Analyze Protein-Protein Interactions with InSyBio Interact

October 2025

Insybio Suite v3.4

InSyBio

Intelligent Systems Biology

User Manual

www.insybio.com

Introduction

Interact is an intracellular interaction analysis tool for the

- analysis of proteins
- analysis of protein interactions
- prediction of proteins' functionality

Proteins are the functional components of many cellular processes and the identification of their physical Protein-Protein Interactions (PPIs) is an area of mature academic research.

InSyBio Interact enables users to combine a variety of sequential, structural, and functional information using a high-performance machine learning technique to predict a confidence and interaction affinity score for each protein pair. These confidence scores indicate the probability of these two proteins interacting and the strength of this interaction. Users are able to tune this confidence score and extract their own datasets as well as obtain only positive or negative interactions. Moreover, InSyBio Suite also provides visualization and meta-analysis of the Protein Interaction Networks. The protein interaction networks are clustered using a novel methodology to extract information about the protein complexes that the proteins form to perform specific tasks. These complexes and the overall PPI networks may be visualized using an interactive visualization module.

With InSyBio Interact you can:

a) Search Positive/Negative PPIs
b) Retrieve information about a specific protein
c) View and download the list of clusters created by the EEMC algorithm
d) Search proteins and Complexes with a GO Term
e) Create and Save your own PPI dataset
f) Create a network from your set of biomarkers (genes, transcripts, proteins)
g) Perform functional enrichment analysis on your set of biomarkers
h) Annotate functionally a specific protein using a patent-pending method [Theofilatos, K., Dimitrakopoulos, C., Mavroudi, S., Korfiati, A. and Alexakos, C., Insybio Inc, 2017.

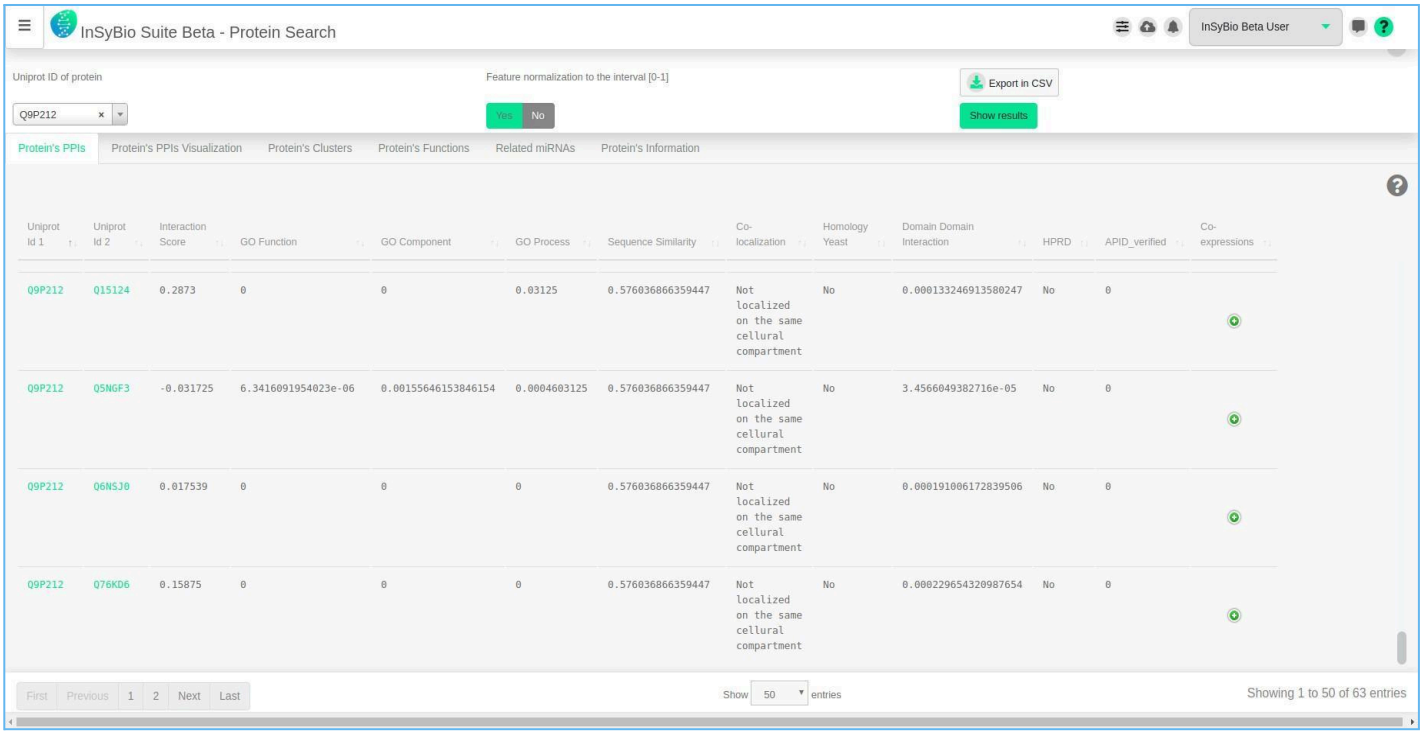
Protein functional and sub-cellular annotation in a proteome. U.S. Patent Application 15/361,461.]

Protein Search

You can extract information about a specific Human protein by providing its UNIPROT identifier.

You can:

a) Request the PPIs that are related to the specific protein and download the results;



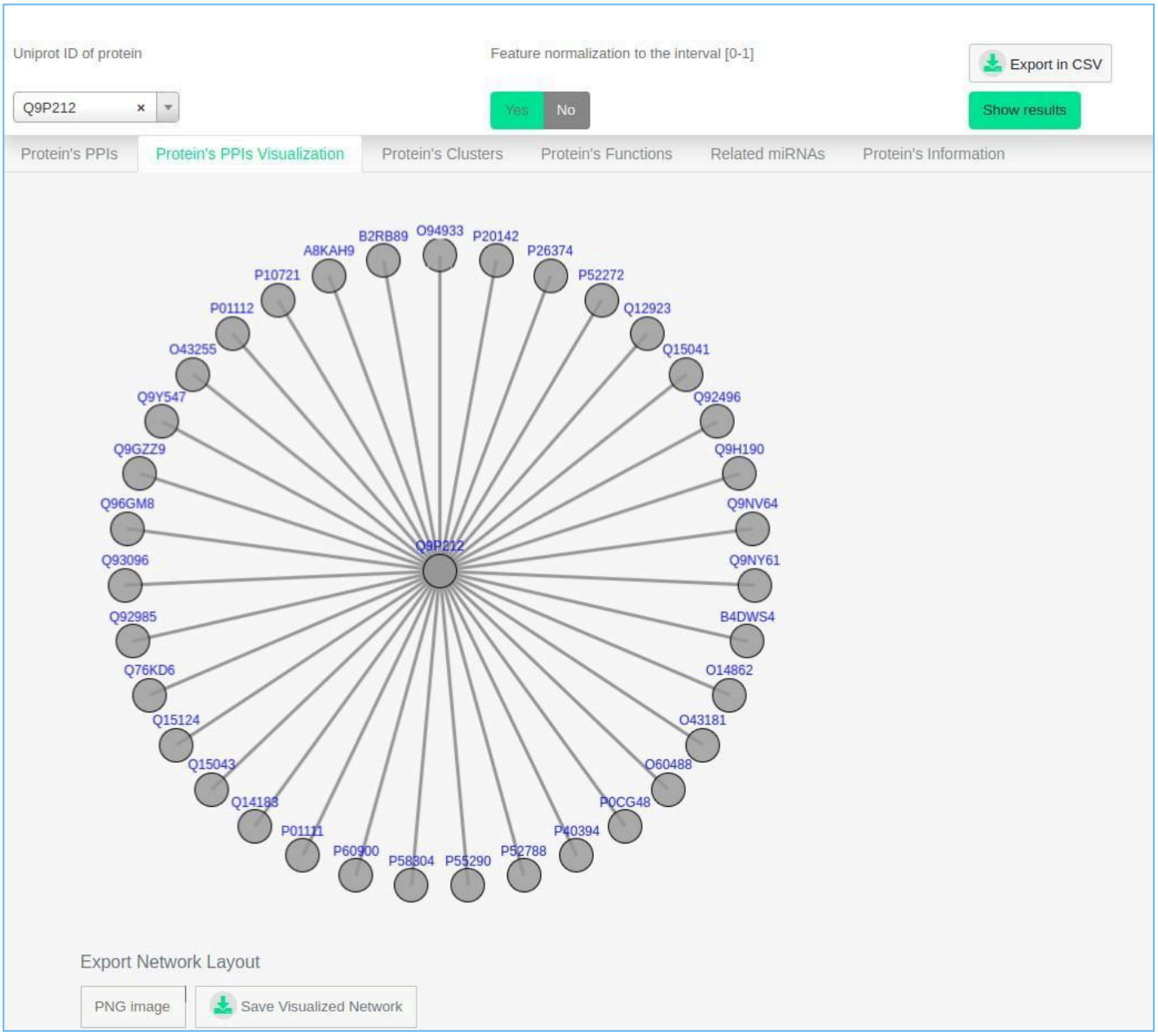
The results are presented on your screen in a browse-able table or you can download them as a TAB delimited CSV file.

For each PPI the following are presented:

- The Uniprot IDs of the two proteins,
- 22 informative sequential, structural and functional features,

- Presence in the APID database,
- The protein-protein interaction score - a confidence score indicating the probability of a protein pair to be an interacting one and the strength of this interaction.

b) Visualize the protein interaction network related to the specific protein and download it in various graph-representation file formats;



The result is an interactive graph displaying the first or the second neighborhood of the protein of interest (after user choice).

