on) and compare the degree distribution of their network with the a random network’s power law distribution. Information in this tab is not available for networks with more than 225000 edges.



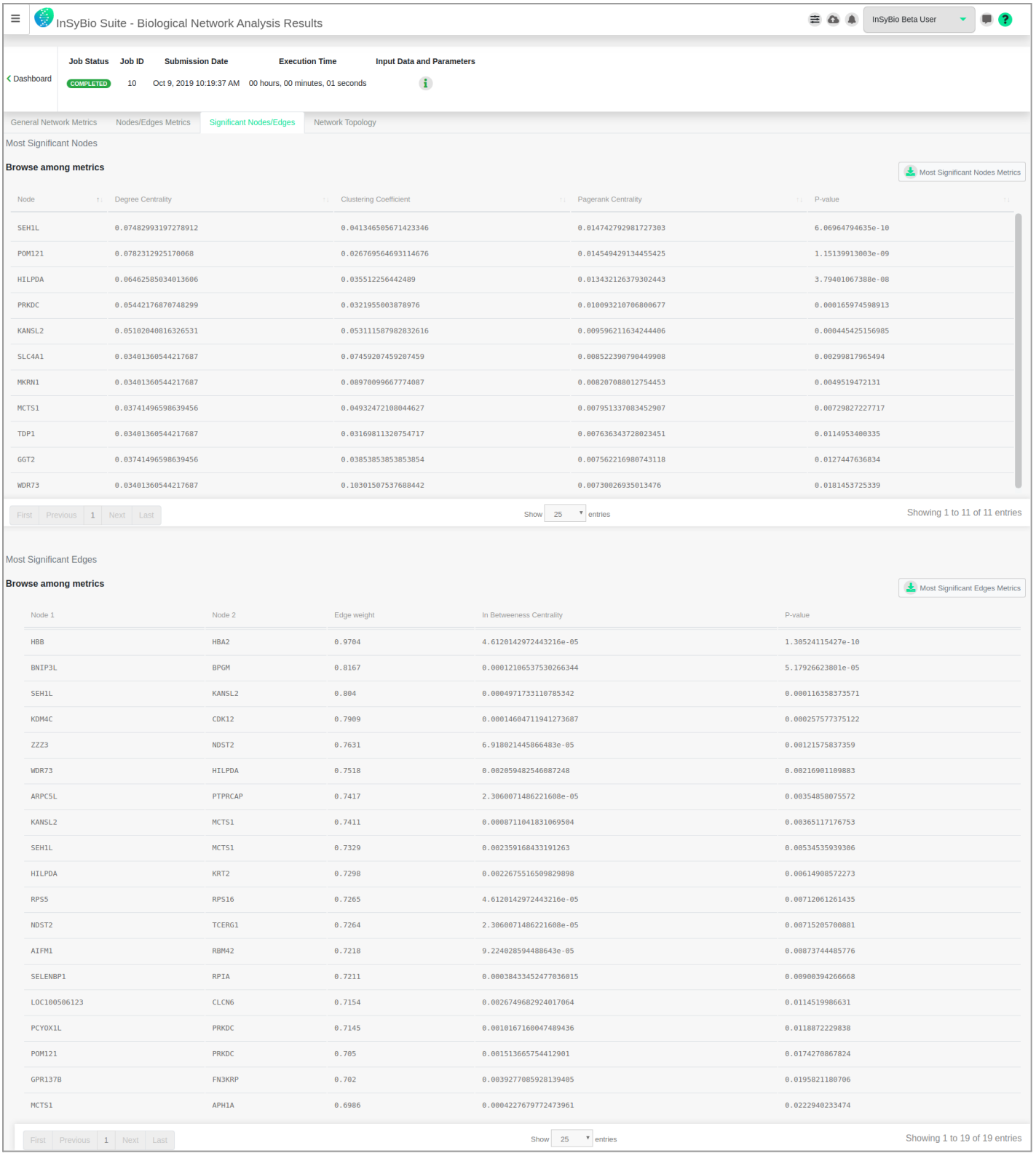
Node/Edges Metrics

In the Node/Edges Metrics tab users can find the metrics for all nodes (degree centrality, clustering coefficient and pagerank centrality) and edges (edge weight and in betweenness centrality) of your network.



