InSyBio Pipelines

March 2022

Insybio Suite v3.0

InSyBio

Intelligent Systems Biology

User Manual

www.insybio.com

Insybio Pipelines

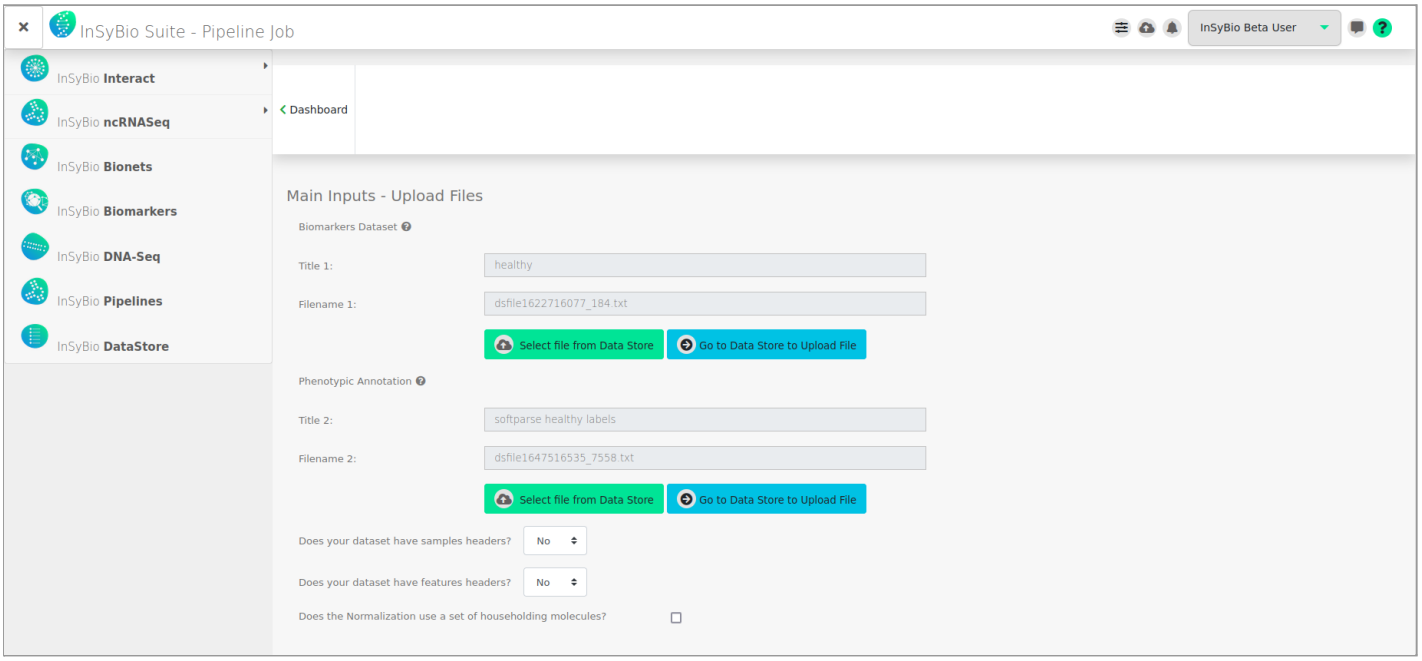
Introduction

Pipelines is a tool that provides an integrated optimized pipeline from raw data until the discovery of biosignatures, networks and predictive models. It takes as input a dataset file and after performing statistical and network analytics, it uses the dataset to train a machine learning model identifying at the same time the optimal biomarkers for the trained model.

To start the calculation:

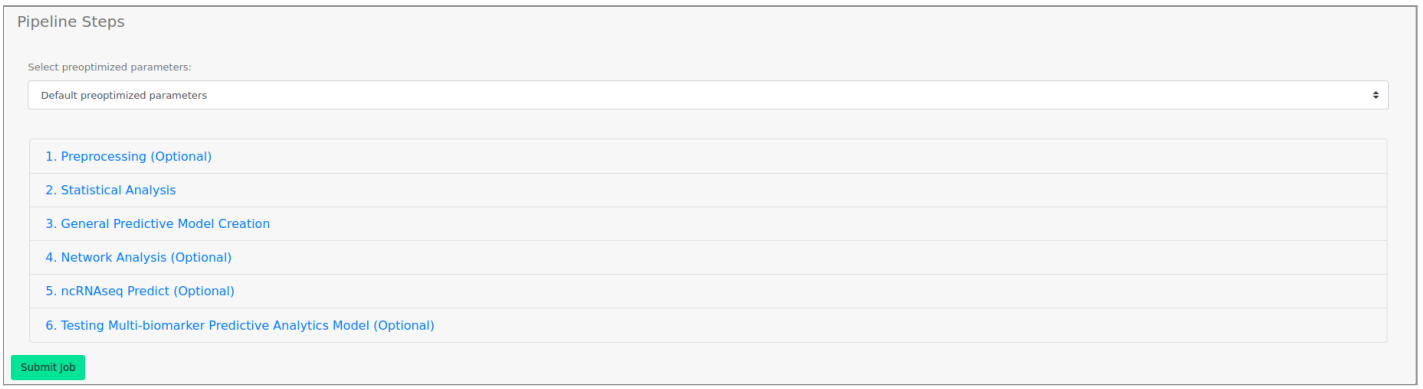
Click in the menu “InSyBio Pipelines” , select the “Add new job” button and then:

- Upload a new Biomarkers Dataset file with features as rows and samples as columns and a corresponding Biomarkers Labels file with all labels in one row, separated by tabs. You are redirected to the Data Store where step by step instructions guide you.
- Or Select a file from the Data Store. There you can find your previously uploaded files or InSyBio pre-uploaded sample datasets.



After that the user will have to insert the information regarding the headers. That is to inform the application if the original dataset has sample headers or feature headers. Optionally, the user will have the option to insert the names of the features that will be used for normalization.

The user should also select which Pipeline Steps he wants to be performed. There are six steps. All steps have pre-optimized parameters and steps for general purpose and pre-optimized parameters and steps for specific application domains that we use as defaults, but the user can configure these values manually. It is advised that the default values are used.



Preprocessing

The first step is the Preprocessing. The user can select if preprocessing should be performed. During preprocessing we filter the dataset, perform normalization, missing values imputation, duplicate measurements averaging and outlier detection with the PCA LOF method.

More specifically, there are two kinds of normalization: arithmetic sample-wise and logarithmic. It should be noted that when the data contain negative numbers the arithmetic normalization should be chosen, since logarithmic normalization method functions only with non-negative data. If “None” is chosen, no normalization takes place.

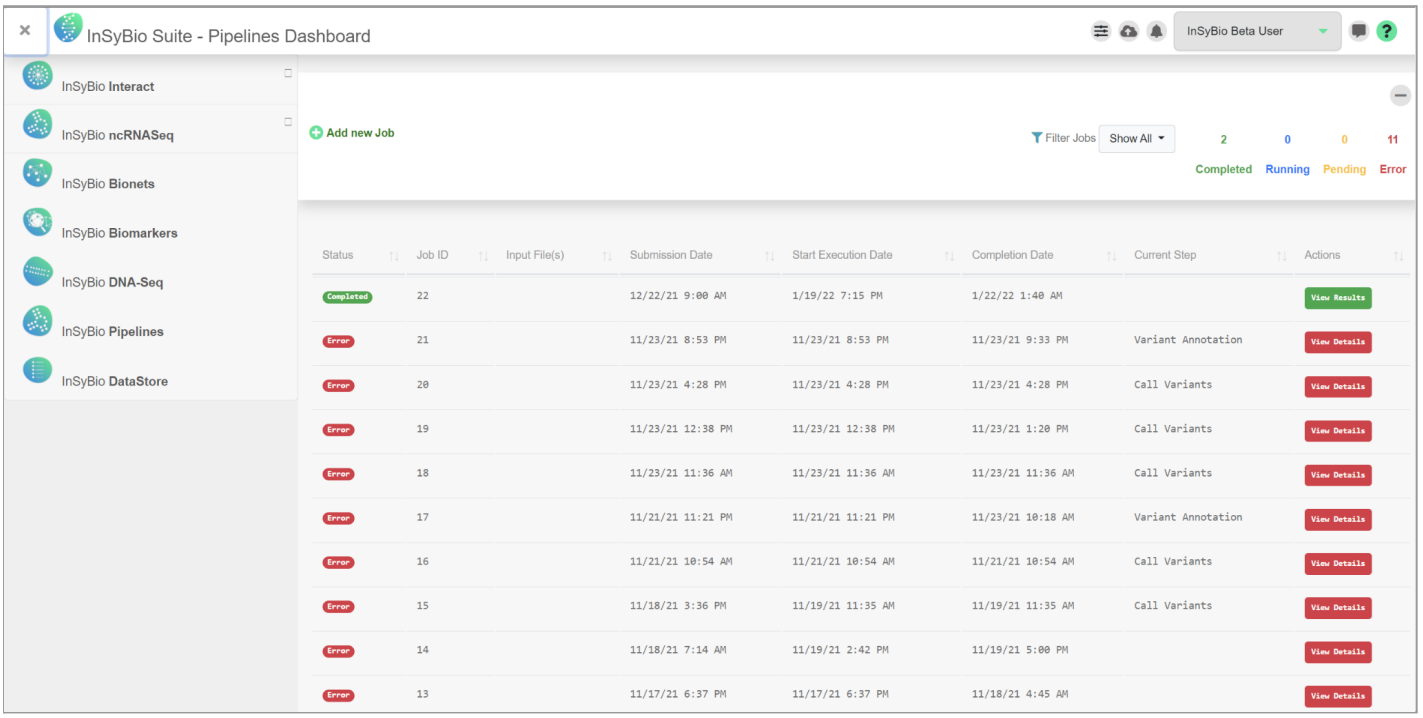
There are two kinds of missing value imputation methods as well: average imputation and KNN imputation. Average imputation is a method in which the missing value on a certain variable is replaced by the mean of the available cases[1]. On the other hand, the key idea of KNN imputation is that a point value can be approximated by the values of the points that are closest to it, based on other variables[2]. The above missing values imputation methods are relevant only for cases where a missing value does not imply a quantification value of zero. In such cases, missing values should be replaced with zeros before uploading the dataset. If “None” is chosen then no missing value imputation takes place.

[1] Single Imputation Methods (iriseekhout.com)
[2] The use of KNN for missing values (towardsdatascience.com)

Furthermore, if a missing values imputation method is being chosen instead of “None”, the duplicate measurements will be averaged.

