

Analyze Protein-Protein Interactions with InSyBio Interact

March 2022

Insybio Suite v3.0



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 score for each protein pair. This confidence score indicates the probability of these two proteins to interact and the strength of this interaction. Users are able to tune this confidence score and extract their own datasets as well as obtain only positives 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 which 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 Ltd, 2017.

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

Search positive/negative PPIs

You can search for positive and negative Human Protein-Protein Interaction Data by specifying a threshold referring to the PPI confidence score.

The screenshot displays the InSyBio Suite Beta - Positive/Negative PPI Search interface. At the top, there is a search bar and a navigation menu. The main area contains search filters and a table of results.

Search Filters:

- Classification threshold value: 0.075
- Define positive/negative results: Positive Negative
- Feature normalization to the interval [0-1]: Yes No
- Export in CSV button
- 211367 Records
- Show results button

Table of Results:

Uniprot Id 1	Uniprot Id 2	Interaction Score	GO Function	GO Component	GO Process	Sequence Similarity	Co-localization
<input type="text" value="Search Uniprot Id 1"/>	<input type="text" value="Search Uniprot Id 2"/>	<input type="text" value="Search Interaction Score"/>	<input type="text" value="Search GO Function"/>	<input type="text" value="Search GO Component"/>	<input type="text" value="Search GO Process"/>	<input type="text" value="Search Sequence Similarity"/>	<input type="text" value="Search Co-localization"/>
P25788	P28870	1.2993	0.275862068965517	0.0769230769230769	0.15625	0.576036866359447	Co-localized on the same cellular compartment
Q9NST8	P29317	1.2992	0	0.0769230769230769	0	2.30414746543779e-59	Probably localized on the same or neighbor cellular compartment
Q8N4N8	O60674	1.2989	0.0114942528735632	0.0769230769230769	0.0625	0.576036866359447	Co-localized on the same cellular compartment
Q5JTC6	O15169	1.2959	0.0114942528735632	0.153846153846154	0.09375	0.576036866359447	Co-localized on the same cellular compartment
P19447	Q8NB84	1.2956	0.0114942528735632	0.0769230769230769	0.03125	0.576036866359447	Co-localized on the same cellular compartment
P06241	P09619	1.2921	0	0.230769230769231	0.03125	3.45622119815668e-35	Not localized on the same cellular compartment

Navigation: First Previous 1 2 3 4 5 ... 4228 Next Last Show 50 entries Showing 1 to 50 of 211,367 entries

You can search by choosing:

- The classification threshold value
- Positive or negative PPIs
- 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.

For every protein pair 22 informative features are calculated and provided including sequential, structural and functional features:

- **Gene Ontology (Co-function, Co-process, Co-localization)** (3 features). Gene Ontology database contains information of three types that include:
 - Molecular function of a gene product,
 - Biological process in which the gene product participates,
 - Cellular component where the gene product actsIn this features the number of common GO terms existing in a protein pair are calculated.
- **Sequence Similarity** (1 feature). This feature was obtained using the NCBI-BLAST+ standalone executable BLASTP to perform the human to human protein alignment. All BLASTP hits obtained with the default parameters and the sequence alignment E-values of each protein-protein pair have been used to fill in the specific feature.
- **Homology Based PPI** (1 feature). For each human protein pair we have found its homology pairs in yeast by using the 0.001 cutoff on the E-value taken by similarity tests that were conducted using BLASTP. If at least one of the resulting homology pairs is referred as interacting in the Yeast organism's DIP dataset, a value equal to "1" is assigned to this feature, otherwise "0".
- **Gene Expression Profile Similarity** (15 features). Fifteen gene expression datasets were used to estimate 15 gene expression profile similarity features. The datasets selected are extracted from NCBI Gene Expression Omnibus (GDS531, GDS534, GDS596, GDS651, GDS806, GDS807, GDS843, GDS987, GDS1085, GDS2855, GDS1402, GDS181, GDS1088, GDS841, GDS3257). For each dataset, the expression profile of each protein of every protein-pair was extracted and the Pearson correlation between the two profiles was estimated.