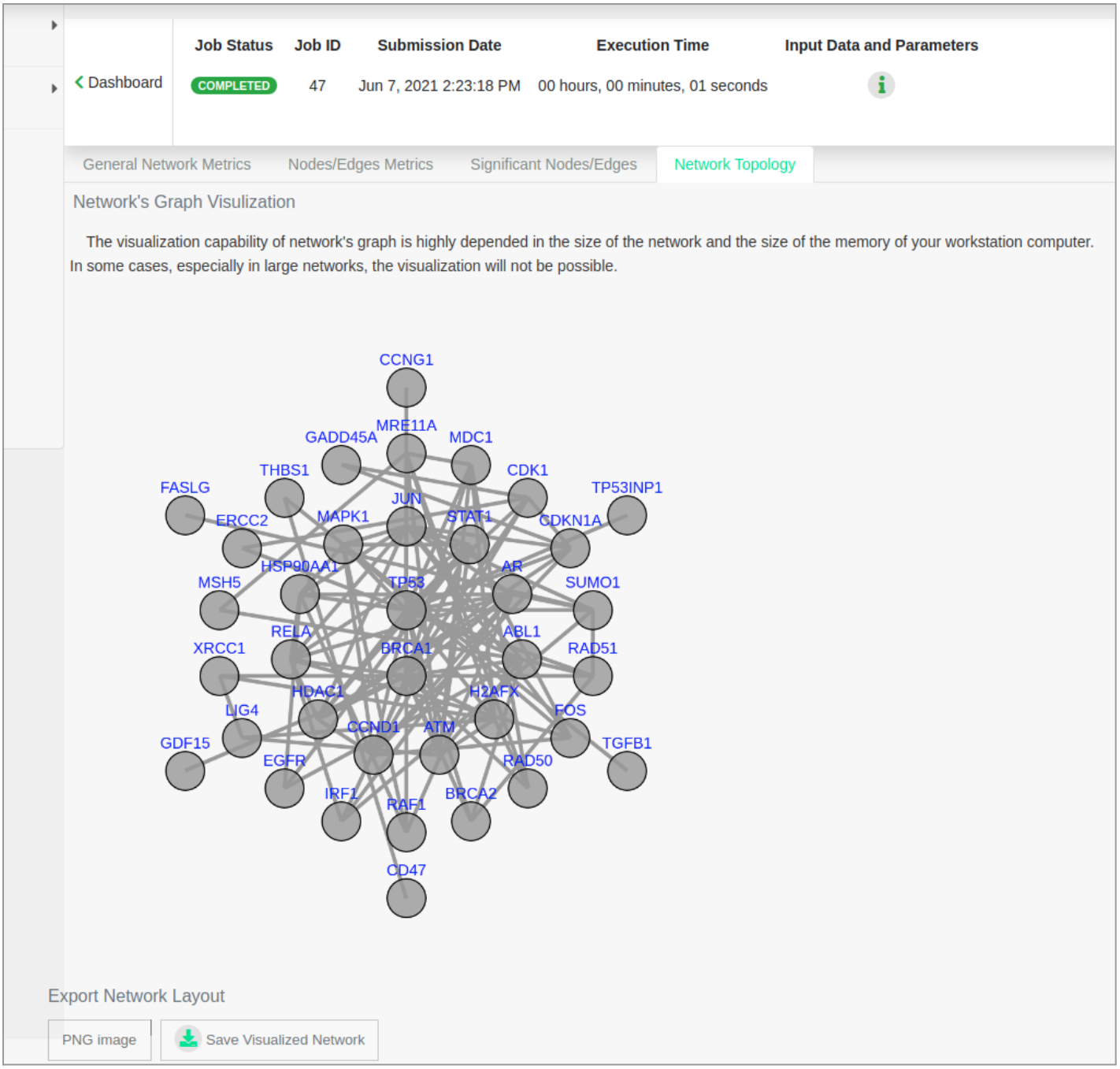
Uncovering Significant Nodes/Edges

In the Uncovering Significant Nodes/Edges tab users can access two tables including the uncovered significant nodes and edges. For each node and edge, the respective metrics and the p-values of their significance a. Significant edges are not available for networks with more than 225000 edges.


InSyBio Suite - Biological Network Analysis Results
InSyBio Beta User
Dashboard
Job Status
Job ID
Submission Date
Execution Time
Input Data and Parameters
COMPLETED
10
Oct 9, 2019 10:19:37 AM
00 hours, 00 minutes, 01 seconds
General Network Metrics
Nodes/Edges Metrics
Significant Nodes/Edges
Network Topology
Most Significant Nodes
Browse among metrics
Most Significant Nodes Metrics
Node
Degree Centrality
Clustering Coefficient
Pagerank Centrality
P-value
SEH1L 0.07482993197278912 0.041346505671423346 0.014742792981727303 6.06964794635e-10
POM121 0.0782312925170068 0.026769564693114676 0.014549429134455425 1.15139913003e-09
HILPDA 0.06462585034013606 0.035512256442489 0.013432126379302443 3.79401067388e-08
PRKDC 0.05442176870748299 0.0321955003878976 0.010093210706800677 0.000165974598913
KANSL2 0.05102040816326531 0.053111587982832616 0.0095962116342444006 0.000445425156985
SLC4A1 0.03401360544217687 0.07459207459207459 0.008522390790449908 0.00299817965494
MKRN1 0.03401360544217687 0.08970099667774087 0.008207088012754453 0.0049519472131
MCTS1 0.03741496598639456 0.04932472108044627 0.007951337083452907 0.00729827227717
TDP1 0.03401360544217687 0.03169811320754717 0.0076363437280234 51 0.0114953400335
GGT2 0.03741496598639456 0.03853853853853854 0.0075622169807431 18 0.0127447636834
WDR73 0.03401360544217687 0.10301507537688442 0.007300269350134 76 0.0181453725339
First
Previous
1
Next
Last
Show 25 entries
Showing 1 to 11 of 11 entries
Most Significant Edges
Browse among metrics
Most Significant Edges Metrics
Node 1
Node 2
Edge weight
In Betweeness Centrality
P-value
HBB HBA2 0.9704 4.6120142972443216e-05 1.30524115427e-10
BNIP3L BPGM 0.8167 0.00012106537530266344 5.17926623801e-05
SEH1L KANSL2 0.804 0.0004971733110785342 0.000116358373571
KDM4C CDK12 0.7909 0.00014604711941273687 0.000257577375122
ZZZ3 NDST2 0.7631 6.918021445866483e-05 0.00121575837359
WDR73 HILPDA 0.7518 0.002059482546087248 0.00216901109883
ARPC5L PTPRCAP 0.7417 2.3060071486221608e-05 0.00354858075572
KANSL2 MCTS1 0.7411 0.0008711041831069504 0.00365117176753
SEH1L MCTS1 0.7329 0.002359168433191263 0.00534535939306
HILPDA KRT2 0.7298 0.0022675516509829898 0.00614908572273
RPS5 RPS16 0.7265 4.6120142972443216e-05 0.00712061261435
NDST2 TCERG1 0.7264 2.3060071486221608e-05 0.00715205700881
AIFM1 RBM42 0.7218 9.224028594488643e-05 0.00873744485776
SELENBP1 RPIA 0.7211 0.0003843343524770360 15 0.00900394266668
LOC100506123 CLCN6 0.7154 0.0026749682924017064 0.0114519986631
PCYOX1L PRKDC 0.7145 0.0010167160047489436 0.0118872229838
POM121 PRKDC 0.705 0.001513665754412901 0.0174270867824
GPR137B FN3KRP 0.702 0.0039277085928139405 0.0195821180706
MCTS1 APH1A 0.6986 0.0004227679772473961 0.0222940233474
First
Previous
1
Next
Last
Show 25 entries
Showing 1 to 19 of 19 entries