To view the results:

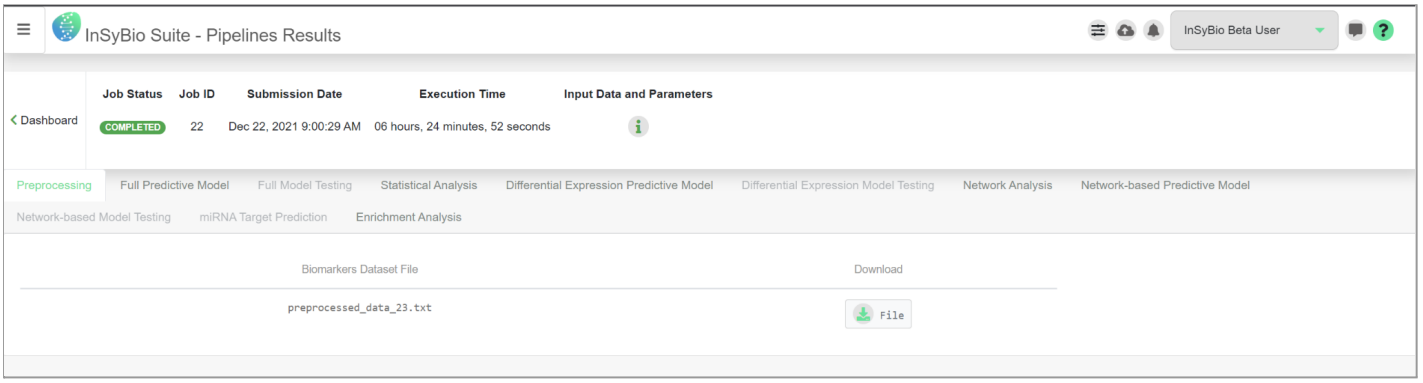
By starting a calculation the Pipelines dashboard is updated with the submitted job. There you can view the status of your current and previous Pipelines calculations. At completion of the calculation you can select the View Details at the Actions column and view the results.



In the Preprocessing tab the user will be able to download the resulting preprocessed file.

Statistical Analysis

The second step is Statistical Analysis. This step is mandatory. At this step, Differential Expression Analysis, Heatmap construction and Spearman correlation table construction are performed.



Only variables annotated as genes/transcripts/proteins will be used for differential expression analysis. If a user has uploaded a phenotypic annotation file with more than two columns then multiple tasks will be created with one column of the phenotypic annotation per file. Every phenotypic column can take two or more values.

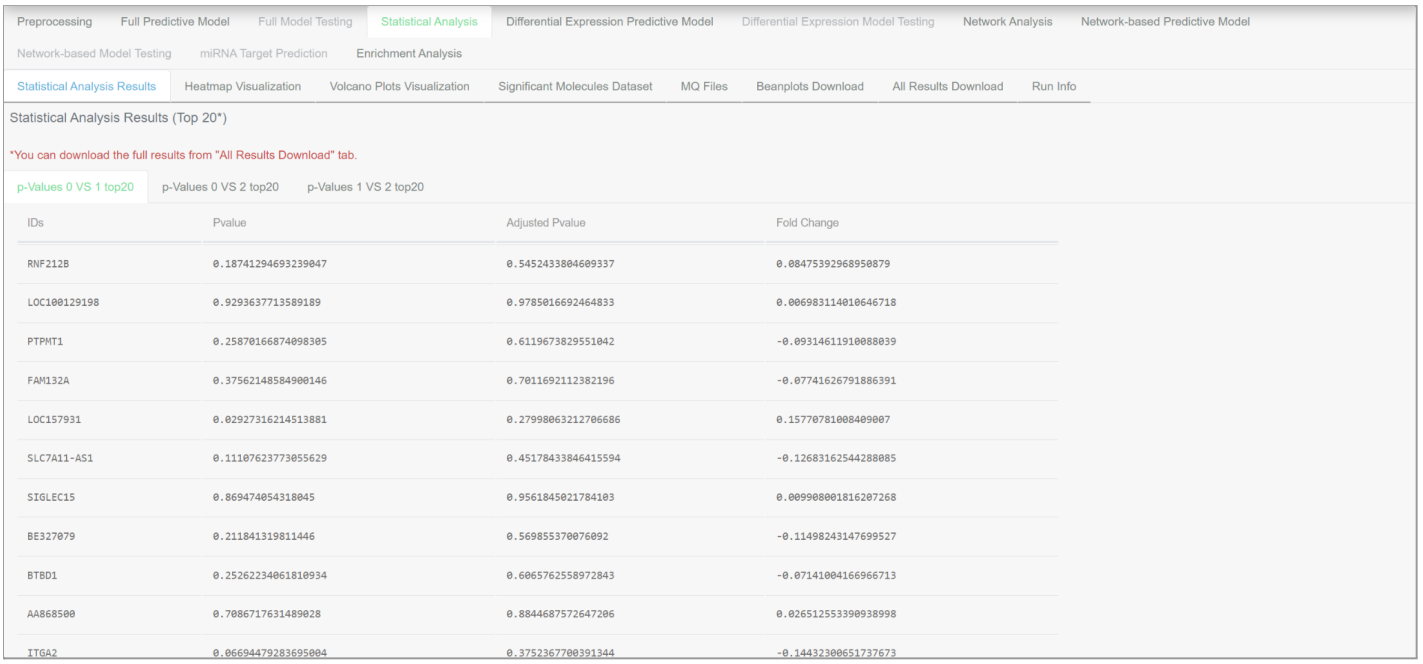
The user has to select the type of analysis to be made on the inserted dataset. There are two types of statistical analysis, paired and unpaired analysis.

Afterwards the user will have to insert the p-value threshold value, which is recommended to be 0.05.

Then, the user will choose the kind of test to be performed on the selected dataset: automatic, parametric Ebayes Test Selection, parametric 2-sided Students T-test (or

One-way ANOVA Test Selection) or non-parametric Kruskal Wallis (or Mann Whitney Test Selection) test. If the automatic version is chosen, then our algorithm will decide which test to run: the parametric or non-parametric.

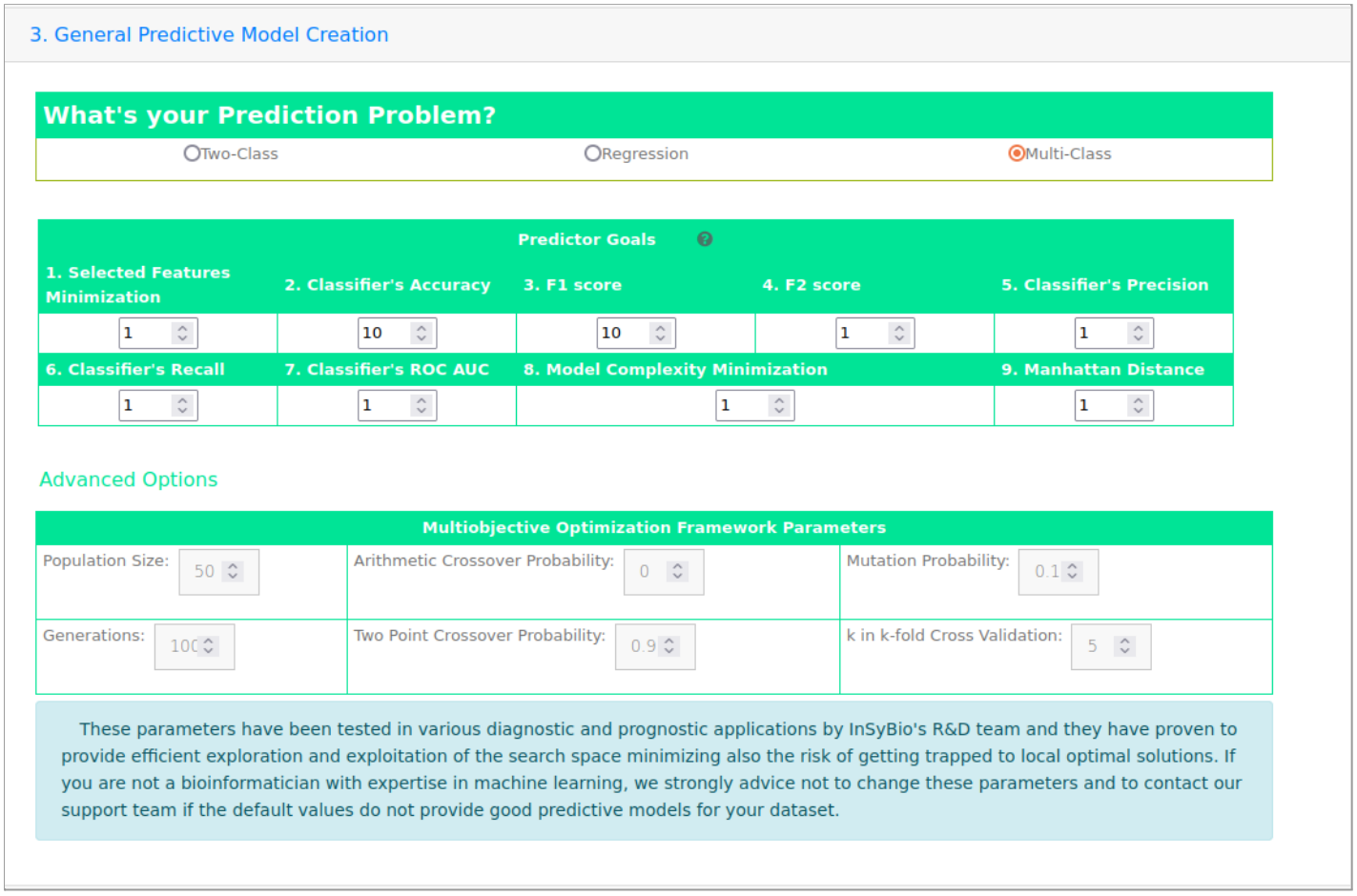
To view the results:



In the Statistical Analysis tab the user will be able to view the Statistical Analysis Results (the top 20 features), the Heatmaps, the Volcano plots, the significant molecules, the molecular quantification (MQ) files and he'll be able to download the Beanplots and all the resulting files. Finally, in the last tab the run information will be displayed.