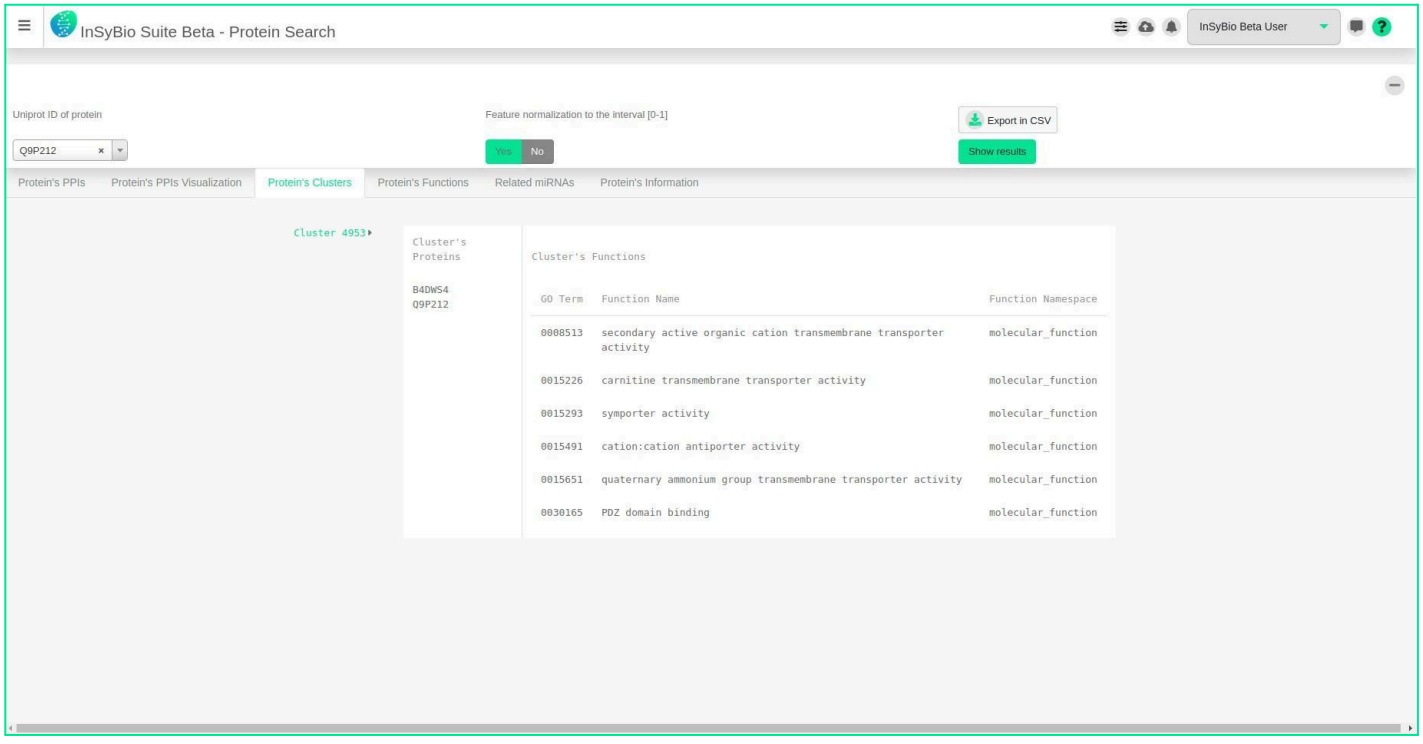
Clicking on a node-protein or on an edge-interaction, the respective information is presented.

Right clicking on a node-protein, its first neighborhood is presented.

There are options for decreasing opacity on mouseover and changing layout (force directed, radial, or circle).

The graph can be saved in various formats (PNG image, SVG image, PDF file, XGMML document, GraphML document, or Simple Interaction Format (SIF) document).

c) Find the clusters related to the specific protein;



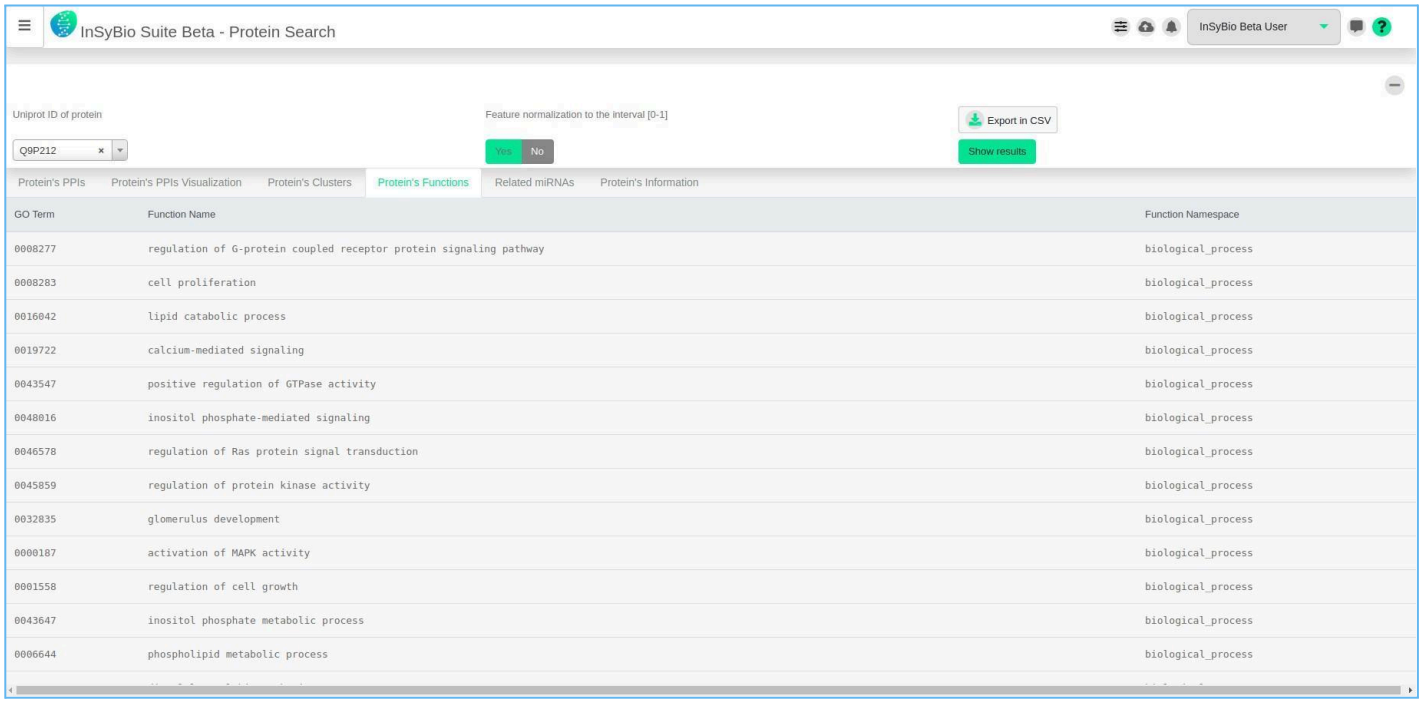
The result is a list of the clusters the protein of interest takes part in.

Clicking on a cluster, the following information is presented:

- The proteins that form the cluster,

- A table of the cluster's functions.

d) List all the functions related to the specific protein;

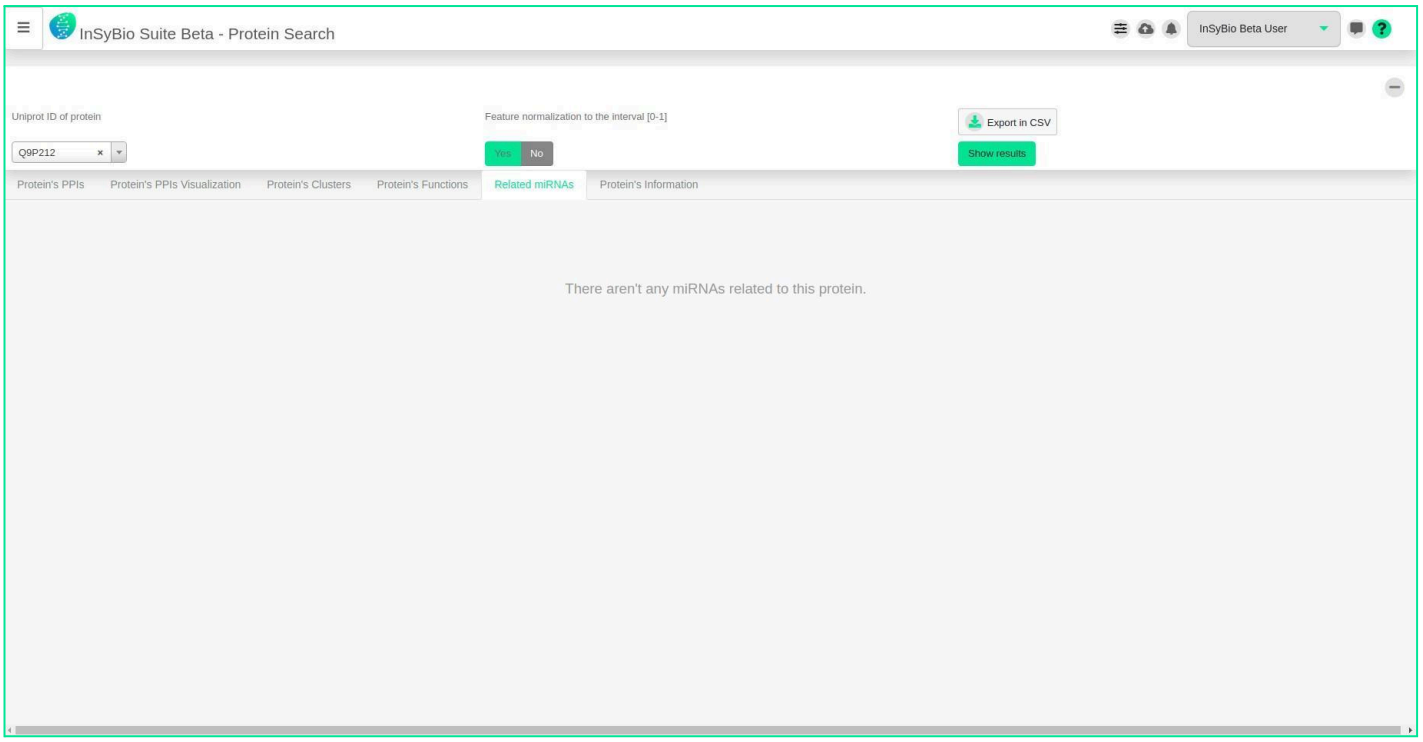


The result is a table with the functions that the protein of interest has.