Biological Network Visualization

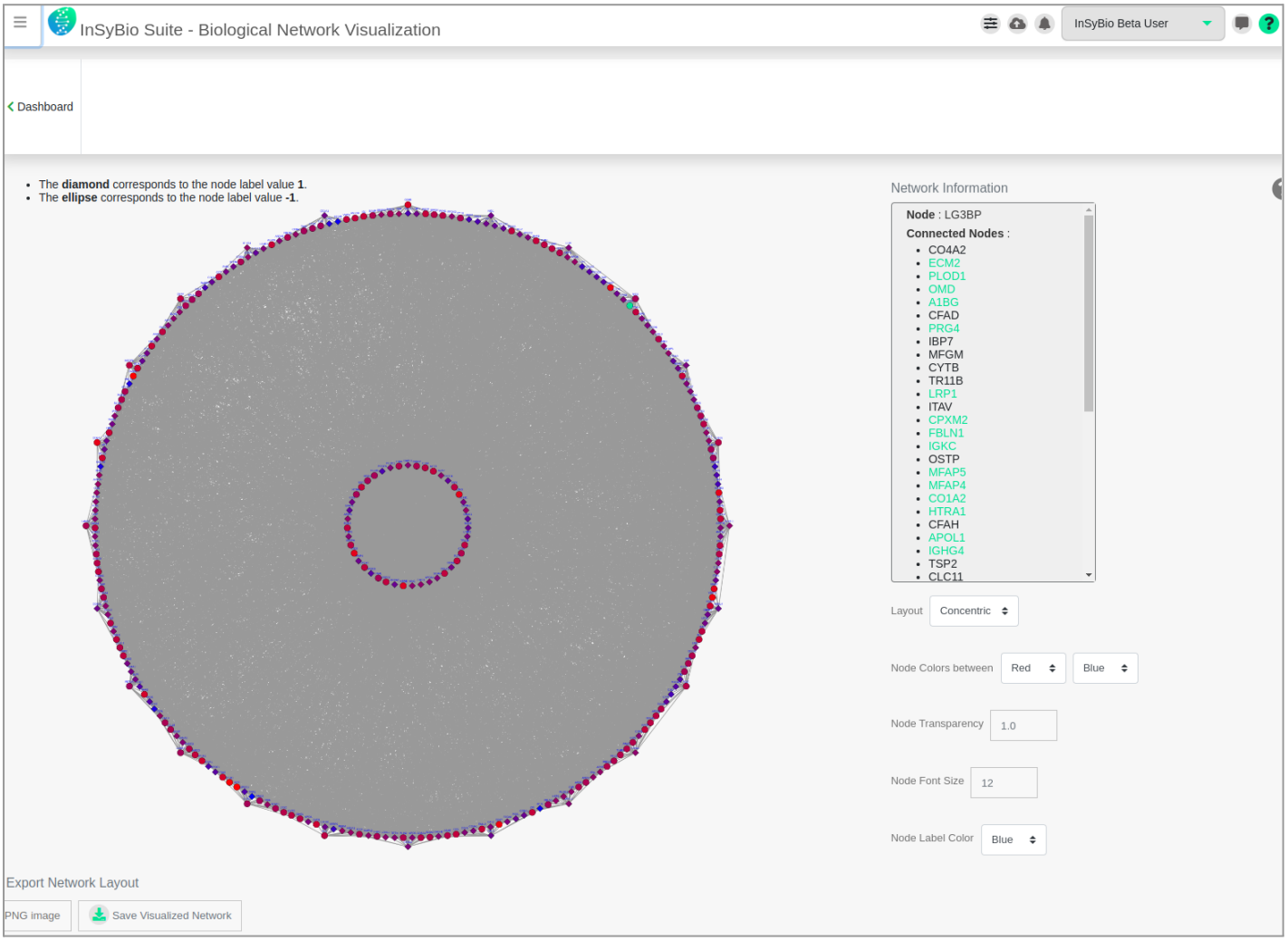
The Biological Network Visualization tab offers an interactive visual representation of the biological network. When networks have more than 10,000 edges, a haircut filter is applied before the visualization of the network. If the haircut filter cannot reduce the number of edges below 10,000 edges then no network visualization is provided. Networks' visualization is based on the Cytoscape plugin [12] and it provides an interactive graphical interface. Users can retrieve information about clicked nodes and edges, export the image in different formats (a PNG, SVG, PDF, XGMML, GraphML or SIF document), decrease opacity on mouseover and view the network using different visualization layouts (force-directed, circle or radial).



Biological Network Visualization

The Biological Network Visualization function offers an interactive visual representation of the biological network. The user has the opportunity to upload a node labels file and a color node scale file with his personal requirements. When networks have more than 10,000 edges, a haircut filter is applied before the visualization of the network. If the haircut filter cannot reduce the number of edges