General Predictive Model Creation

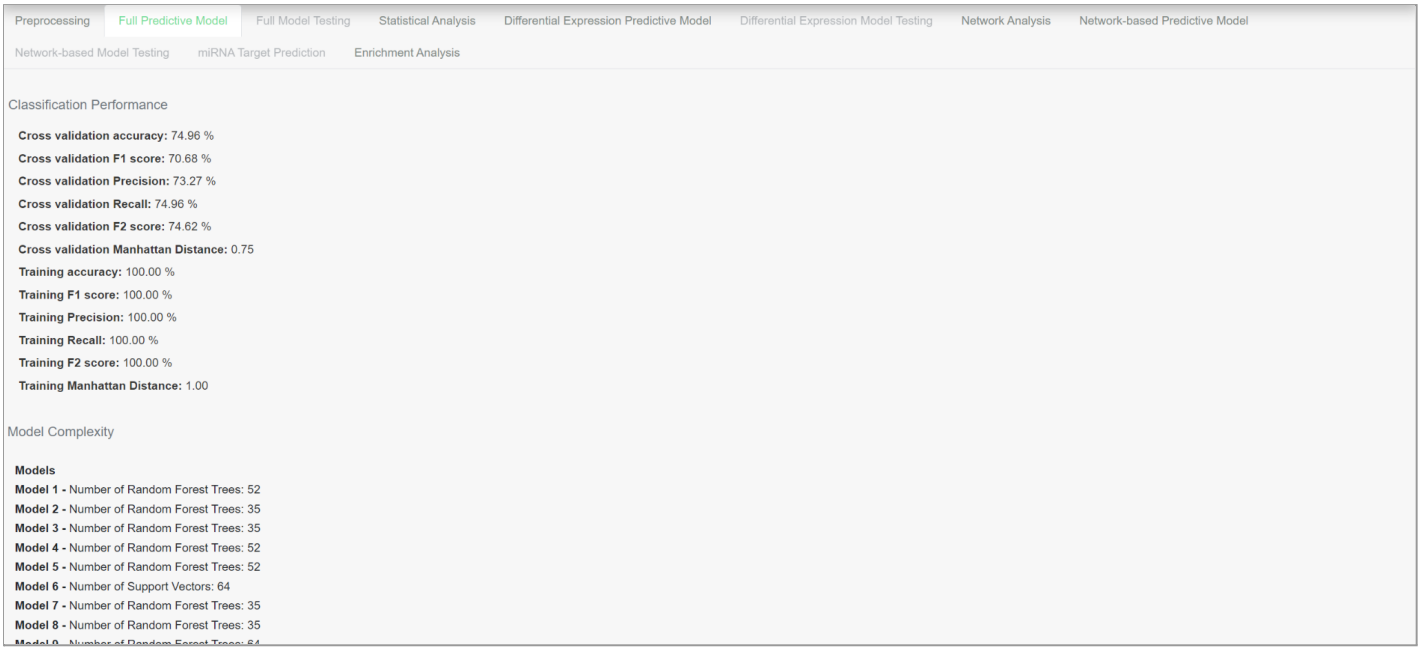
The third step is the General Predictive Model Creation. This step is also mandatory and allows users to train their own predictors using the biomarkers dataset, the phenotypic annotation and the parameters that they selected.



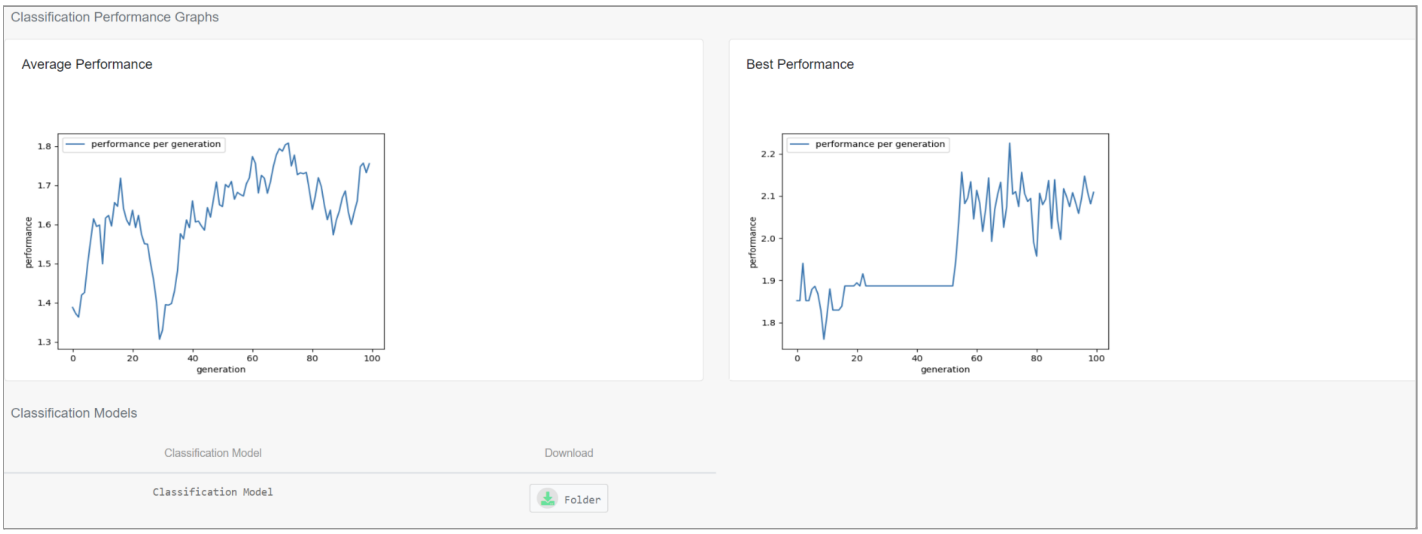
The user chooses the kind of prediction problem he has at hand. Later, he can alter the weights of the predictor goals. It is advisable to use the default values. The higher the weight, the more significant the goal.

Finally, the user can alter the multi-objective optimization framework parameters at the Advanced Options. Those are the population size, the number of generations, the arithmetic crossover probability, the two point crossover probability, the mutation probability and the number of folds k for the cross validation.

To view the results:



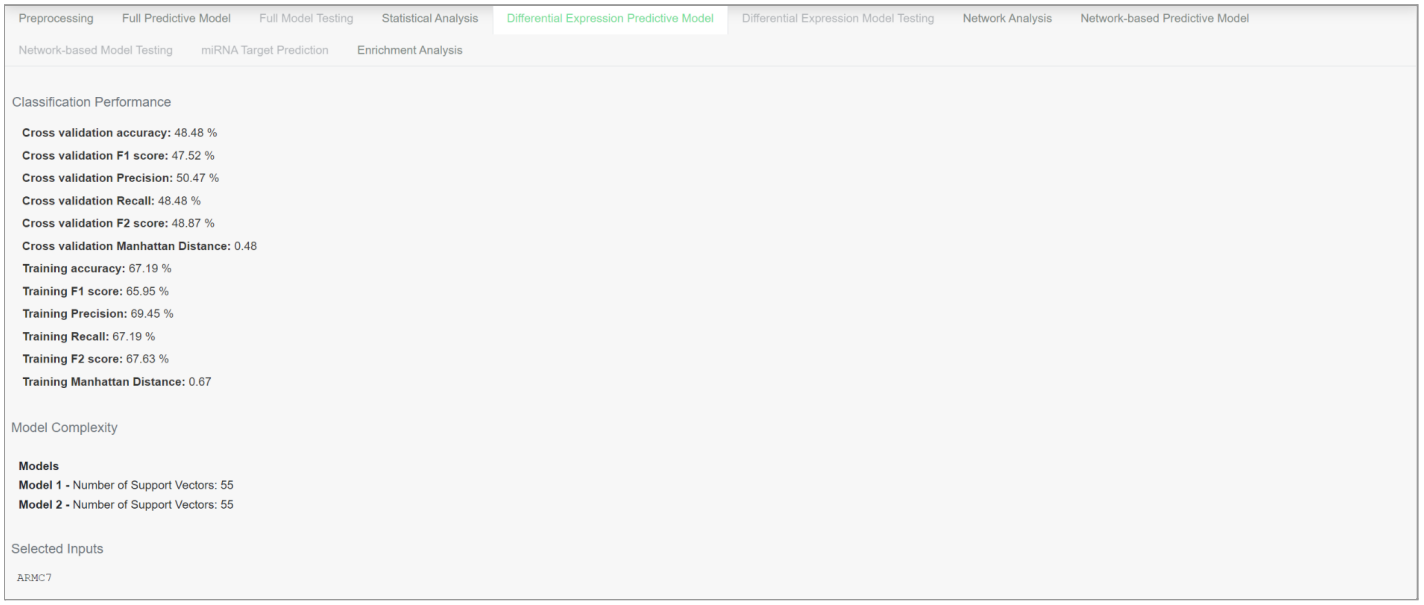
,ISOC1,YPEL3,RPS24,RECK,RBM15,HID1,ZNF816-
ZNF321P,AF187554,BET1L,PNMA3,ZP2,AK026701,TUBB4B,ARHGAP6,AI733177,RASA3,PRDM12,AF090925,NMRAL1P1,TIMM13,UNC5A,AQP12B,C7orf43,AV736725,AIP,AI7987
24,UBAP2L,PAPLN,BE044484,AW172407,LINC00588,LOC100133920,SNF8,PQLC2L,AK022067,CDC26,1556333_at,BC016176,ASMTL,237424_at,AI469935,LINC01535,AF086
565,AI733457,SFMBT1,TMEM26-
AS1,TFAP4,240896_at,HECW2,AI017540,LOC102724312,RBBP5,C5orf66,CENPH,RAD54B,IFITM3,GUCY2D,BC041050,SEPN1,RPL38,HRG,BC035326,244489_at,PLEKHA6,RAB
3GAP1,AL831920,RHOT2,IQCF5,ZFP62,IFNA5,PET100,SEH1L,OPN5,ATG16L1,AI457449,AL043143,LOC101928510,AW779654,MRPL21,FOLH1,MZT2A,KLHDC9,AI075053,VSTM
4,AI022850,TAL1,228586_at,SLC12A6,AI084064,NUDT19,PABPN1,PMF1-
BGLAP,AW452392,AK022254,BE675275,MYO18A,KLF5,PDCD6IP,RHBG,EFNB3,228896_at,BE858194,CASD1,IL19,N40199,CLSTN2-
AS1,KDF1,PHLDA2,SNX5,LOC100287896,R20660,PRKY,MEIOC,MEX3C,UBL4A,KIF13B,ATG4D,PCA3,MTA1,AI074594,JUNB,NKX2-
5,MRPS17,GNL3,C4orf45,MPL,AP2A1,LOC339685,ACTL9,TSSK3,JPH4,AF339810,BC039410,AW974816,H95280,KLHDC7B,208144_s_at,1554281_at,CCDC151,EP400,ADORA2
B,AI076370,PRKD2,LFNG,LOC100506885,ARL4D,ERGIC2,X84340,PDCD2L,GFRA2,KPNA6,PTPN5,STRA6,PIK3R6,UBXN11,GGH,KIAA0907,RRM2,AW082221,AK021495,ANXA2P3,
AK091415,207916_at,PMS2CL,ATAT1,EPHA5-
AS1,NSUN4,C22orf29,240654_at,TNNC2,CLEC4F,N23258,AL832887,C6orf118,AJ012498,212883_at,S100A12,MICALL1,ZNF600,C9orf131,GJA1,AV700385,TRAF2,TMEM10
8,TMEM51,LOC100129380,PPP1R36,LTC4S,APC,SYS1,IL36A,FAM210A,AK023800,RNF152,AL050136,BC031975,LOC728805,FAF1,HUWE1,AI821649,1566268_at,CUL9,CDC42
SE1,C10orf11,NPM1,IFNA14,RASSF4,LOC101060835,AU156181,ZNF831,AW237316,MARS2,FOXO3B,CEP76,TNRC6C-
AS1,EXOSC4,TMSB10,SGCD,TRIM74,LOC101927040,C3orf35,LOC155060,228549_at,LOC400620,AF147426,LOC101927929,RSL1D1,243160_at,ZNF362,CC2D2A,AF130093,D
BT,AI419968,AI923201,DHTKD1,ZDHHC4,SEPT9,BC041976,PLCL2,BC040306,RANBP1,MIEF1,AI674915,AW139719,244047_at,NDUFB8,SZT2,PAXIP1,NSMCE1,TSC22D1-
AS1,SUPT6H,IPO4,LOC101927379,SLC45A2,HAT1,RNPEPL1,LOC81787,CDK5RAP3,ANGPT4,OXSM,MARK1,TMEM184C,SNAPC5,228156_at,BF508839,GATAD1,BC031957,SUFU,FL
J16779,MFSD4B,1556828_at,DOK7,EGR3,GUSB,240451_at,LOC285847,AI700768,THOC1,BC013633,POP7,CROCC,CH17-
360D5.1,NPFF,PLGRKT,COX8C,SHQ1,208421_at,TFAP2A,MIR4784,LRFN4,BC042181,C5orf51,234454_at,AK024973,AI733345,BE463783,1570653_at,LZIC,AI360167,WDF
Y2,TMCC3,KRT76,AW015426,LOC285902,WDHD1,MIR3917,AA010315,SMG9,LOC101928886,LOC101928787,CD8A,SPRY3,241217_x_at,LILRA2,CDT1,PLAGL2,ZNF446,C17orf8
9,BG403405,AK098724,AF085897,FAM96B,AU150817,TBC1D5,MYH14,C17orf98,BAG6,RRAS,CDCA2,ZNF24,UNG,PHF7,ZCCHC4,GJA5,SNORD18C,BLOC1S2,216943_at,BF51121
2,AL359594,PPP1R9B,AW271060,TMEM171,TMEM8B,SRF,AW572853,AI247365,216568_x_at,BRD3,CCDC93,CROT,PMFBP1,AI922972,LOC101929741,CNTNAP1,LOC728084,CDC
A5,ARAP1,BET1,SRP54,U90905,HAUS1,BC037412,MIR4640,WTAP,XRCC2,BE504838,PRORSD1P,SEC16A,VTI1A,KDM1A,FUT1,ULK3,AW205775,COLGALT2,ZNF467,GLI3,C1R,HS
D17B12,PITPNC1,MYZAP,LOC100507194,MAGEC1,AI421677,LDB1,IL10RB,SLC24A1,LOC100289045,FLAD1,1560905_at,PNMT,INPP5B,AI821694,WDR25,AI923713,SNORD3D,
MEA1,MPPED1,PLXNA3,LINC01468,242192_at,CDR1,HSD17B1,ACTL7A,TUBB2B,AF075064,LOC100310756,LOC101927770,FRK,GGCT,PLRG1,FAM110B,LINC01004,TRBC1,PTTG
2,SECISBP2,CFAP73,VPS4B,GAS1,GRIA4,217676_at,234372_at,AI377746,LOC102724782,AI760332,RBM45,SUN5,AK026468,AP1S2,AU148090,FAM86B3P,DMRTB1,PCDHGC4
,DCAKD,C1orf27,MS4A2,HPCAL4,ESRP1,ATG4A,C10orf10,PAK6,TUBB1,ZSCAN25,MTM1,FAM155A,242223_at,RNF185-
AS1,B3GAT3,RIDA,NOB1,CAAP1,CYP19A1,SMG5,TWF2,TULP3,PSENEN,TMX2-
CTNND1,AA648962,RALA,MURC,TAF12,LINC01622,238586_at,LINC00427,C2orf76,UBE2C,AI138418,DDX49,AF007143,STEAP3,BE066500,LOC105370629,SCO2,AK000106,M
SH2,S81578,216421_at,GRB10,AI822140,BC042815,W96141,ADD3-
AS1,GTF2IP7,ST18,SPTBN2,MRO,BF847120,H27618,LINC00282,AL162010,AA814006,ARHGDIB,AKT2,220862_s_at,TMEM258,AI280131,LOC101928521,NDUFA7,ZNF646,LNX
2,H07100,BCL7C,NSUN2,AV659223,MIR205