- **Co-localization** (1 feature). PLST tool (Scott et al, 2004) was used to predict the local cellular compartments in which each protein is likely to function. For every pair of proteins, a feature is calculated by taking the value of 1 if the two proteins share at least one local compartment and the value of 0 if they do not have any common local cellular compartment.
- **Domain-Domain Interactions** (1 feature). All possible domain pairs (downloaded from InterPro) were evaluated using the training set and the hypergeometric distribution to locate pairs that are interacting with high probability (p-values less than 0.05). Then for every protein pair, their domain combinations are computed and the number of the interacting ones is measured and used as the feature's value.

Retrieve information about a specific protein

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

You can:

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

Uniprot Id 1	Uniprot Id 2	Interaction Score	GO Function	GO Component	GO Process	Sequence Similarity	Co-localization	Homology Yeast	Domain Domain Interaction	HPRD	APID_verified	Co-expressions
Q9P212	Q15124	0.2873	0	0	0.03125	0.576036866359447	Not localized on the same cellurall compartment	No	0.000133246913580247	No	0	
Q9P212	Q5NGF3	-0.031725	6.3416091954023e-06	0.00155646153846154	0.0004603125	0.576036866359447	Not localized on the same cellurall compartment	No	3.4566649382716e-05	No	0	
Q9P212	Q6NSJ8	0.017539	0	0	0	0.576036866359447	Not localized on the same cellurall compartment	No	0.000191006172839586	No	0	
Q9P212	Q76KD6	0.15875	0	0	0	0.576036866359447	Not localized on the same cellurall compartment	No	0.000229654320987654	No	0	

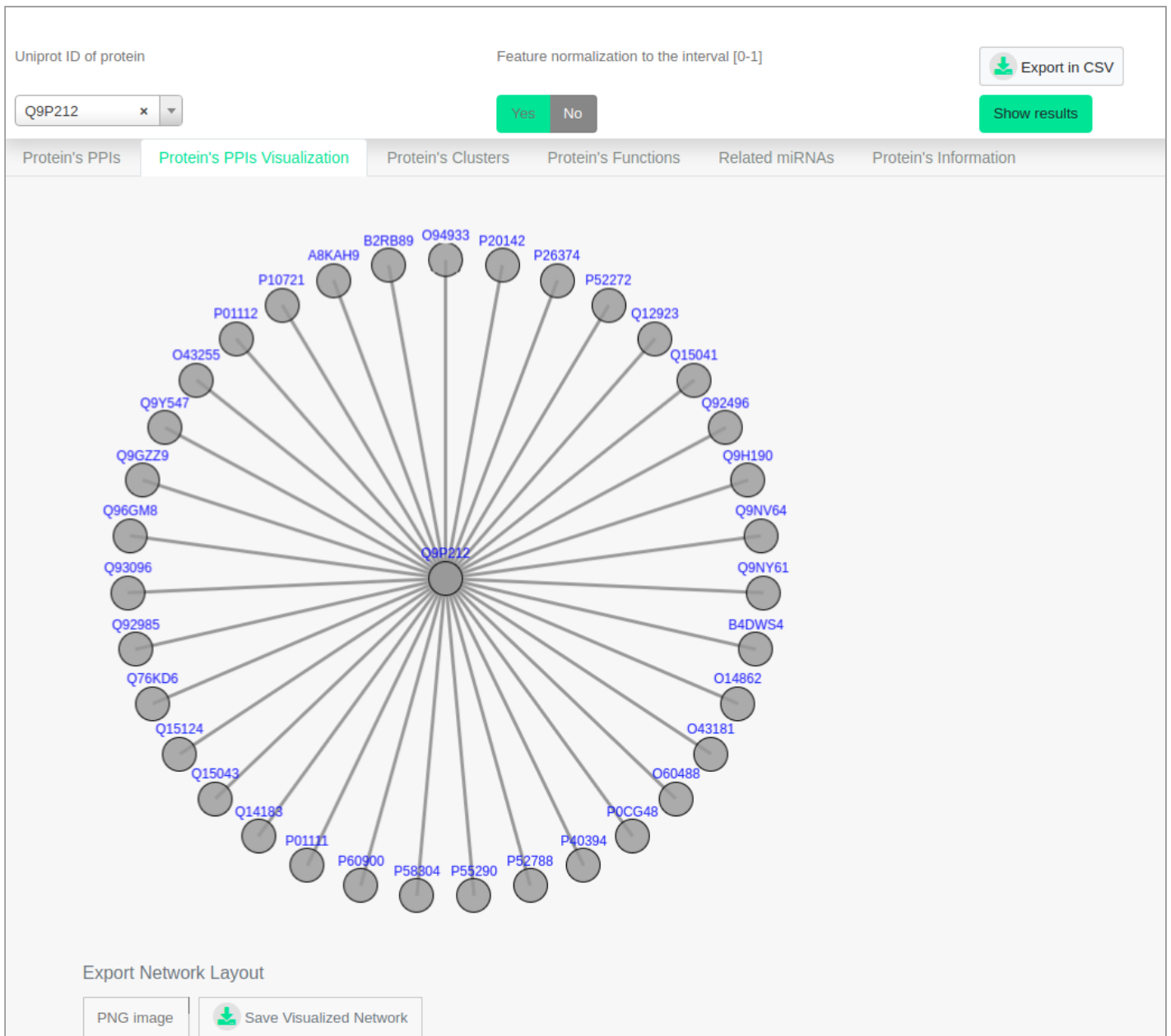
The results are presented on your screen in a browse-able table or you can download them as a TAB delimited CSV file.