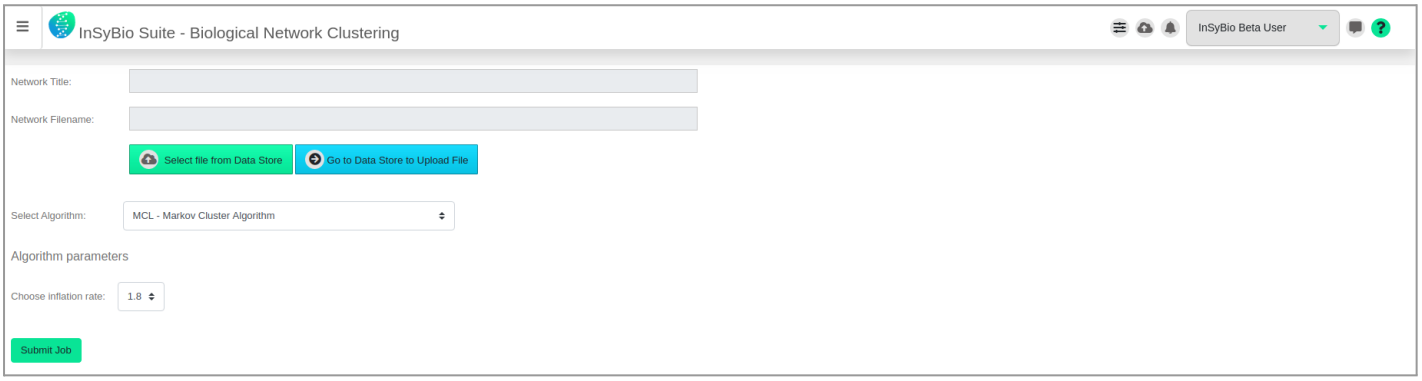
below 10,000 edges then no network visualization is provided. Networks' visualization is based on the Cytoscape plugin [12] and it provides an interactive graphical interface. Users can retrieve information about clicked nodes and edges, export the image in different formats (a PNG, SVG, PDF, XGMML, GraphML or SIF document), decrease opacity on mouseover and view the network using different visualization layouts (concentric, force-directed fcose, breadthfirst, circle, random, grid, cose), The user can also the node colors, the node transparency, the node font size and the node label color.



Biological Network Clustering

With this tool, you can analyze your Biological Network to extract complexes of similar nodes (e.g. protein complexes). Weighted and unweighted Biological Networks can be handled. Three options are supplied for the prediction of Biological Network complexes: Markov Clustering (MCL) [3], Restricted Neighborhood Search Clustering (RNSC) [4] and Clustering with Overlapping Neighborhood Expansion (ClusterONE) [22]. You can name and enter your network via a file or by a saved snapshot in your Data Store and specify the algorithm parameters:

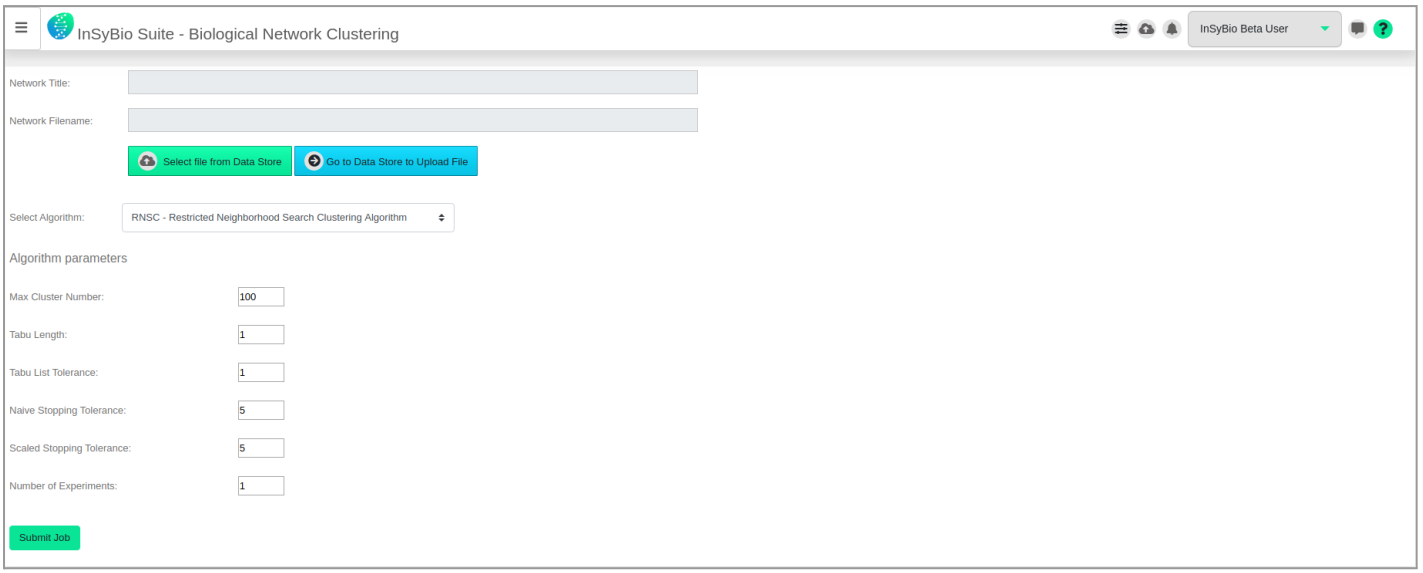
MCL algorithm



- Inflation Rate parameter (default value 1.8)

The output is a list of the clusters created from the algorithm. For each cluster, the node IDs are listed.

RNSC algorithm



- Maximum number of clusters (default value 100),
- Tabu length (default value 1),
- Tabu list tolerance (default value 1),
- Naive stopping tolerance (default value 5),
- Scaled stopping tolerance (default value 5), and
- Number of experiments (default value 1)

Each type of graph has a fairly similar response to the changing tabu length, and RNSC clusters each quite well (and quite quickly) with a tabu length of around n/100, where n the number of nodes in the graph.