In the Full Predictive Model tab the user will be able to view the classification (or regression) performance of the cross validation and of the training set. For the two-class prediction problem the user will be able to see the accuracy, sensitivity, specificity, f1 score, f2 score, ROC AUC score. For the multi-class prediction problem the user will be able to see the accuracy, f1 score, f2 score, precision, recall and Manhattan distance and for the regression prediction problem the user will be able to view the root mean square error, the relative absolute error, the root relative squared error, the R2 (coefficient of determination) regression score, the explained variance score and the Spearman Correlation.

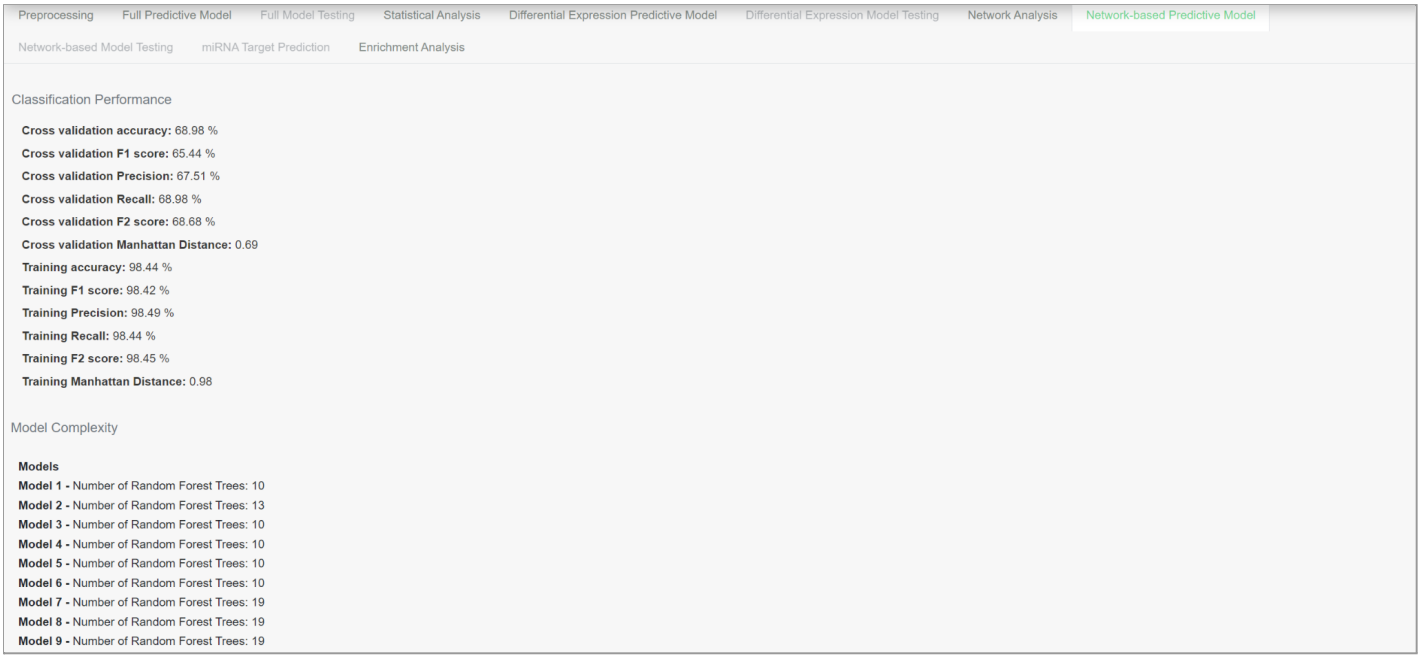
Additionally, the complexity of every model, which is the total number of support vector machines and the number of trees for RandomForest and also, the average and the best performance of the trained model are being displayed.

Differential Expression Predictive Model Results:



At the Differential Expression Predictive Model tab you can view the same results as at the Full Predictive Model tab. The difference between them is the input to this same step. For the Differential Expression prediction step the input is original dataset with only the significant biomarkers selected.

Network-based Predictive Model Results:



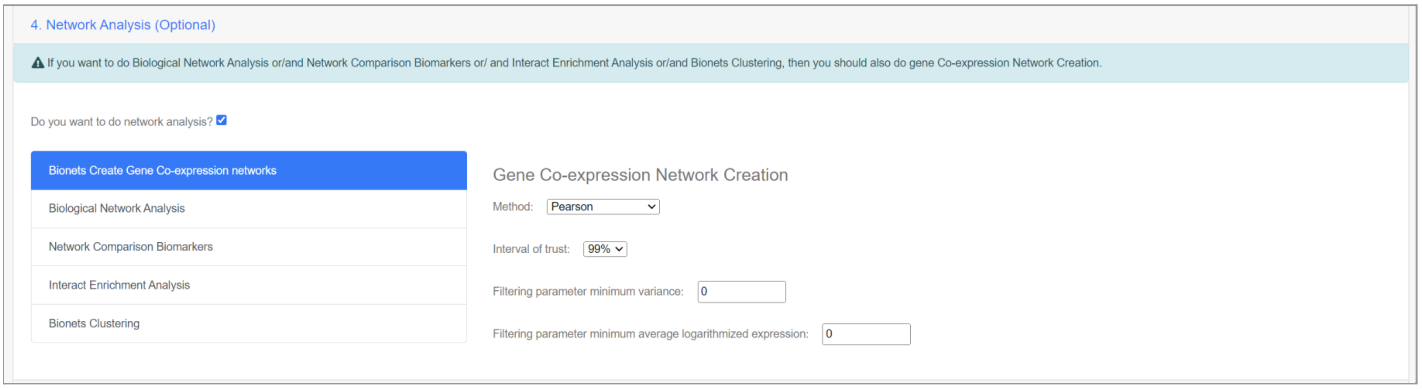
At the Network-Based Predictive Model tab you can view the same results as at the Full Predictive Model tab. The difference between them is the input to this same step. For the Network-Based prediction step the inputs are the original dataset with only the significant biomarkers and the output of the Network Comparison

Network Analysis

The fourth step is Network Analysis. This step is optional and it consists of five steps:

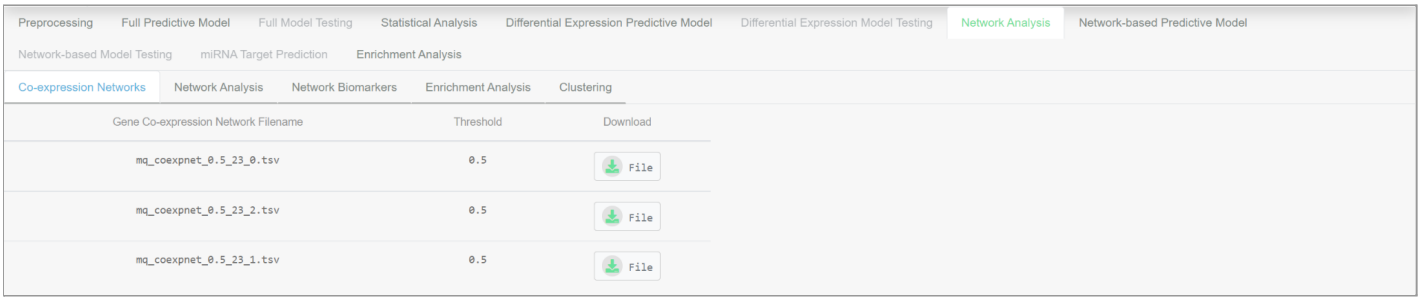
- Bionet Create Gene Co-expression Networks.
- Biological Network Analysis
- Network Comparison Biomarkers
- Interact Enrichment Analysis
- Bionets Clustering

At each step, the default value for every parameter is selected. These values can be configured manually by the user.