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 screenshot shows the InSyBio Suite Beta - Protein Search interface. The search bar contains 'Q9P212'. The 'Feature normalization to the interval [0-1]' dropdown is set to 'No'. The 'Export in CSV' button is visible. The 'Protein's Clusters' tab is active, displaying a table with 'Cluster's Proteins' (B4DWS4, Q9P212) and 'Cluster's Functions' (GO Term, Function Name, Function Namespace).

GO Term	Function Name	Function Namespace
0008513	secondary active organic cation transmembrane transporter activity	molecular_function
0015226	carnitine transmembrane transporter activity	molecular_function
0015293	symporter activity	molecular_function
0015491	cation:cation antiporter activity	molecular_function
0015651	quaternary ammonium group transmembrane transporter activity	molecular_function
0030165	PDZ domain binding	molecular_function

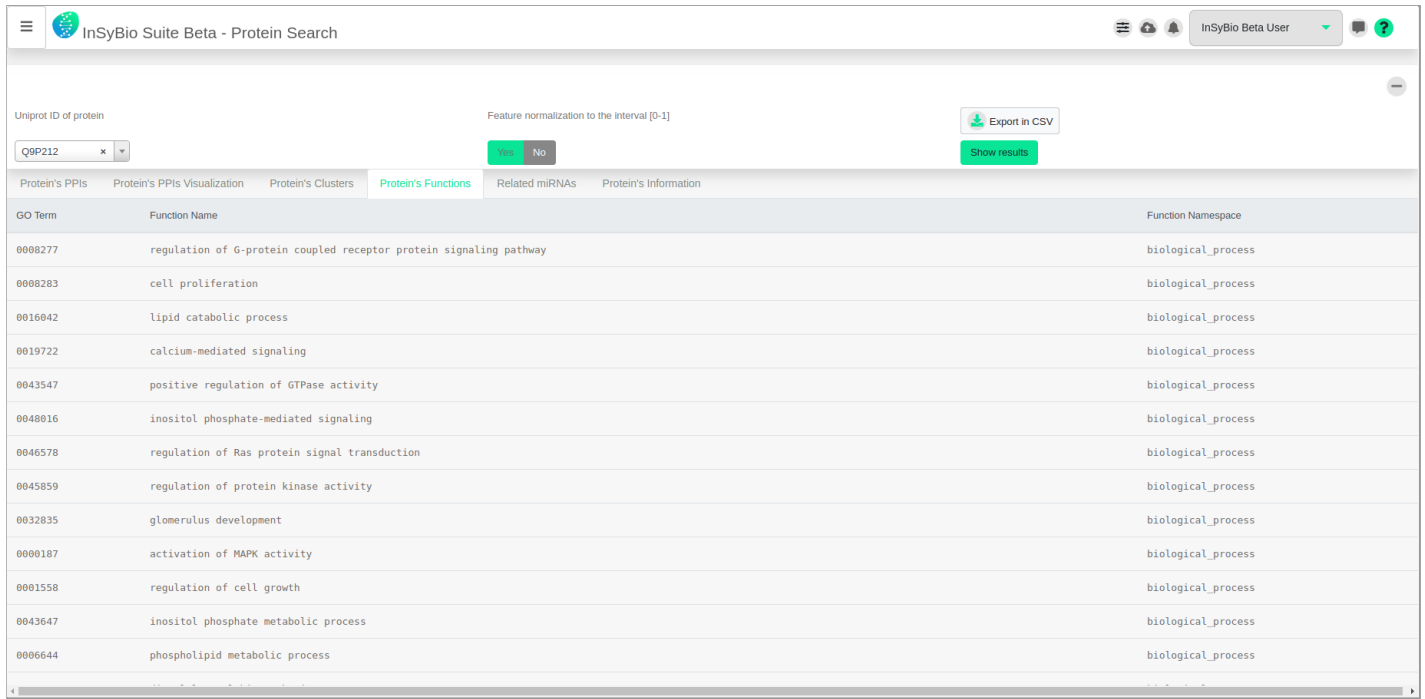
The result is a list of the clusters the protein of interest takes part in.