For each function, the following information is presented:

- The function's GO term,
- The function's name, and
- The function's namespace.

e) List all the miRNAs related to the specific protein;

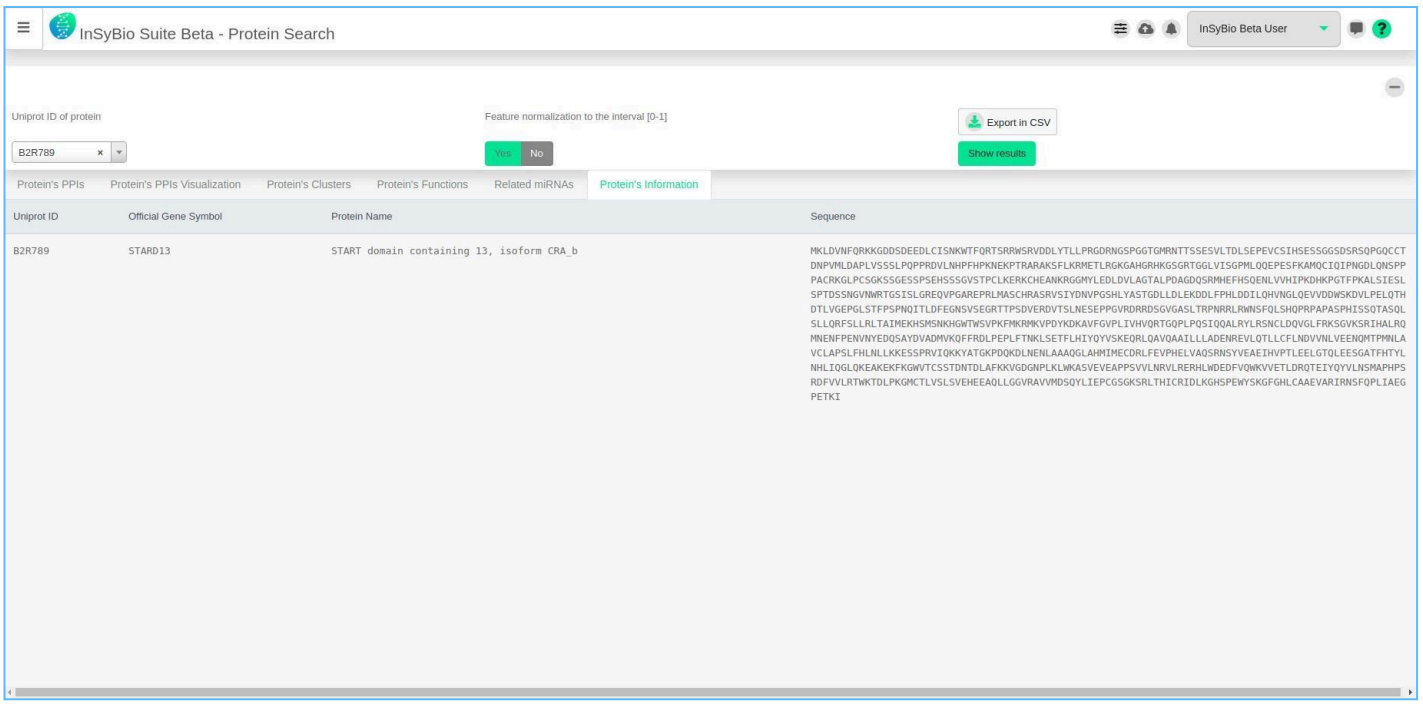


The result is a table with the miRNAs that regulate the expression of the protein of interest.

For each miRNA, the following information is presented:

- The miRNA's accession number, and
- The miRNA's id.

f) Find the protein's information;



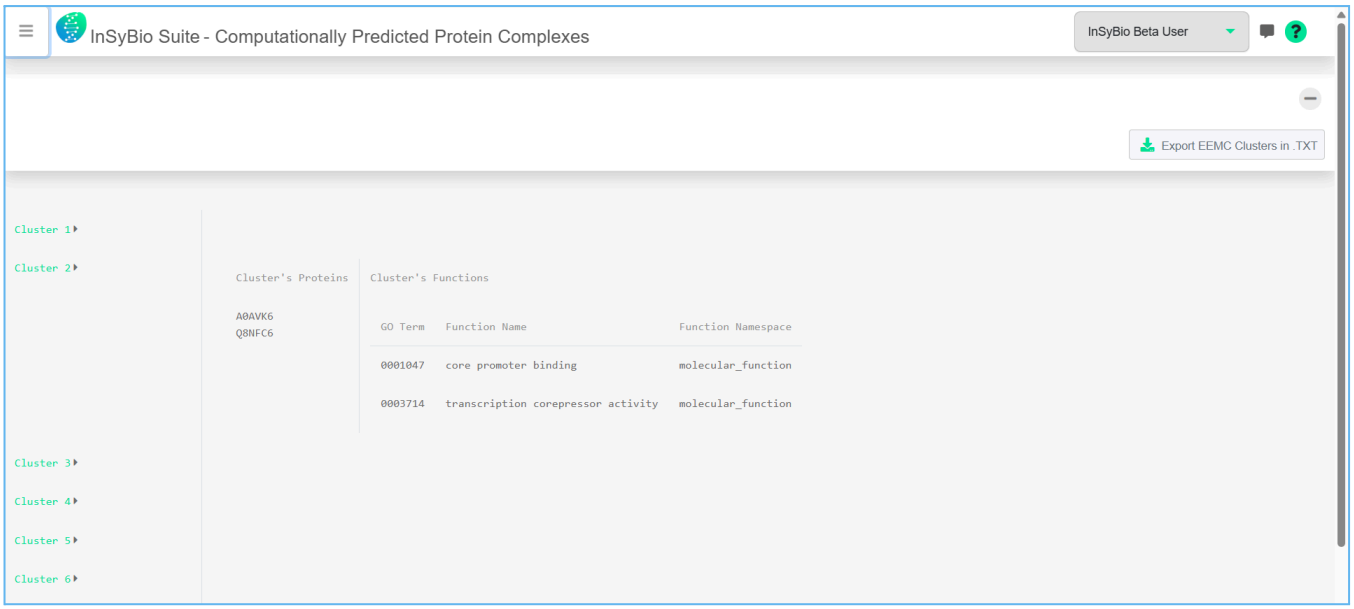
The result is a table containing information about the protein of interest.

For each protein, the following information is presented:

- The UniProt ID,
- The official gene symbol
- The protein name, and
- The protein's sequence.

Computationally Predicted Protein Complexes

You can view a list of all the clusters computed from the algorithm EEMC. EEMC (Evolutiona Enhanced Markov Clustering) is a novel fully unsupervised methodology that InSyBio team has propos for the prediction of protein complexes from weighted PPI graphs. It is a hybrid combination of adaptive evolutionary algorithm and a state-of-the-art clustering algorithm named Enhanced Mark Clustering. When applied to new human datasets EEMC' s performance was encouraging in t prediction of protein complexes which consist of proteins with high functional similarity. From EEMC analysis, we suggest new potentially true Human protein complexes which should be further validat using experimental techniques.



The result is a list of the clusters calculated using the EEMC algorithm.

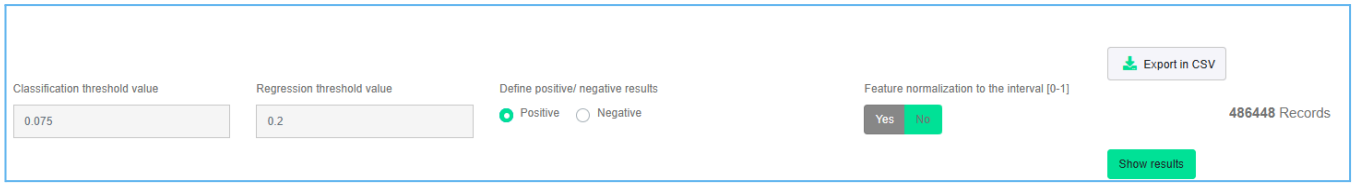
Clicking on a cluster, the user can be informed about the proteins that form the cluster and the cluster's functions.

The entire set of clusters can be downloaded in a txt file with the following format:

```
cluster_id,proteins,go_terms
1,A0AV96 Q9C0H2,0005254
2,Q8NFC6 A0AVK6,0001047 0003714
3,A0AVT1 Q7Z6D5,0019780 0016874
...
```

Positive/Negative PPI Search

You can search positive and negative Human Protein-Protein Interaction Data by specifying threshold referring to the PPI classification confidence score and the affinity of the interaction.