To view the results:

In the Network Analysis tab you can view the five different tabs for each step of the



Network Analysis steps.

Gene co-expression network creation

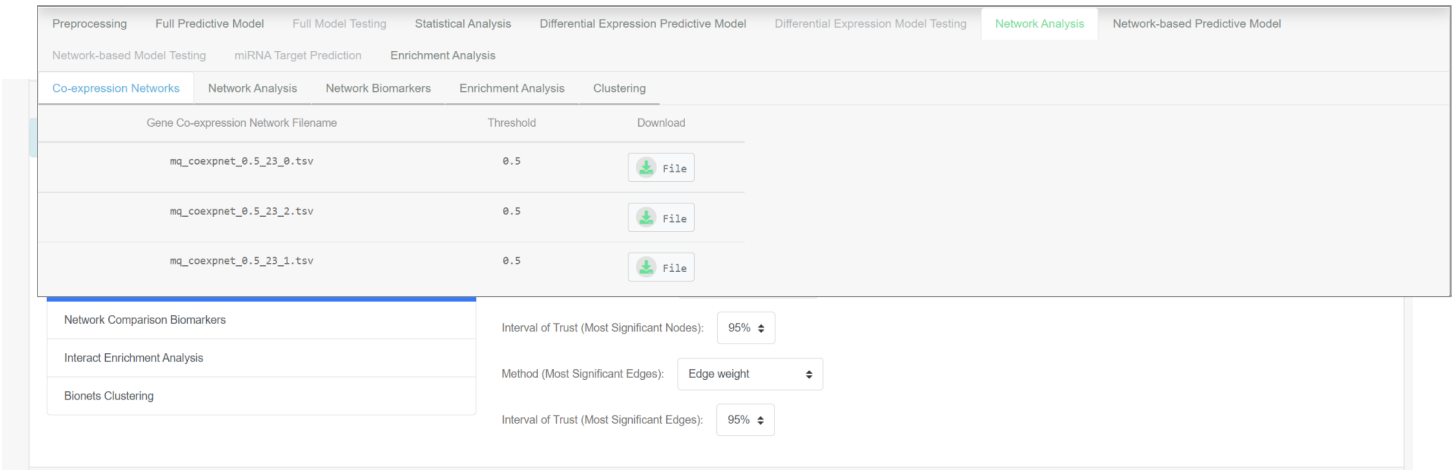
Gene expression files, either uploaded directly by the users or generated through soft files parsing, can be used to generate weighted gene co-expression networks. Experienced users can tune the parameters of the algorithms used for this step and select the most suitable algorithm for them. Three options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

- **Pearson Correlation:** This method adds an edge to a network if the Pearson correlation of the nodes adjacent to the edge exceeds a threshold.
- **Mutual information:** 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:** 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.

To view the results:

In the Co-expression Networks tab you can view and download the created networks.



Biological Network Analysis

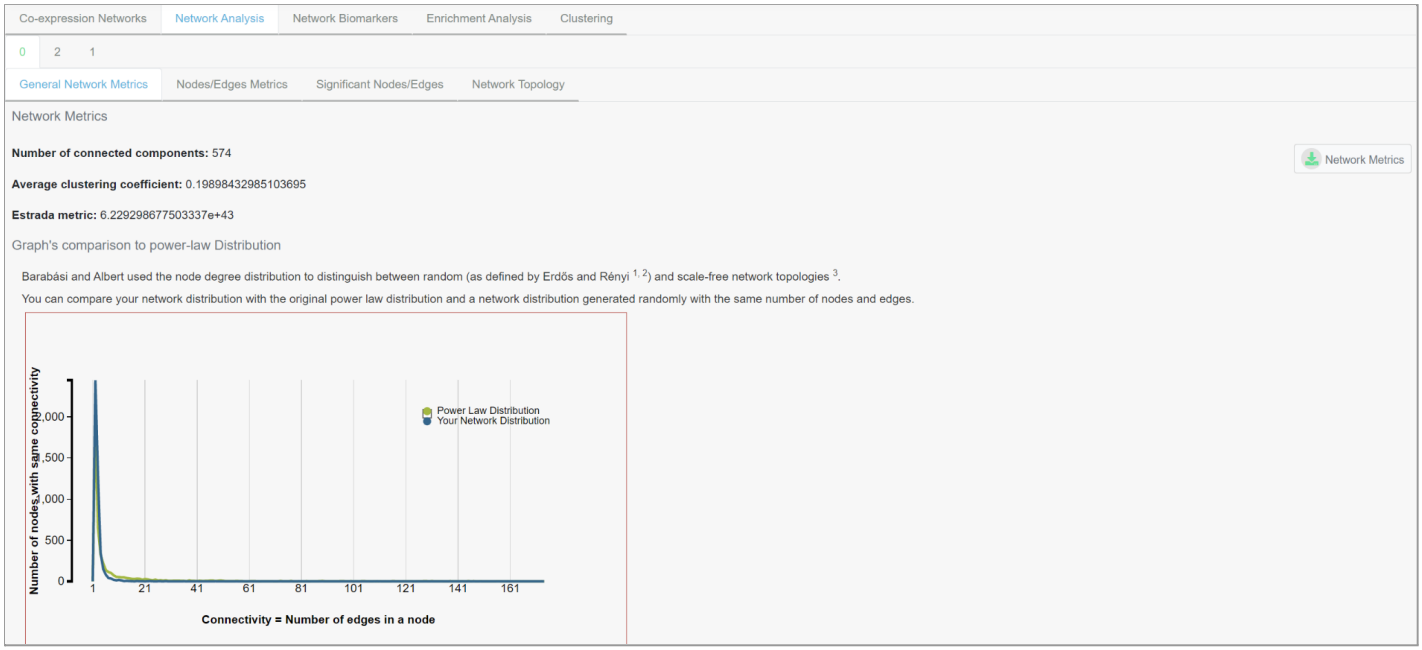
In order to analyze the biological network, the users can tune the following parameters if they are experienced:

- Method for selecting significant nodes (Pagerank (default), Clustering Coefficient, degree centrality])
- Confidence interval for locating significant nodes
- Method for selecting significant edges (Edge weight (default), Inbetweeness centrality)
- Confidence interval for locating significant edges

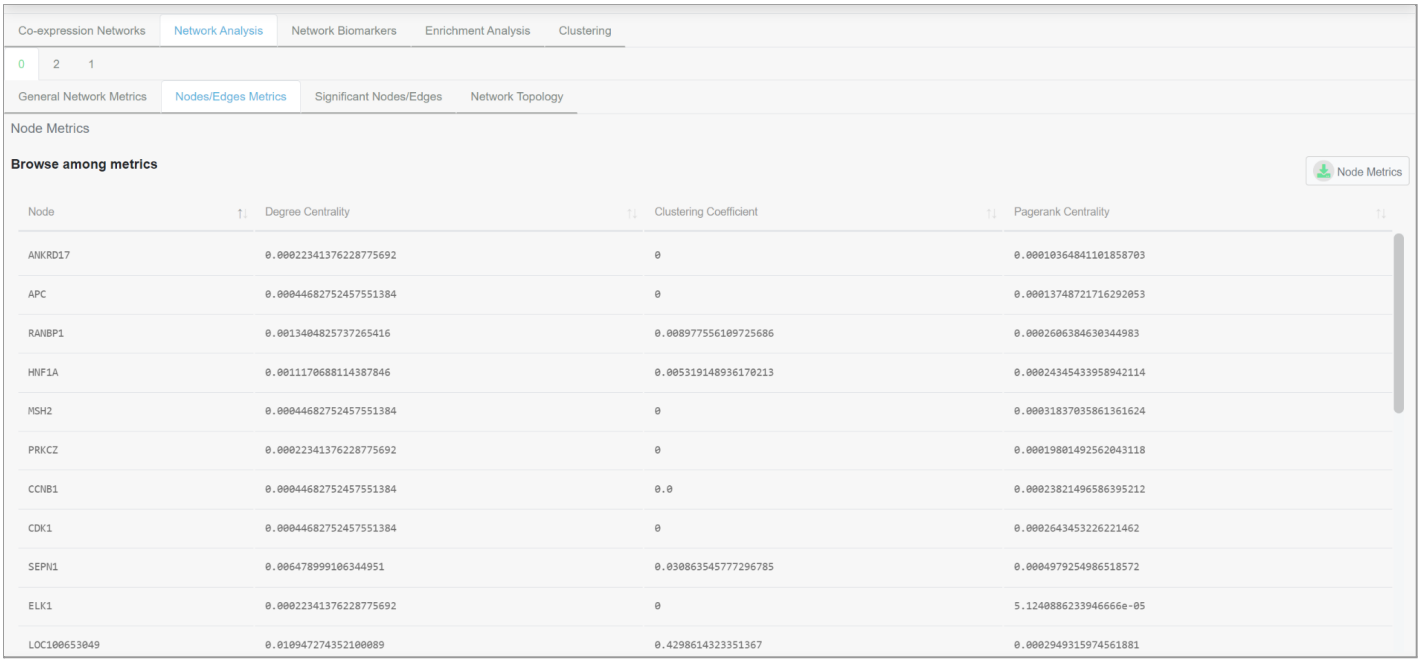
To view the results:

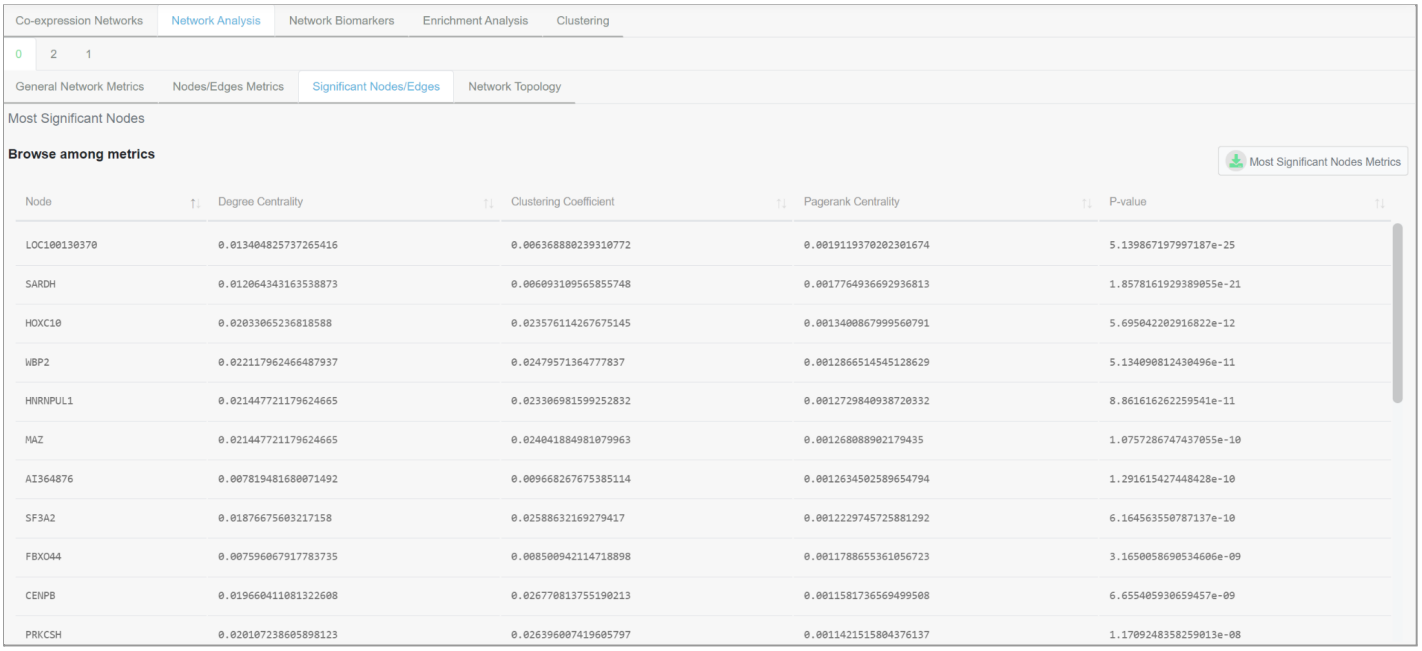
Four new tabs are generated, General Network Analysis, Node/Edges Metrics, Significant Nodes/Edges and Network Topology.

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

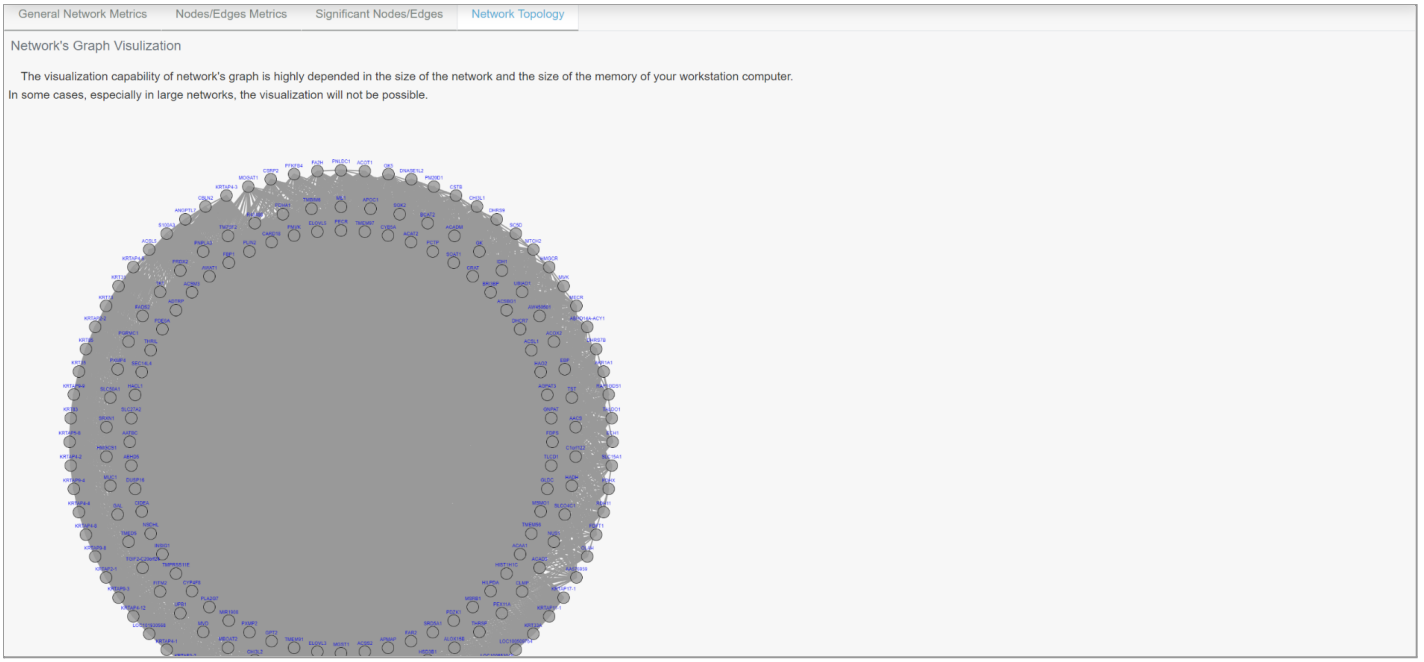


At the Node/Edges Metrics tab users can find the metrics for all nodes (degree centrality, clustering coefficient and pagerank centrality) and edges (edge weight and in betweenness centrality) of your network.

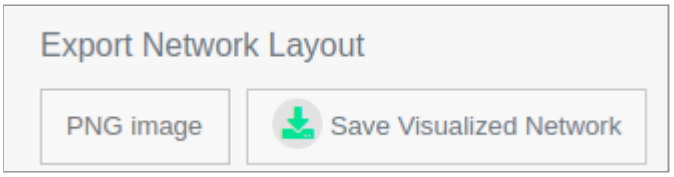




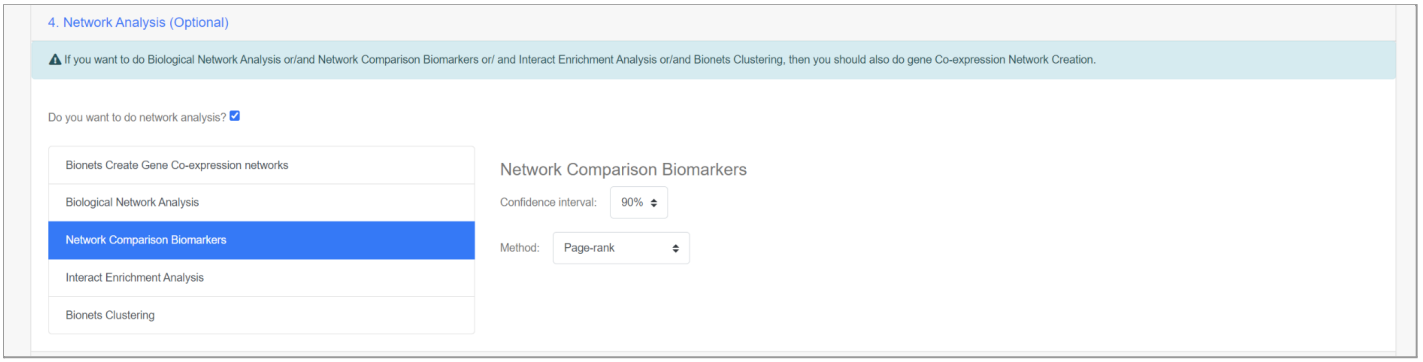
At the Significant Nodes/Edges tab users can access two tables including the significant nodes and edges. For each node and edge, the respective metrics and the p-values of their significance a. Significant edges are not available for networks with more than 225000 edges.



The Network Topology 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 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, JPG), decrease opacity on mouseover and view the network using different visualization layouts (force-directed, circle or radial).



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.

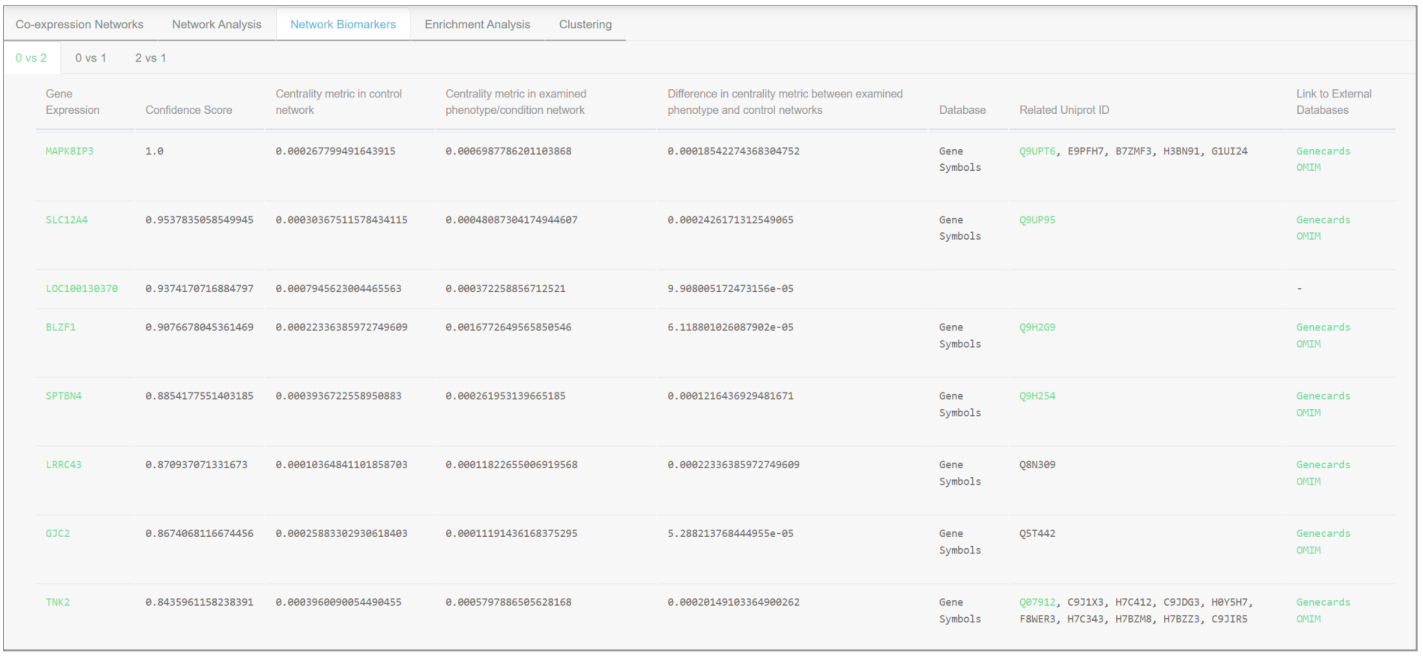
When having two biological networks of different conditions, users can use them to predict network biomarkers with an InSyBio's novel methodology. In particular, a certain network metric is selected and InSyBio BioNets attempts to detect network's nodes with significantly altered values for this network 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
- Clustering Coefficient
- Pagerank method

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.