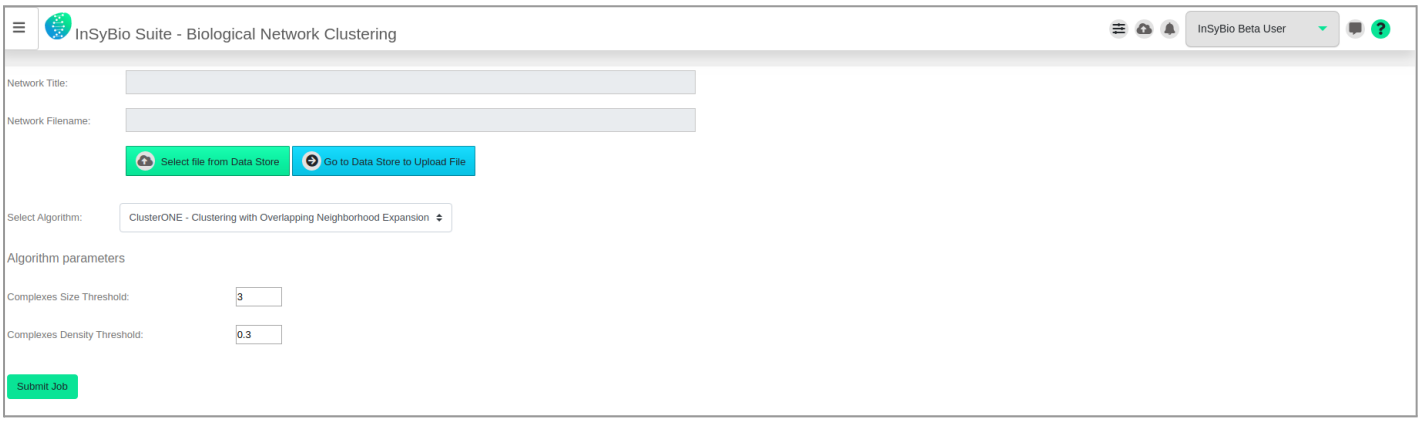
King et al [4] showed that a diversification frequency of n/50 and length diversification = n/5 (n the number of nodes in the graph) yield fairly good results in terms of both score and time, but no set of parameters is clearly the best for all classes of graphs. For more information on Parameter Training and Statistical Results please consult [5].

The RNSC has the capability to efficiently searching many minima in the search space. In this section we will consider a standard tabu list, i.e. one with tabu list tolerance equal to 1, since it was shown in [5] that extending the tabu tolerance did not offer an advantage for the problem.

For the naive and scaled stopping tolerance, there is always a choice between speed and quality. For all graph types, the final scaled cost decreases as the stopping tolerance is increased and the standard deviation of the cost decreases similarly.

The output is a list of the clusters created from the algorithm. For each cluster, the node IDs are listed.

ClusterONE algorithm

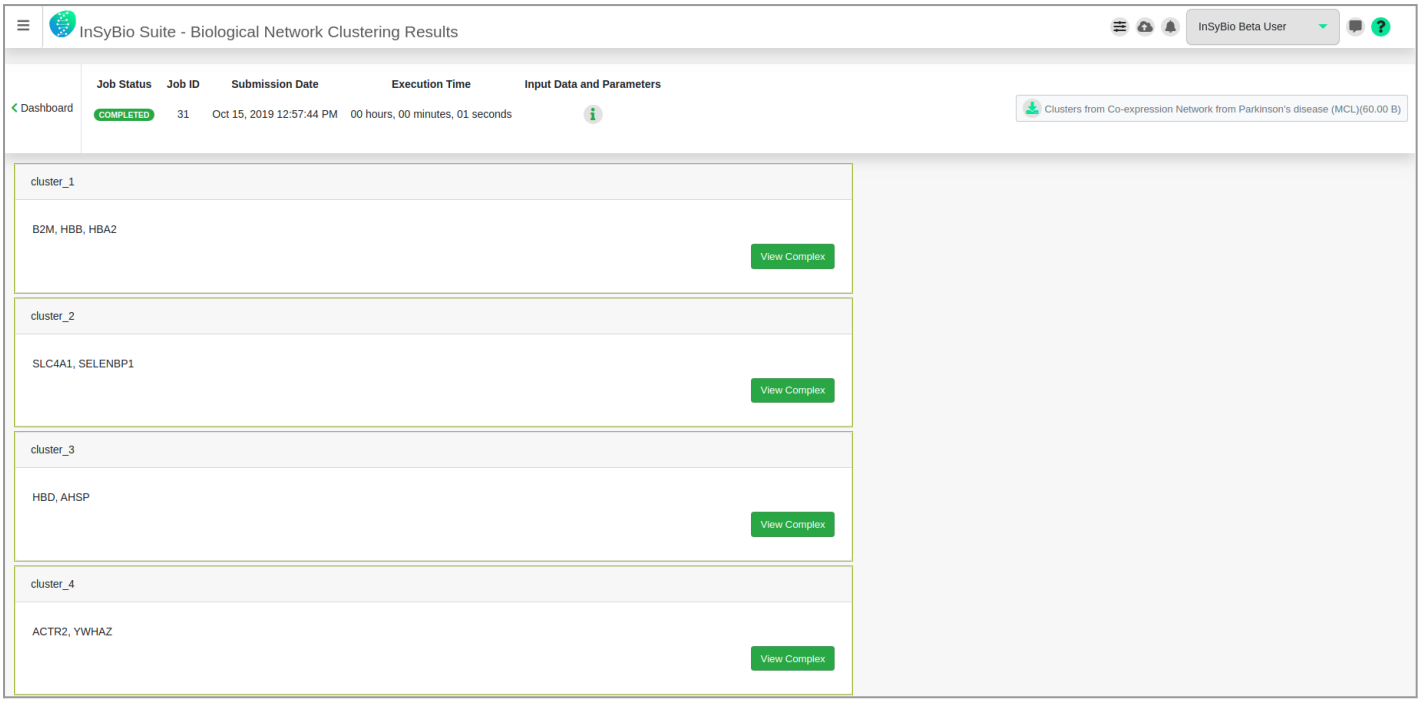


- Complexes size threshold: (default value 3)
- Complexes density threshold: (default value 0.3)

The output is a list of the clusters created from the algorithm. For each cluster, the node IDs are listed.