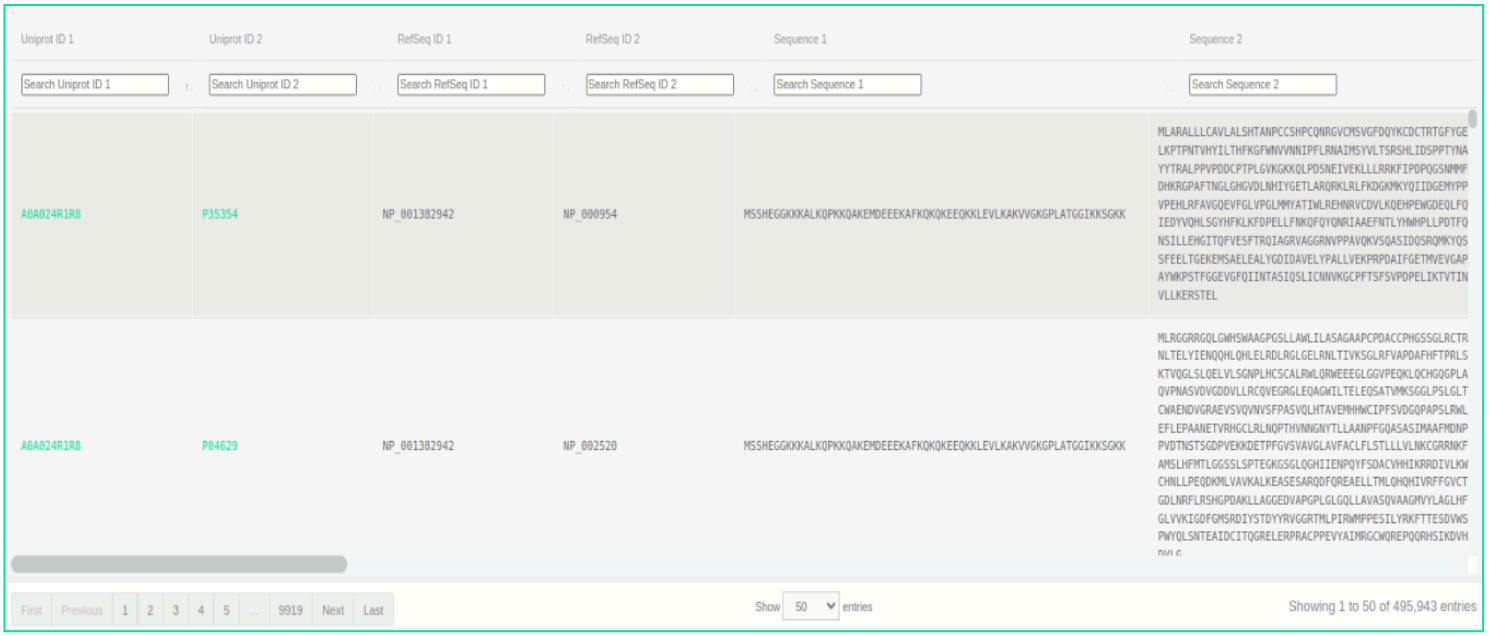


You can search by choosing:

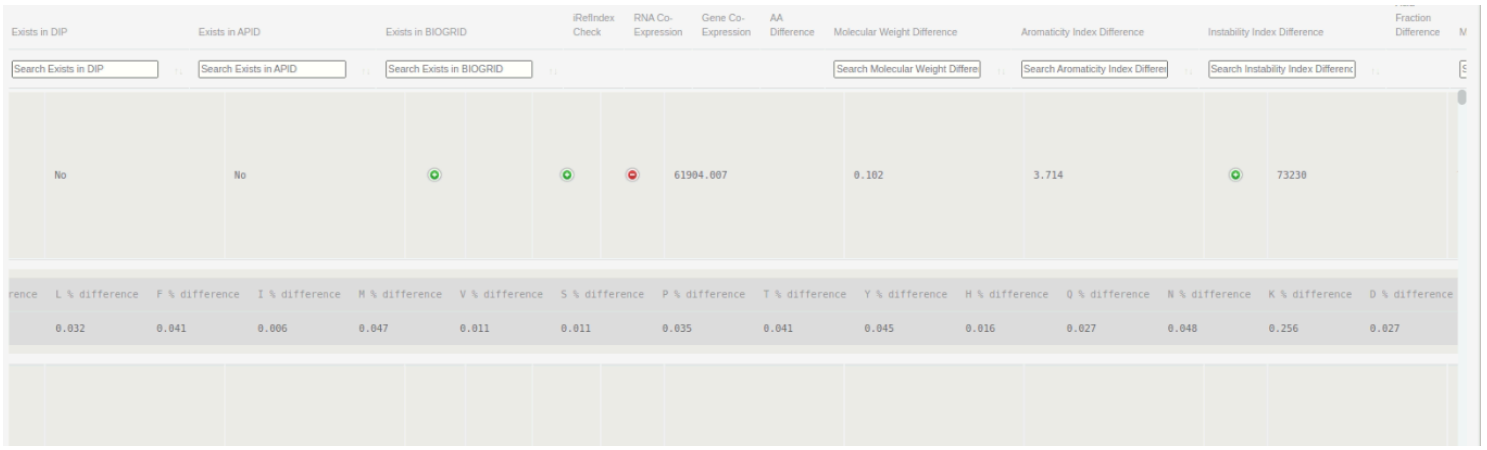
- A classification confidence threshold value
- A regression (affinity) threshold value
- Positive (interacting) or Negative (non-interacting) PPIs
- Normalization to the [0,1] interval or not

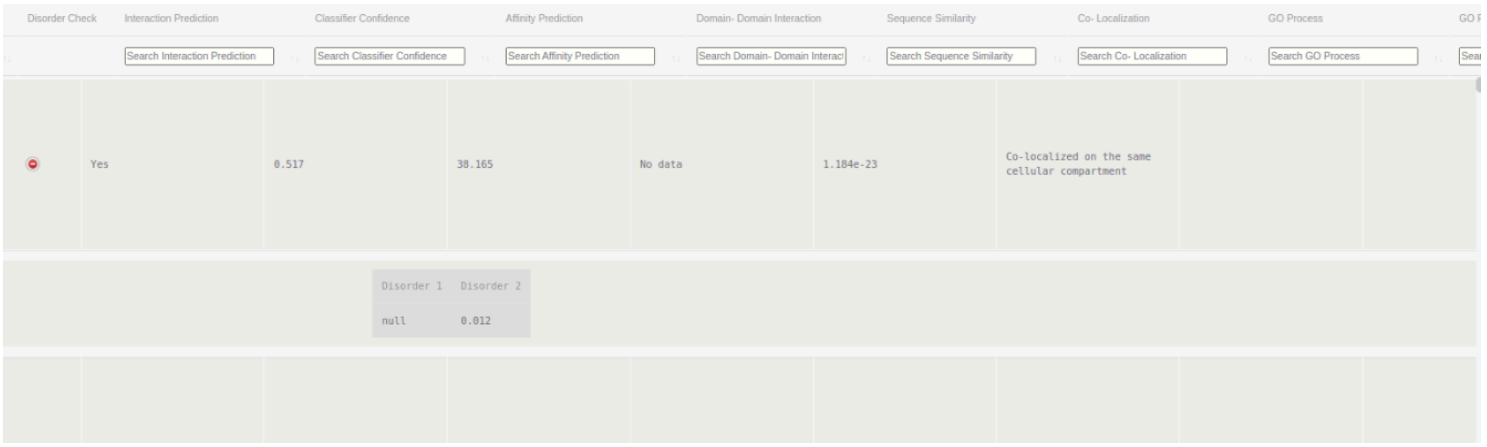
And you get a table view of the selected PPIs along with their informative features:

General View of the Table

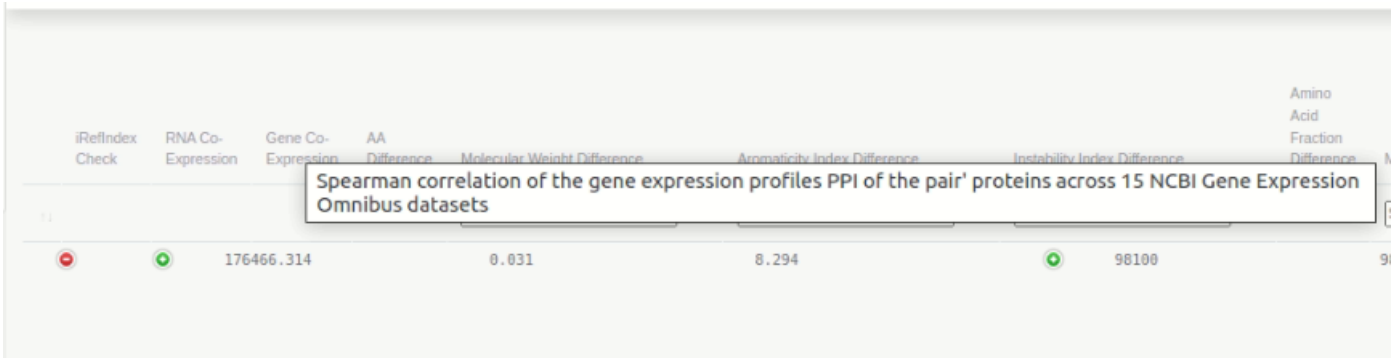


Feature Groups with Pop-Up features





Hover cursor over features for more info



The results are presented on your screen in a browsable table, or you can download them as a TAB-delimited CSV file.

- For each PPI, the following are presented:
- The UniProt IDs, RefSeq IDs, and Sequences of the two proteins
- 75 total informative sequential, structural, and functional features
- Presence in other databases (iRefIndex, Mint, APID, BIOGRID, DIP)
- Two protein-protein interaction scores - the classifier's probability score and affinity predictions for each interaction.

For every protein pair, 75 informative features are calculated and provided, including sequential, structural, and functional features:

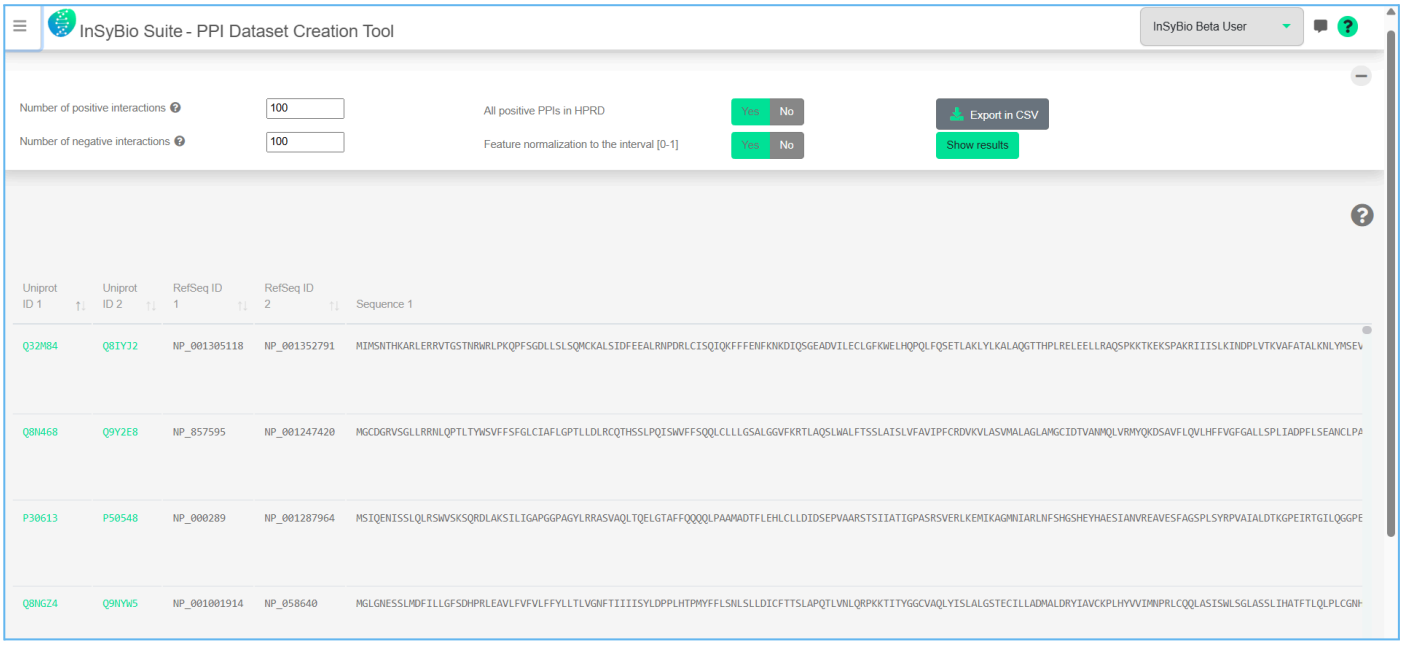
- **PPI descriptive features** (6 features): UniProt IDs, RefSeq IDs, and Sequences of the two proteins.
- **Disorder Check** (3 features): Checks the intrinsic disorder of the two proteins. If one of the two proteins has disorder >0.5, the rest of the features aren't calculated for that pair.
- **Interaction Scores** (2 features): Classifier Confidence and Affinity Prediction.
- **Domain- Domain Interaction** (1 feature): Presence of known domain interactions extracted from Pfam Database, a database of protein families and domains, each represented by multiple sequence alignments and hidden Markov models (HMMs). Interactions between Pfam IDs are derived from the 3did DB.
- **Sequence Similarity** (1 feature): The E-value of the sequence similarity
- **Co-Localization** (1 feature): Subcellular co-localization of the two proteins in the pair. The database used for extracting information about protein localizations in the eukaryotic cell is eSLDB.
- **GO term similarity** (GO Process, GO Function, GO Component – 3 features): The similarity of the proteins in the pair based on the similarity of their respective GO terms. The GO terms are filtered, keeping only those that are indicative of a biological process, of a molecular function, and of a cellular component. Similarity is measured using the Lin method.
- **Orthology in other organisms** (Orthology Mouse, Drosophila, Yeast, Ecoli – 4 features): Existence of a corresponding orthologous protein in other organisms. PPIs from the other organisms are derived from the DIP database, and the mapping of homologous proteins was done via NCBI- HomoloGene Dataset.
- **Existence in other Databases** (Exists in DIP, in APID, in BIOGRID, in Mint – 4 features): These columns indicate if each PPI is recorded in the DIP, APID, BIOGRID, and Mint databases as

well.

- **iRefindex Check** (3 features): These columns indicate if each PPI is recorded in the iRefIndex database, which was originally used for training and testing the classification models. For each PPI also appearing in iRefidex, the experimental method of detection (Detection Method) and the PMID of the citation (PMID of Interaction) are also referenced.
- **RNA Co-Expression** (2 features): These features calculate the similarity between each protein in the pair, in terms of their RNA expression profiles. The RNA expression datasets used were GSE227375 and GSE228702 from the NCBI Gene Expression Omnibus. Spearman correlation is used for calculating the expression similarity of each PPI pair.
- **Gene Co- Expression** (15 features): These features represent the similarity of the two proteins in terms of their Gene Expression among fifteen NCBI Gene Expression Omnibus datasets (GDS531, GDS534, GDS596, GDS651, GDS806, GDS807, GDS843, GDS987, GDS1085, GDS2855, GDS1402, GDS181, GDS1088, GDS841, GDS3257). Spearman correlation is used for calculating the expression similarity of each PPI pair.
- **AA Difference** (20 features): The difference in the percentage of every amino acid in the PPI proteins' sequences.
- **Molecular Weight Difference** (1 feature): The difference in molecular weight between the PPI proteins.
- **Aromaticity Index Difference** (1 feature): The difference in aromaticity index between the PPI proteins.
- **Instability Index Difference** (1 feature): The difference in instability index between the PPI proteins.
- **Amino Acid Fraction Difference** (3 features): The difference in fraction of total amino acids that are contained in the helix (Helix Difference), in the turn (Turn Difference), and in the sheet (Sheet Difference)
- **Molar Extinction Coefficient Difference** (2 features): The difference in molar extinction coefficient (MEC) when it is calculated assuming cysteines reduced (MEC Reduced Difference), and assuming cysteine residues (MEC Residues Difference).
- **GRAVY Index Difference** (1 feature): The difference in GRAVY (Grand Average of Hydropathy) between the PPI proteins.
- **Protein Charge Difference** (1 feature): The difference in protein charge at pH=7 between the PPI proteins.

PPI Datasets

You can create your own datasets by using a simple form to choose the number of positive and negative examples.



There are options for:

- All positive PPIs being in HPRD or not,
- The application of normalization to the features to the interval [0-1].

The results are presented on your screen in a browse-able table or you can download them as a TAB delimited CSV file.

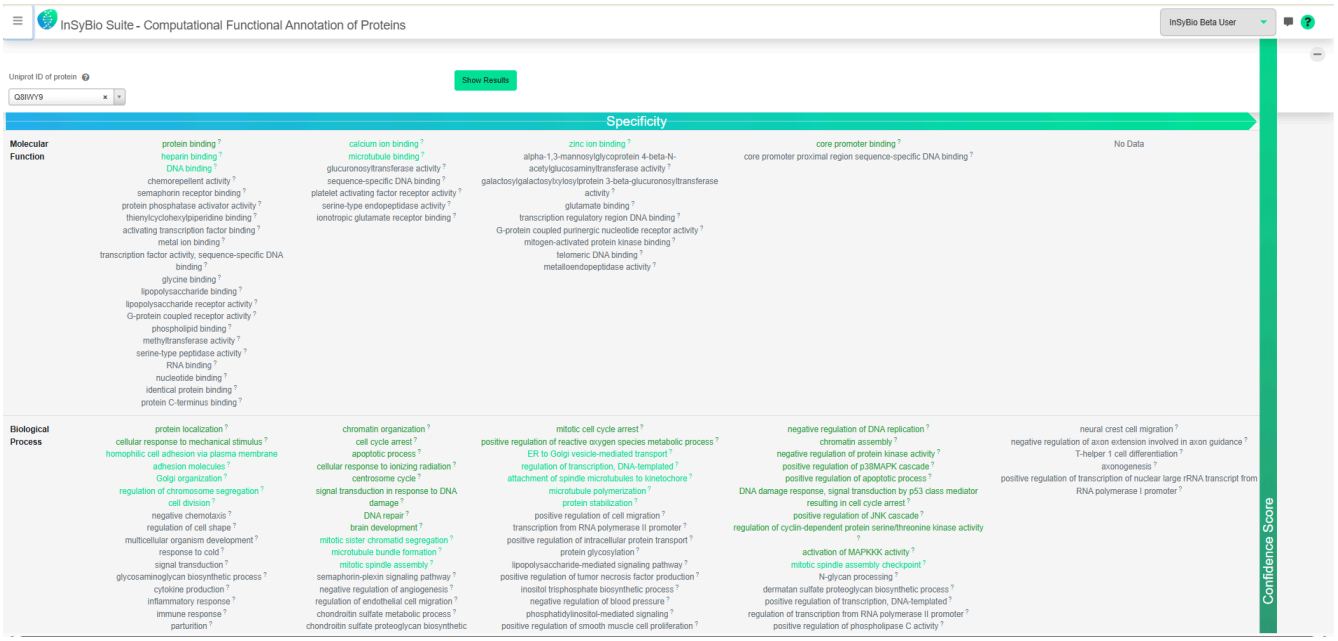
For each PPI the following are presented:

- The Uniprot IDs of the two proteins,

- 22 informative sequential, structural and functional features,
- Presence in the APID database,
- The protein-protein interaction score - a confidence score indicating the probability of a protein pair to be an interacting one and the strength of this interaction.

Computational Functional Annotation of Proteins

You can view the confirmed and predicted functional annotation of a given protein, from our database.



You can provide the Uniprot ID of your protein of interest, and the results of our Functional Annotation Patented method will appear. The associated GO terms with the protein are separated into 3 rows, one for each major term group (molecular function, biological process and cellular component terms). Going from left to right, the specificity of the term increases. Each term' s position in a cell represents its predicted confidence score in descending order, dark green are confirmed associations (score 1), light green high confidence and gray medium confidence predicted associations (with score near 0.5).