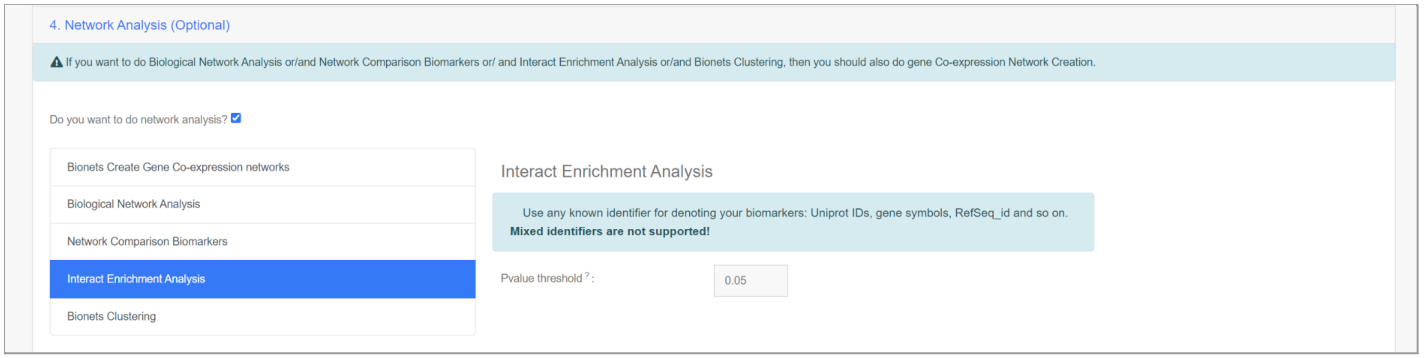
To view the results:



Gene Expression	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
MAPK8IP3	1.0	0.000267799491643915	0.000698778620110386 8	0.000185422743683047 52	Gene Symbols	Q9UPT6, E9PFH7, B7ZMF3, H3BN91, G1UI24	Genecards OMIM
SLC12A4	0.9537835058549945	0.000303675115784341 15	0.000480873041749446 07	0.000242617131254906 5	Gene Symbols	Q9UP95	Genecards OMIM
LOC100130370	0.9374170716884797	0.000794562300446556 3	0.000372258856712521	9.908805172473156e-05			-
BLZF1	0.9076678045361469	0.000223363859727496 09	0.001677264956585054 6	6.118801026087902e-05	Gene Symbols	Q9H2G9	Genecards OMIM
SPTBN4	0.8854177551403185	0.000393672255895088 3	0.000261953139665185	0.000121643692948167 1	Gene Symbols	Q9H254	Genecards OMIM
LRRC43	0.870937071331673	0.000103648411018587 03	0.000118226550069195 68	0.000223363859727496 09	Gene Symbols	Q8N309	Genecards OMIM
GJC2	0.8674068116674456	0.000258833029306184 03	0.000111914361683752 95	5.288213768444955e-05	Gene Symbols	Q5T442	Genecards OMIM
TNK2	0.8435961158238391	0.000396009005449045 5	0.000579788650562816 8	0.000201491033649002 62	Gene Symbols	Q07912, C9J1X3, H7C412, C9JDG3, H0Y5H7, F8WER3, H7C343, H7BZM8, H7BZZ3, C9JIR5	Genecards OMIM

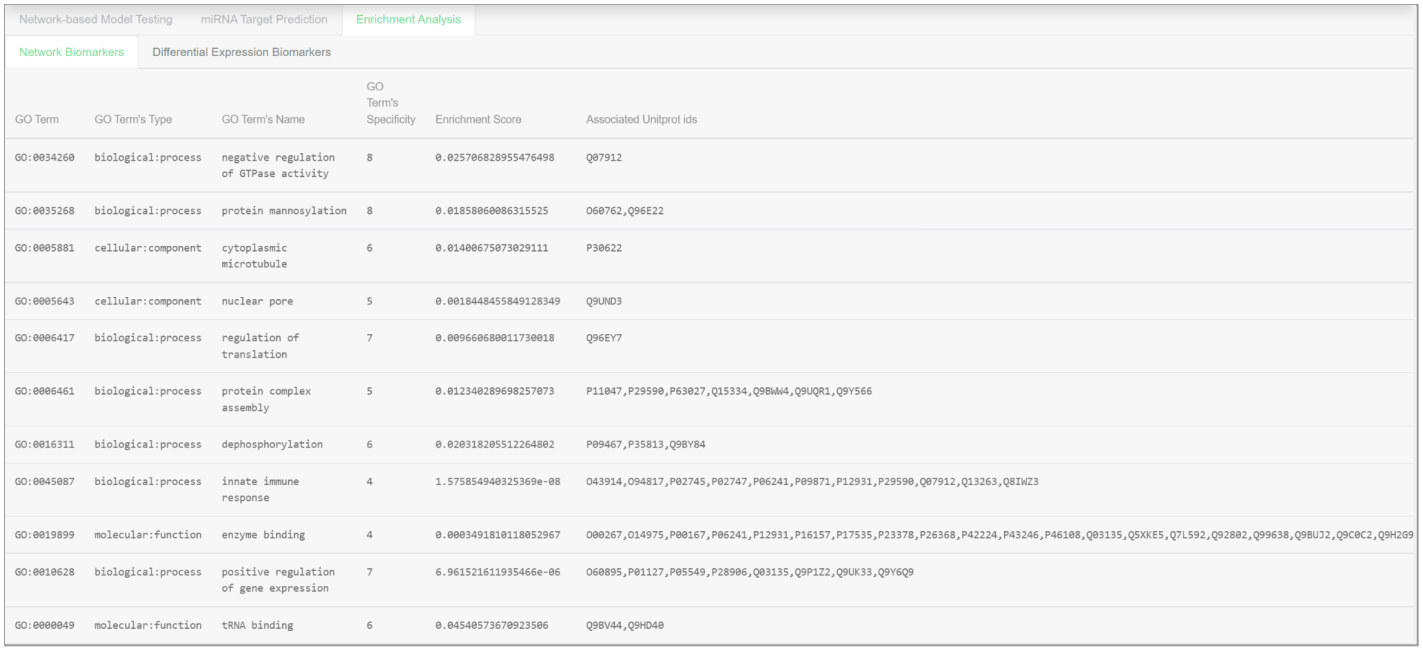
At the Network Biomarkers tab the results are presented in a table with Gene, Confidence Score, Centrality metric in control network, Centrality metric in examined phenotype/condition network, Difference in centrality metric between examined phenotype and control networks, Database, Related Uniprot ID, 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.

Interact Enrichment Analysis



You can perform enrichment analysis with hypergeometric distribution on a given a list of proteins, genes or transcripts and produce a list of GO terms associated with the list, with their term specificity and score in the distribution. You can also provide your custom annotation, term, term type and functional annotation of molecules files, that will be appended to the default files to perform the enrichment. You can define a pvalue threshold for the biomarker to GO terms association output.

To view the results:



Network-based Model Testing | miRNA Target Prediction | Enrichment Analysis

Network Biomarkers | Differential Expression Biomarkers

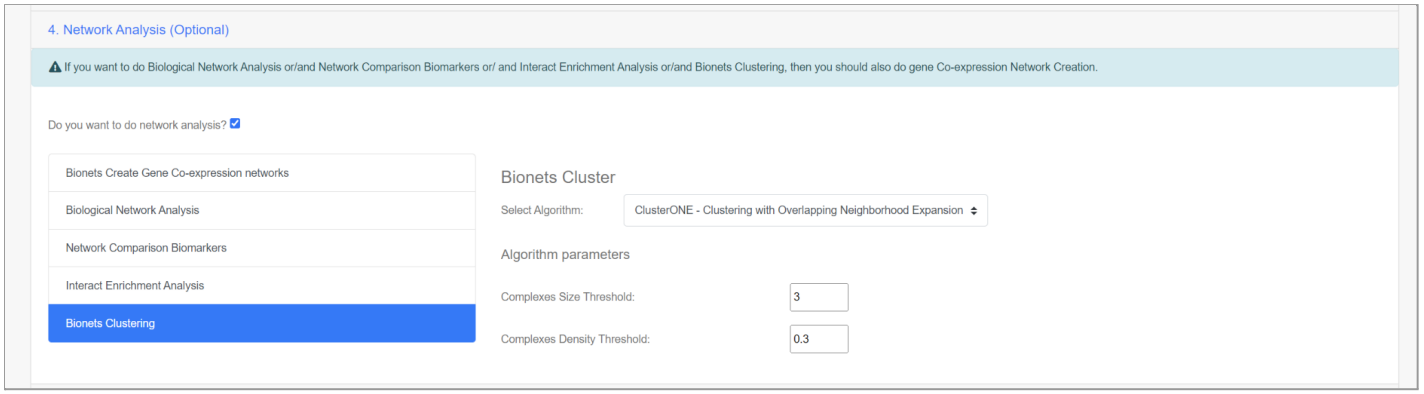
GO Term	GO Term's Type	GO Term's Name	GO Term's Specificity	Enrichment Score	Associated Unitprot ids
GO:0034260	biological:process	negative regulation of GTPase activity	8	0.025706828955476498	Q07912
GO:0035268	biological:process	protein mannosylation	8	0.01858060086315525	O60762,Q96E22
GO:0005881	cellular:component	cytoplasmic microtubule	6	0.01400675073029111	P30622
GO:0005643	cellular:component	nuclear pore	5	0.0018448455849128349	Q9UND3
GO:0006417	biological:process	regulation of translation	7	0.009660680011730018	Q96EY7
GO:0006461	biological:process	protein complex assembly	5	0.012340289698257073	P11047,P29590,P63027,Q15334,Q9BWW4,Q9UQR1,Q9Y566
GO:0016311	biological:process	dephosphorylation	6	0.020318205512264802	P09467,P35813,Q9BY84
GO:0045087	biological:process	innate immune response	4	1.575854940325369e-08	O43914,O94817,P02745,P02747,P06241,P09871,P12931,P29590,Q07912,Q13263,Q8IWZ3
GO:0019899	molecular:function	enzyme binding	4	0.0003491810118052967	O00267,O14975,P00167,P06241,P12931,P16157,P17535,P23378,P26368,P42224,P43246,P46108,Q03135,Q5XKE5,Q7L592,Q92802,Q99638,Q9BUJ2,Q9C0C2,Q9H2G9
GO:0010628	biological:process	positive regulation of gene expression	7	6.961521611935466e-06	O60895,P01127,P05549,P28906,Q03135,Q9P1Z2,Q9UK33,Q9Y6Q9
GO:0000049	molecular:function	tRNA binding	6	0.04540573670923506	Q9BV44,Q9HD40

At the Enrichment Analysis tab you can view the results that are a list of GO terms, terms type and name, specificity, enrichment score, associated Uniprot ids and input ids.