View Results

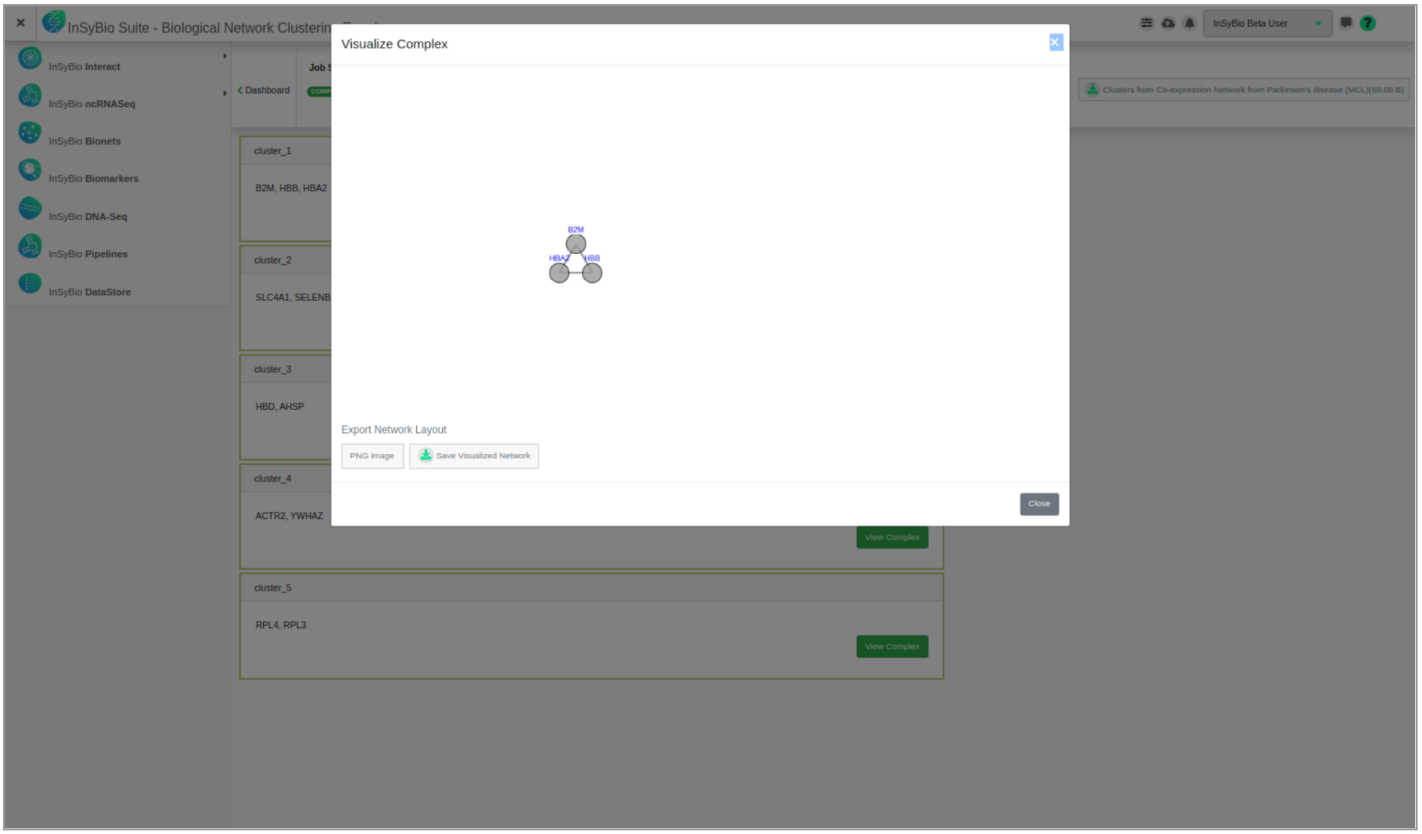
After starting an analysis job you can go to “BioNets Jobs Dashboard”, where you can view the status of your current and previous BioNets jobs. After the completion of the analysis you can select the “View Results” at the Actions column and view the Network Clusters.



The results are presented on your screen as a list or you can download them as a TAB delimited tsv file.