GO Functional Annotation

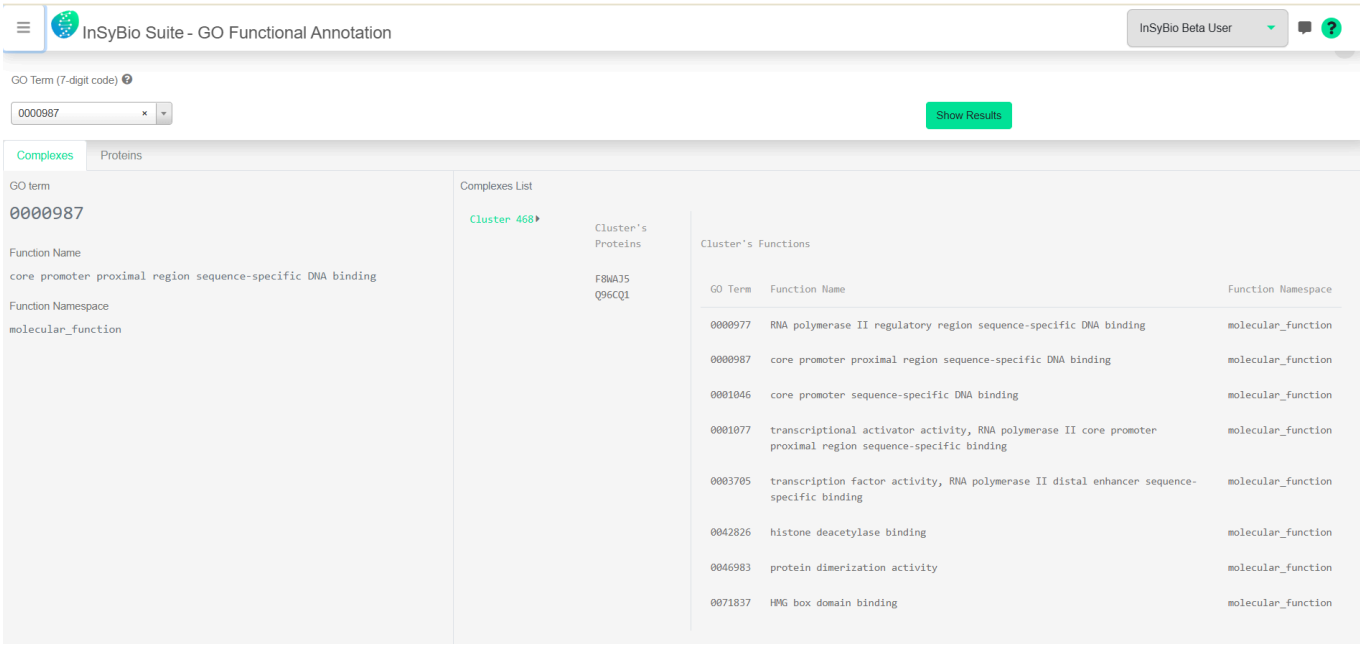
You can derive the details of your function of interest by providing its GO-term.

For each function, the following information is presented:

- The function's GO term,
- The function's name, and
- The function's namespace.

Selecting the respective tab, you can also view the related:

a) Complexes;

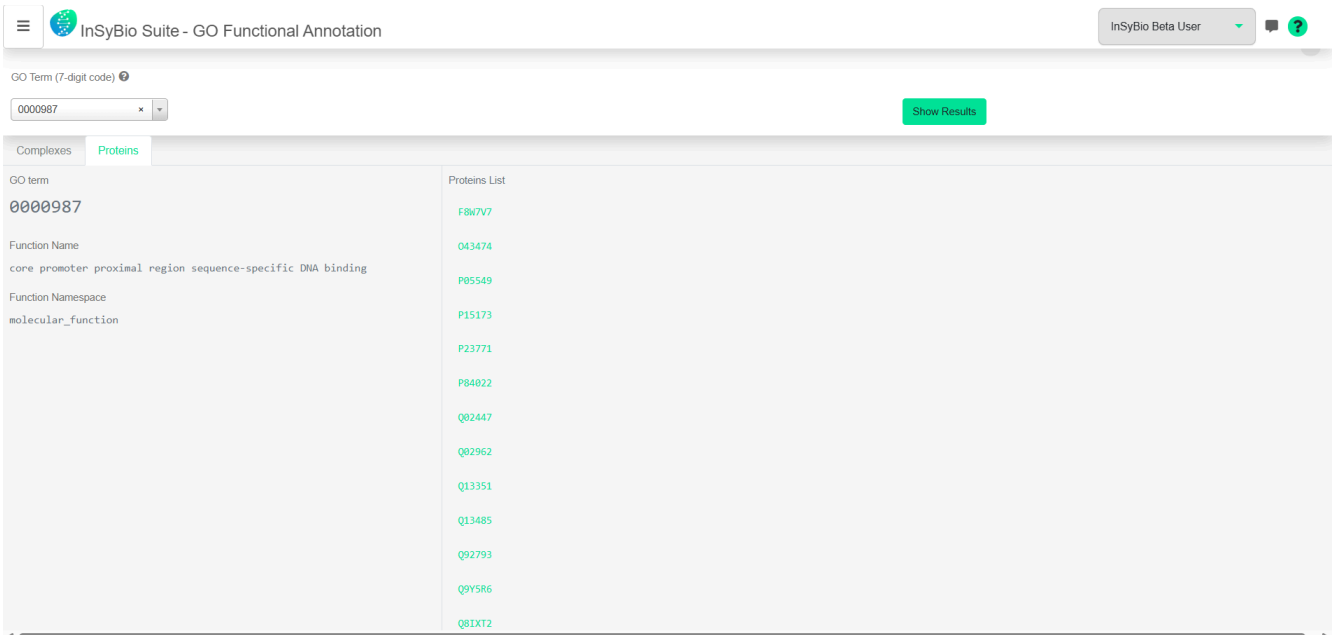


The result is a list of the complexes having the function of interest.

Clicking on a cluster, the following information is presented:

- The proteins that form the cluster, and
- A table of the cluster's functions.

b) Proteins;

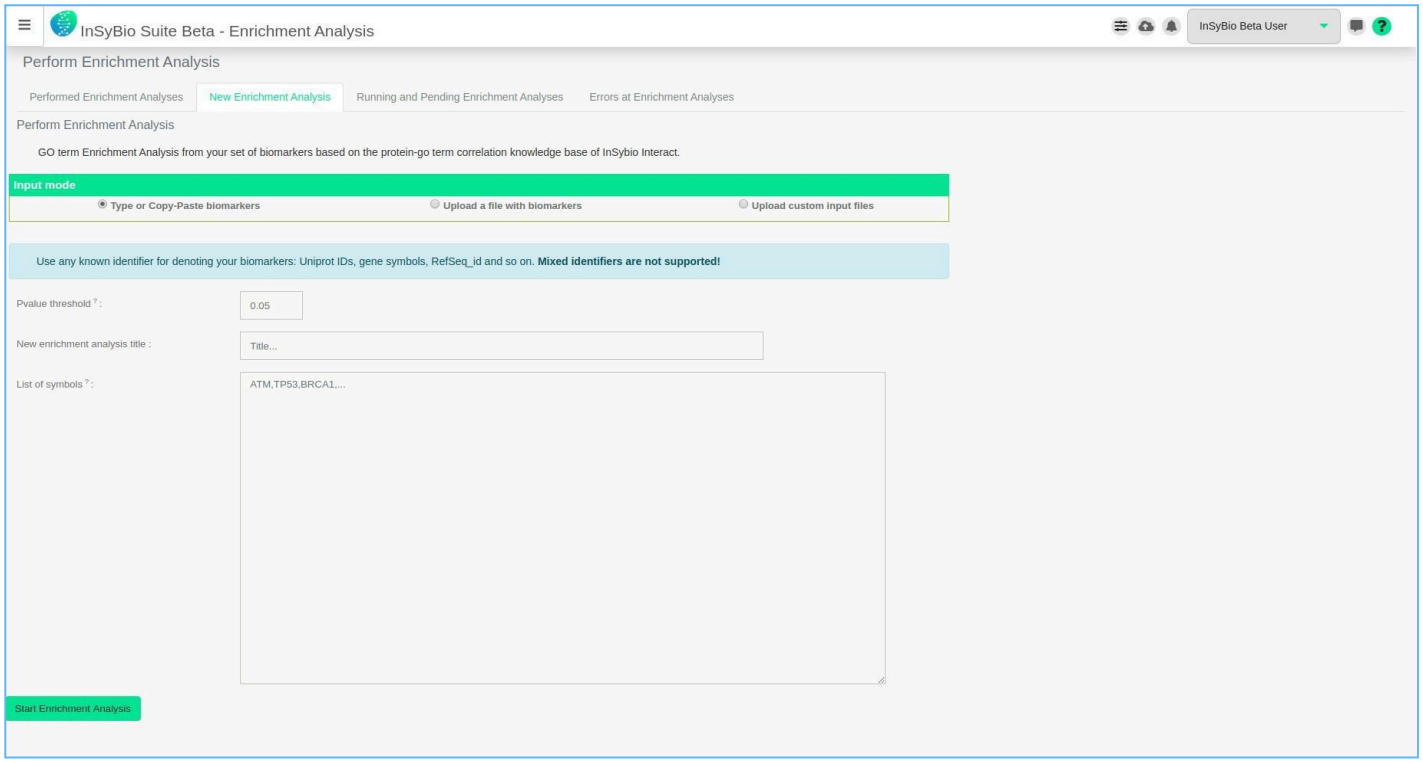


The result is a list of the proteins having the function of interest. Clicking on a protein the user will be guided to the protein's page.