Bionets Clustering

At this step, you can analyze your Biological Network to extract complexes of similar nodes (i.e protein complexes), Weighted and unweighted Biological Networks.

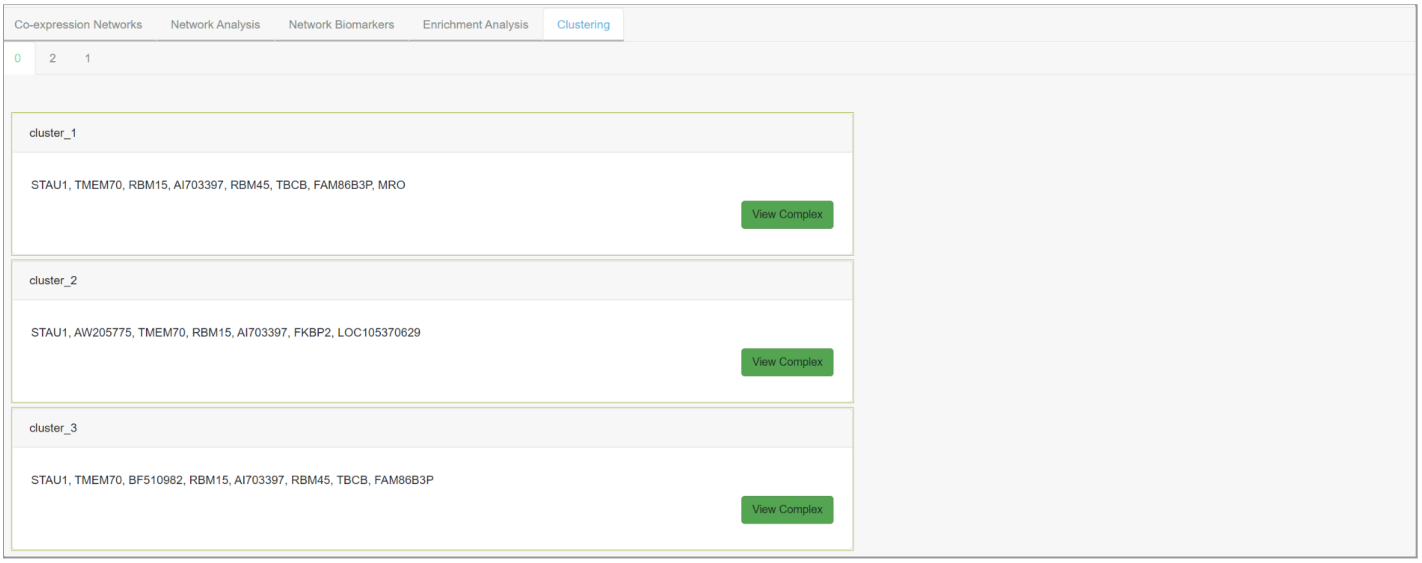
For the prediction of Biological Network complexes one option is supplied:



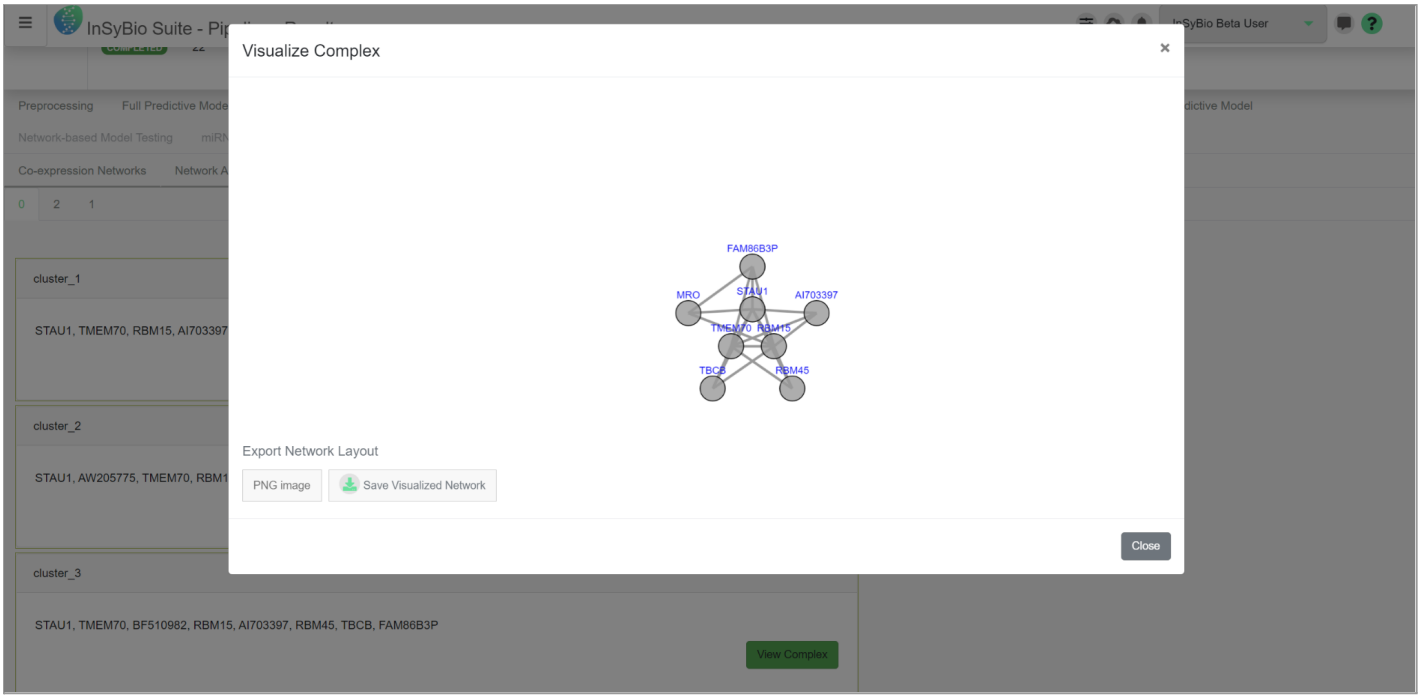
Clustering with Overlapping Neighborhood Expansion (ClusterONE).

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

To view the results:



At the Clustering tab you can view the different network clusters that are computed. You can also visualize them by clicking “View Complex”.



ncRNAseq Predict

The fifth step is the ncRNAseq Predict step. This step is also optional and with this step you can computationally predict potential miRNA targets at given Genes or Transcripts and given miRNAs. BLAST is performed in order to find potential target sites, and then the computational intelligent technique, which was applied for the prediction of miRNAs (hybrid combination of Genetic Algorithms and epsilon-SVRs), and 124 informative features are used in order to calculate a prediction score.

For this step:

- Select miRNAs and the Genes you want to search for potential targets by searching in our Database and adding them to the miRNA List and Genes List or add them manually to their Lists and separating them with commas.

To view the results:

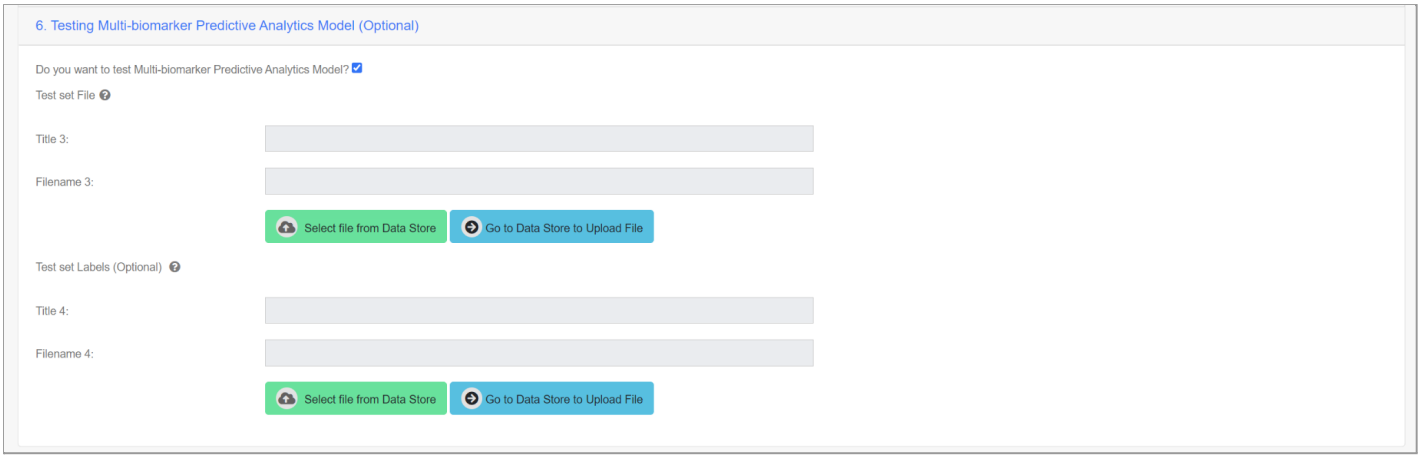
At the tab miRNA Target Prediction the results are presented on your screen in a browse-able table, with each miRNA and gene pair in a row with their confidence score. By pressing Details at the Actions Column the specific scores between the miRNA and the gene's transcripts can be viewed. If no target sites are found "No targets found!" is presented at the score column. If one or more target sites are found you can view its UTR sequence, with the target sites of the miRNA highlighted. Multiple target sites are marked with green color and unique target sites are marked with light blue.

Testing Multi-biomarker Predictive Analytics Model

The sixth and final step is the Testing Multi-biomarker Predictive Analytics Model step. This step is also optional and allows the users to test the predictors that they have trained in the previous "training" step.

The first input file is the test dataset, which can be preprocessed or not preprocessed. The second input are the testset labels, which is optional. If the user inserts the testset labels then he'll receive as an output along with the predicted labels the performance metrics of the prediction.

It should be noted that the input dataset must have the format of the previous functionalities, that is it should have as rows the features and as columns the samples.



At the end, the user should click the Submit Job button to start the job.

To view the results:

At the Full Model Testing tab the user will be getting the following results. He'll view the predicted labels and the performance metrics. For the two class classification problem (two-class, multi-class) he'll view the test set accuracy, specificity, sensitivity, f1 score, f2 score and ROC AUC score. For the multiclass classification problem he'll view the test set accuracy, f1 score, precision, recall, f2 score and Manhattan Distance. For the regression he'll view the testset mean squared error, the test set relative absolute error and the root relative squared error.

How to get InSyBio Pipelines

To request a free one month full (evaluation) version of InSyBio Suite please email us at info@insybio.com .

To purchase InSyBio Interact commercial version 3.0 please contact us at sales@insybio.com.

About Us

InSyBio Ltd is a bioinformatics pioneer company (www.insybio.com) in precision medicine and nutrition, 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