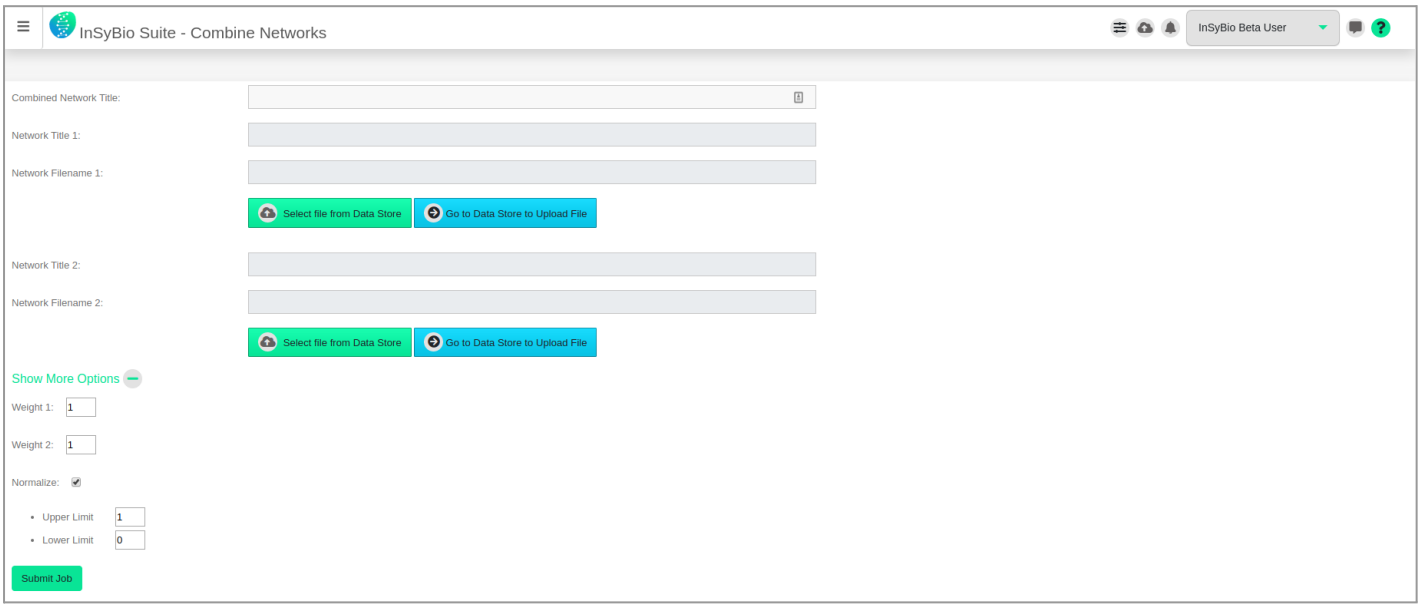
Clusters Visualization

For each computed cluster the participating nodes are presented and it can be visualized. Clusters' visualization is based on the Cytoscape plugin [12] and it provides an interactive graphical interface. Users can retrieve information about clicked nodes and edges, export the image in different formats (a PNG, SVG, PDF, XGMML, GraphML or SIF document), decrease opacity on mouseover and view the network using different visualization layouts (force-directed, circle or radial).


InSyBio Suite - Biological Network Clusterin
InSyBio Beta User
InSyBio Interact
InSyBio ncRNASeq
InSyBio Bionets
InSyBio Biomarkers
InSyBio DNA-Seq
InSyBio Pipelines
InSyBio DataStore
Dashboard
Clusters from Co-expression Network from Parkinson's disease (MCL)(60.00 B)
cluster_1
B2M, HBB, HBA2
cluster_2
SLC4A1, SELENB
cluster_3
HBD, AHSP
cluster_4
ACTR2, YWHAZ
View Complex
cluster_5
RPL4, RPL3
View Complex
Visualize Complex
B2M
HBA2
HBB
Export Network Layout
PNG image
Save Visualized Network
Close

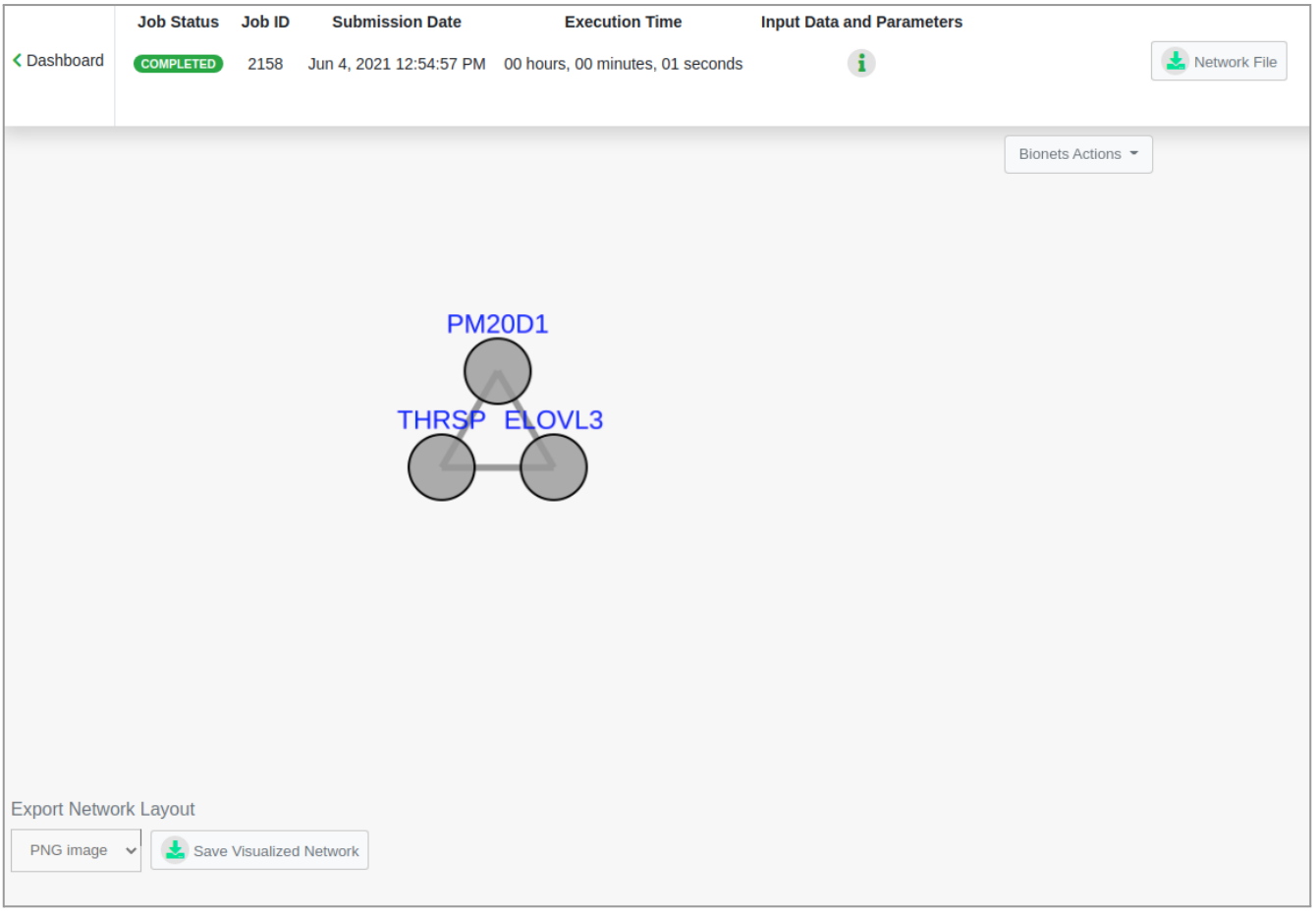
Combine Networks

You can select from Data Store two files, representing Biological Networks, to combine with same or different node ids types. If they have the same node type they are merged as is, if they have different node type they are changed to GeneSymbols ids using the Insybio Interact knowledgebase. You can perform normalization to the edge scores, according to your specifications and also use different weights for the combination of networks.