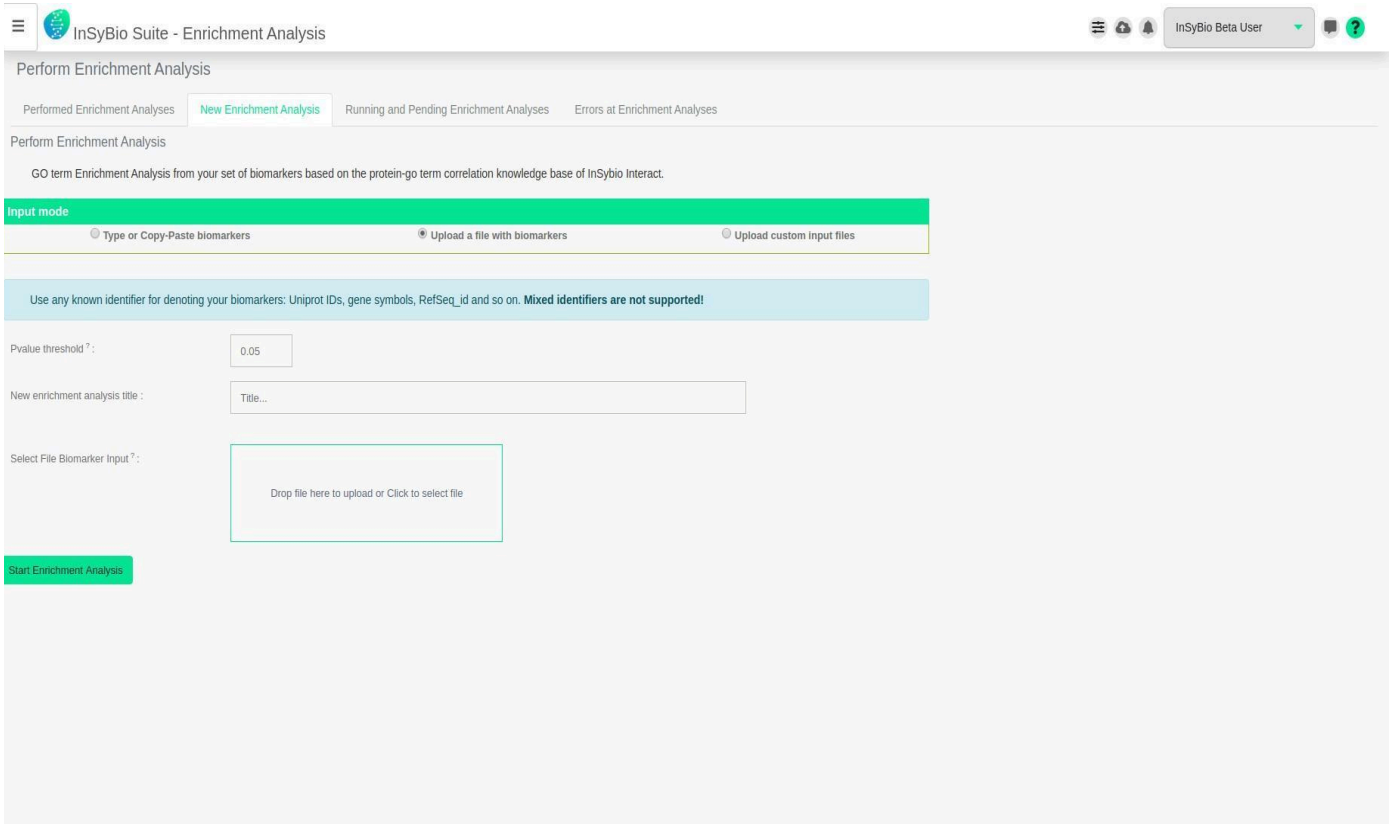
Enrichment Analysis Functionality

You can perform enrichment analysis with hypergeometric distribution on a given 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.

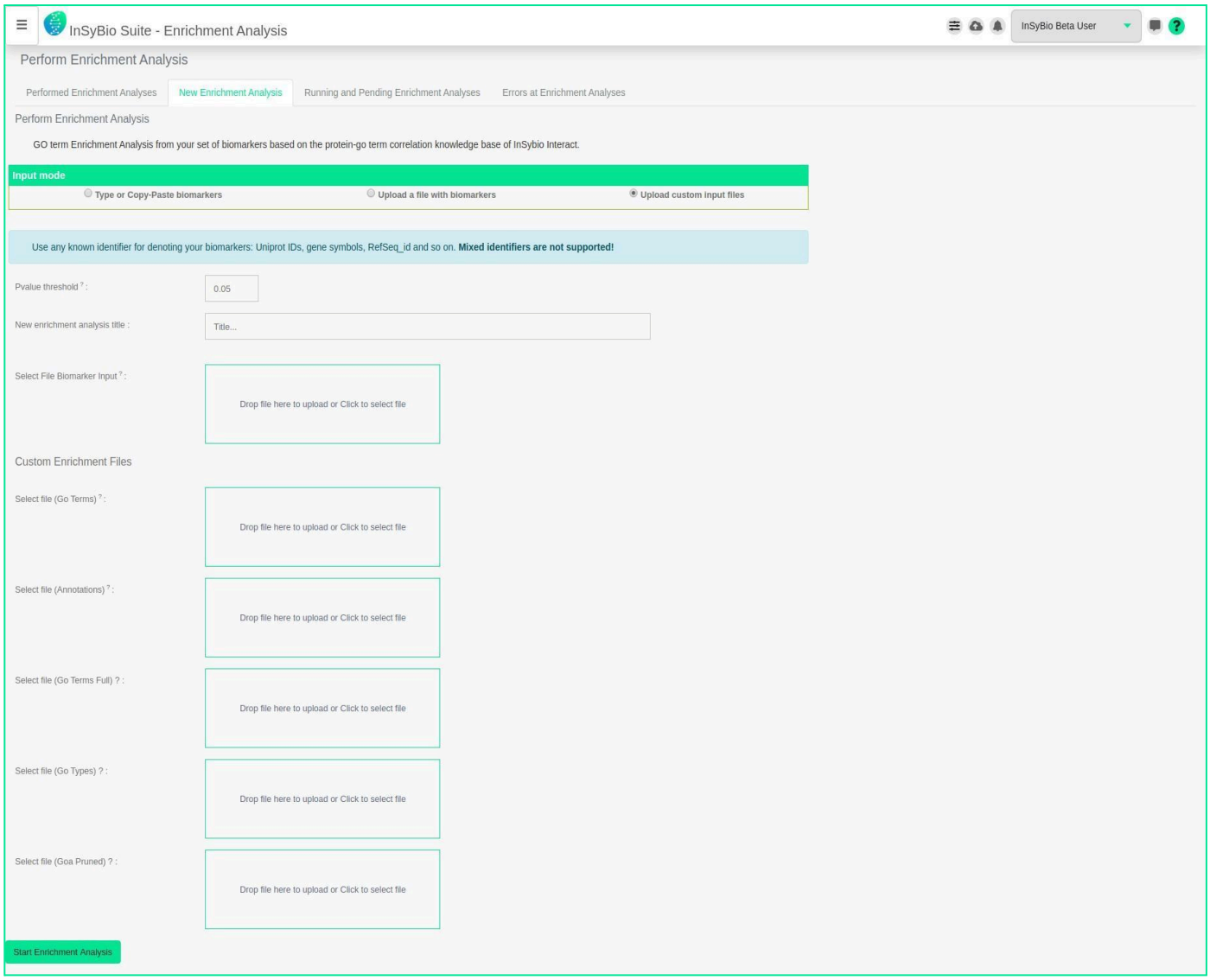
In the tab New Enrichment Analysis you can:



- Either Type or Copy-Paste biomarkers or



- Upload a file with biomarkers or

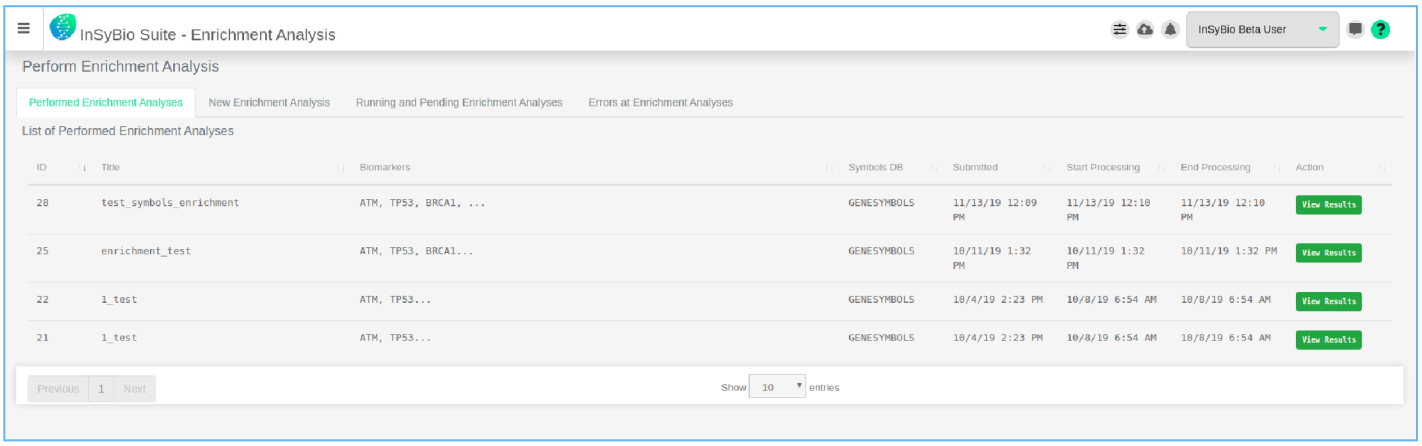


- Upload a file with biomarkers and other custom annotation, term, term type and functional annotation of molecules files