View Results

In order to access your network you can go to “BioNets Jobs Dashboard”, where you can view the status of your current and previous BioNets jobs. After the completion of the graph you can select the “View Results” at the Actions column and view the combined network results.


Dashboard
Job Status
Job ID
Submission Date
Execution Time
Input Data and Parameters
COMPLETED
2158
Jun 4, 2021 12:54:57 PM
00 hours, 00 minutes, 01 seconds
Network File
Bionets Actions
PM20D1
THRSP
ELOVL3
Export Network Layout
PNG image
Save Visualized Network

References

1. Kohalmi, S. E., Reader, L. J., Samach, A., Nowak, J., Haughn, G. W., & Crosby, W. L. (1998). Identification and characterization of protein interactions using the yeast 2-hybrid system. In Plant Molecular Biology Manual (pp. 95-124). Springer Netherlands.
2. Puig, O., Caspary, F., Rigaut, G., Rutz, B., Bouveret, E., Bragado-Nilsson, E., ... & Séraphin, B. (2001). The tandem affinity purification (TAP) method: a general procedure of protein complex purification. Methods, 24(3), 218-229.
3. Van Dongen, S. M. (2001). Graph clustering by flow simulation.
4. King, A. D., Pržulj, N., & Jurisica, I. (2004). Protein complex prediction via cost-based clustering. Bioinformatics, 20(17), 3013-3020.
5. King, A. (2005). An efficient cost-based graph clustering algorithm. McGill University, Montreal.
6. Page, L., Brin, S., Motwani, R., & Winograd, T. (1999). The PageRank citation ranking: bringing order to the Web.
7. Tong, H., Faloutsos, C., & Pan, J. Y. (2006). Fast random walk with restart and its applications.
8. Saramäki, J., Kivelä, M., Onnela, J. P., Kaski, K., & Kertesz, J. (2007). Generalizations of the clustering coefficient to weighted complex networks. Physical Review E, 75(2), 027105.
9. Opsahl, T., Agneessens, F., & Skvoretz, J. (2010). Node centrality in weighted networks: Generalizing degree and shortest paths. Social Networks, 32(3), 245-251.
10. Lu, L., & Zhang, M. (2013). Edge Betweenness Centrality. In Encyclopedia of Systems Biology (pp. 647-648). Springer New York.
11. de la Peña, J. A., Gutman, I., & Rada, J. (2007). Estimating the Estrada index. Linear Algebra and its Applications, 427(1), 70-76.
12. Lopes, C. T., Franz, M., Kazi, F., Donaldson, S. L., Morris, Q., & Bader, G. D. (2010). Cytoscape Web: an interactive web-based network browser.Bioinformatics, 26(18), 2347-2348.

13. Edgar, R., Domrachev, M., & Lash, A. E. (2002). Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. Nucleic acids research, 30(1), 207-210.
14. Zhang, B., & Horvath, S. (2005). A general framework for weighted gene co-expression network analysis. Statistical applications in genetics and molecular biology, 4(1).
15. Song, L., Langfelder, P., & Horvath, S. (2012). Comparison of co-expression measures: mutual information, correlation, and model based indices. BMC bioinformatics, 13(1), 328.
16. Hatfield, G., Hung, S. P., & Baldi, P. (2003). Differential analysis of DNA microarray gene expression data. Molecular microbiology, 47(4), 871-877.
17. Breitling, R., & Herzyk, P. (2005). Rank-based methods as a non-parametric alternative of the T-statistic for the analysis of biological microarray data. Journal of bioinformatics and computational biology, 3(05), 1171-1189.
18. Cui, X., & Churchill, G. A. (2003). Statistical tests for differential expression in cDNA microarray experiments. Genome Biol, 4(4), 210.
19. Safran, M., Dalah, I., Alexander, J., Rosen, N., Stein, T. I., Shmoish, M., ... & Lancet, D. (2010). GeneCards Version 3: the human gene integrator. Database,2010, baq020.
20. Amberger, J., Bocchini, C., & Hamosh, A. (2011). A new face and new challenges for Online Mendelian Inheritance in Man (OMIM®). Human mutation,32(5), 564-567.
21. Piñero, J., Queralt-Rosinach, N., Bravo, À., Deu-Pons, J., Bauer-Mehren, A., Baron, M., ... & Furlong, L. I. (2015). DisGeNET: a discovery platform for the dynamical exploration of human diseases and their genes. Database, 2015, bav028.
22. Nepusz, T., Yu, H., & Paccanaro, A., (2012). Detecting overlapping protein complexes in protein-protein interaction networks. Nature Methods volume 9, pages 471–472 (2012)
23. Xu, W., Hou, Y., Hung, Y. S., & Zou, Y. (2013). A comparative analysis of Spearman's rho and Kendall's tau in normal and contaminated normal models. Signal Processing, 93(1), 261-276.

How to get InSyBio BioNets

To request a free one month license of InSyBio Suite please email us at info@insybio.com .

To purchase InSyBio Interact commercial version 3.1 please contact us at sales@insybio.com.

About Us

InSyBio Ltd is a bioinformatics pioneer company (www.insybio.com) in personalized healthcare, that focuses on developing computational frameworks and tools for the analysis of complex life-science and biological data in order to develop predictive integrated biomarkers (biomarkers of various categories) with increased prognostic and diagnostic aspects for the personalized Healthcare Industry.

InSyBio Suite consists of tools for providing integrated biological information from various sources, while at the same time it is empowered with robust, user-friendly and installation-free bioinformatics tools based on intelligent algorithms and methods.

COPYRIGHT NOTICE