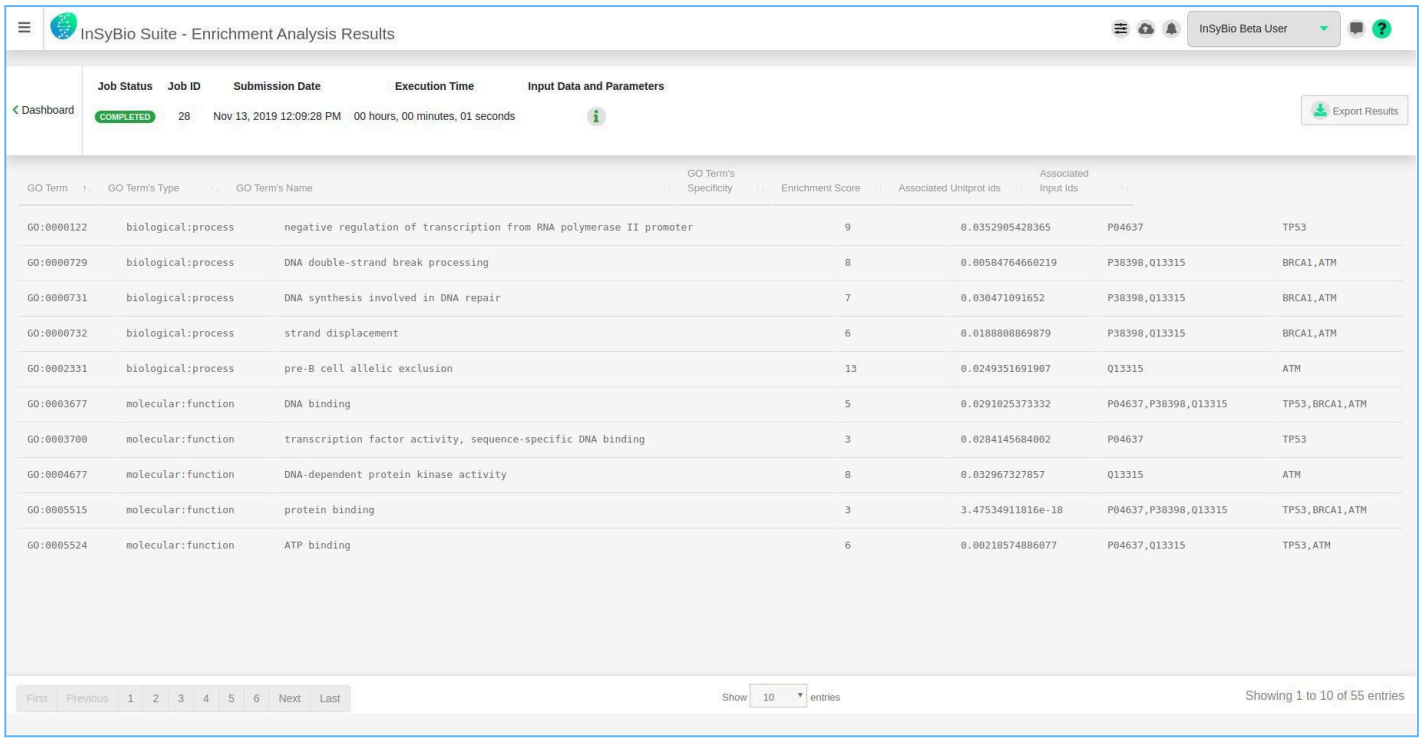
You can use any known identifier for denoting your biomarkers: Uniprot IDs, gene symbols, RefSeq_id and so on. Mixed identifiers are not supported! Note that each symbol should be in one line or separated by comma.

You can also define a pvalue threshold for the biomarker to GO terms association output.

After the enrichment analysis, the job will appear in the Performed Enrichment Analyses tab.



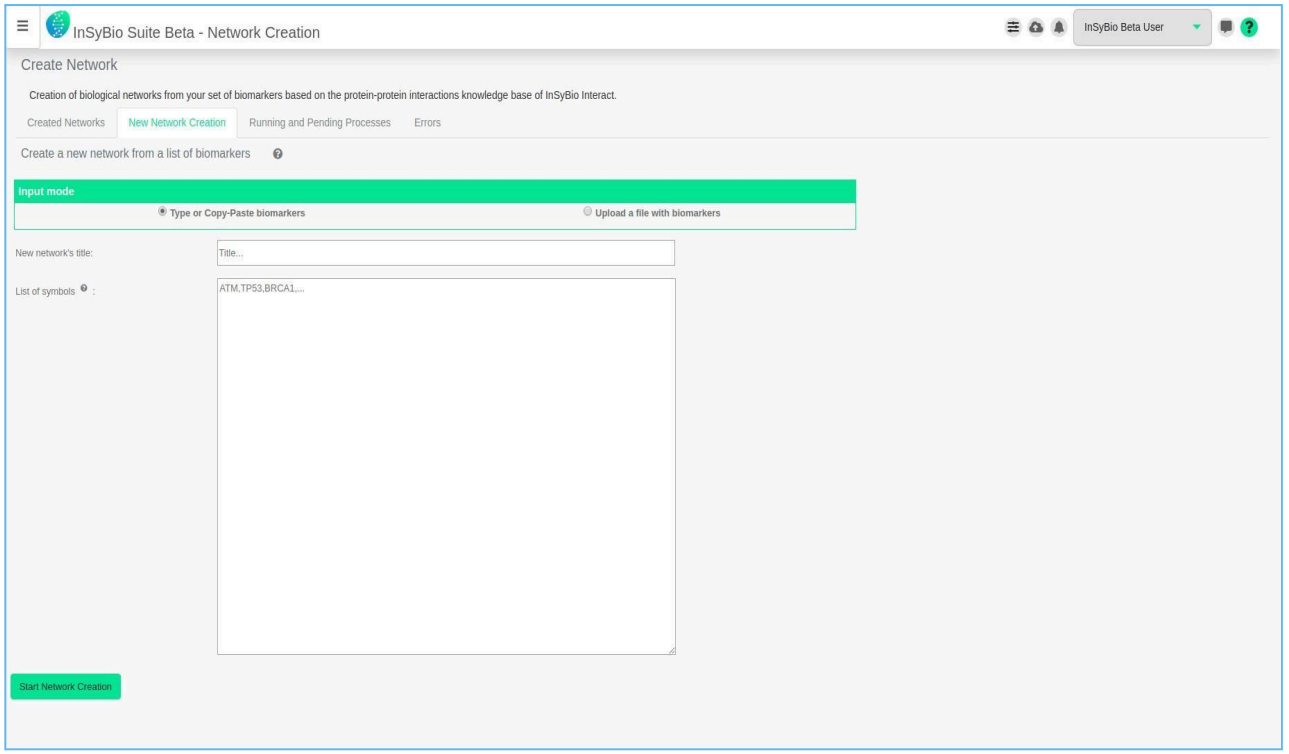
By clicking View Results you will be redirected to the Enrichment Analysis Results window, where you can view the results and download them in a tab delimited file. The results are a list of GO terms, terms type and name, specificity, enrichment score, associated Uniprot ids and input ids.



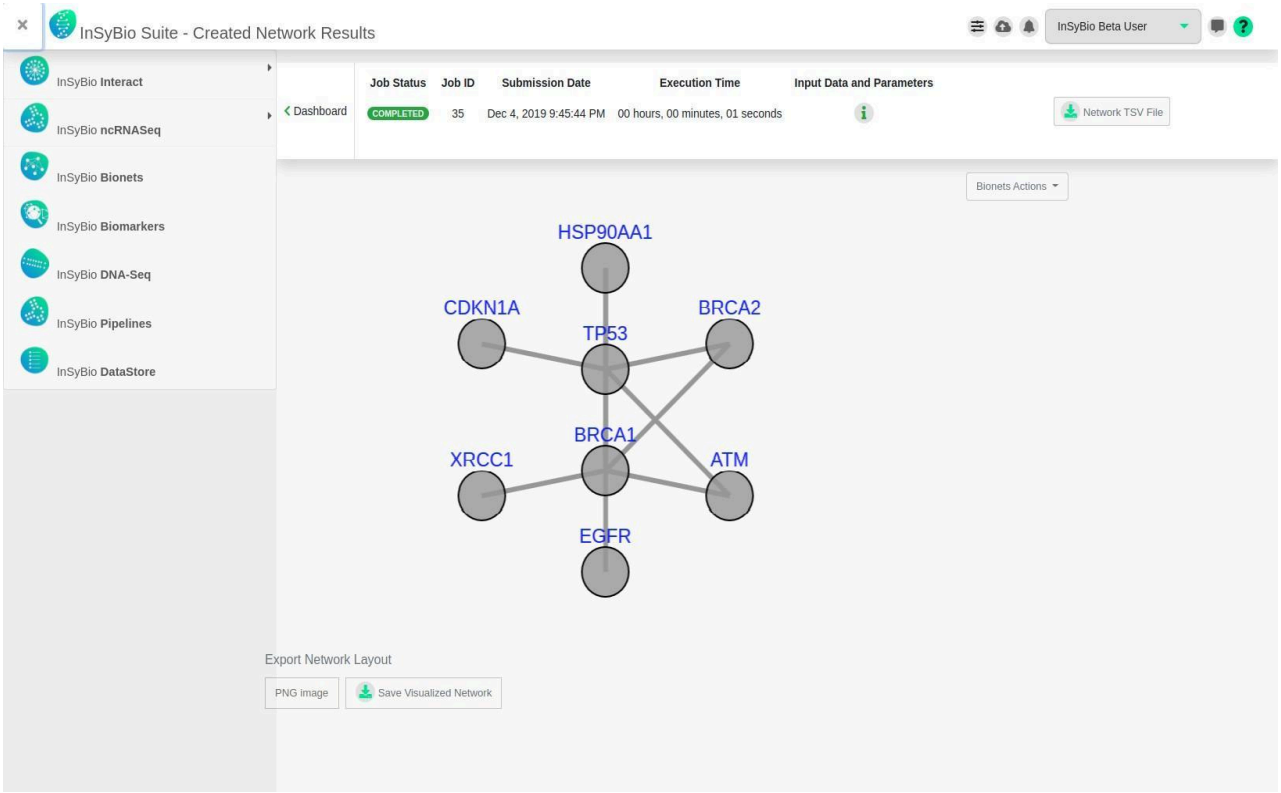
Network Creation

You can create and view biological networks from your set of biomarkers based on the protein-protein interactions knowledge base of InSyBio Interact.

In the tab New Network Creation you can either Type or Copy-Paste the biomarkers or Upload a file with biomarkers. You can use any known identifier for denoting your biomarkers: UniProt IDs, gene symbols, RefSeq_id and so on. Mixed identifiers are not supported! Note that each symbol should be in one line or separated by comma (without extra spaces or symbols between the commas).



After the network creation, the job will appear in the Created Networks tab. By clicking View Results a new page is opened.



Networks' visualization 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).

You can download the network in a TSV format and perform a list of supported actions in InSyBio BioNets tool.

Cloud computing Infrastructurer and Security Certifications

InSyBio Suite and all its tools are running over the cloud computing as a service infrastructure of Vultr (https://www.vultr.com), at the Amsterdam (Netherlands) facilities, offering the following security attestations and certifications (SOC 2+ (HIPAA), PCI (Merchant), CSA Star Level 1, ISO/IEC 20000-1:2018, ISO/IEC 27001:2022, ISO/IEC 27017:2015, ISO/IEC 27018:2019).

How to get InSyBio Interact

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

To purchase InSyBio Interact commercial version 3.3 please contact us at sales@insybio.com.

About Us

InSyBio Inc 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