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;



GO Term	Function Name	Function Namespace
0008277	regulation of G-protein coupled receptor protein signaling pathway	biological_process
0008283	cell proliferation	biological_process
0016042	lipid catabolic process	biological_process
0019722	calcium-mediated signaling	biological_process
0043547	positive regulation of GTPase activity	biological_process
0048016	inositol phosphate-mediated signaling	biological_process
0046578	regulation of Ras protein signal transduction	biological_process
0045859	regulation of protein kinase activity	biological_process
0032835	glomerulus development	biological_process
0000187	activation of MAPK activity	biological_process
0001558	regulation of cell growth	biological_process
0043647	inositol phosphate metabolic process	biological_process
0006644	phospholipid metabolic process	biological_process

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 screenshot shows the InSyBio Suite Beta - Protein Search interface. At the top, there is a search bar with the text 'Uniprot ID of protein' and a dropdown menu containing 'Q9P212'. To the right of the search bar, there is a 'Feature normalization to the interval [0-1]' section with a 'No' button. Further right, there are two buttons: 'Export in CSV' and 'Show results'. Below the search bar, there is a navigation menu with tabs: 'Protein's PPIs', 'Protein's PPIs Visualization', 'Protein's Clusters', 'Protein's Functions', 'Related miRNAs' (which is highlighted), and 'Protein's Information'. The main content area below the navigation menu displays the message: 'There aren't any miRNAs related to this 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 screenshot shows the InSyBio Suite Beta - Protein Search interface. The search bar contains the Uniprot ID 'B2R789'. The 'Feature normalization to the interval [0-1]' is set to 'No'. The 'Export in CSV' and 'Show results' buttons are visible. The 'Protein's Information' tab is selected, displaying a table with the following data:

Uniprot ID	Official Gene Symbol	Protein Name	Sequence
B2R789	STAR13	START domain containing 13, isoform CRA_b	MKLDVNFQRKGGDSDEEDLCLISNKKWTFQRTSRRWSRVDDLYLLPRGDRNGSPGGTGMNRTTSSSESVLTDLSEPEVCSIHSESSGGSDRSRQPGGCT DNPVMDAPLVSSLPQPRDVLNHPFHKNEKTRARAKSFLKRMETLLGKGAHRHKGSGRTGGLVIGPMLQQEPESFKAMQCIQIPNGDLQNSPP PACRKLPCSGKSSGESSPSEHSSGVSTPCLKERKCHEANKRGMYLEDLVLAGTALPDAGDQSRMHEFHSQENLVVHIKDKHGTFPKALSIKESL SPTDSSNGVWRGTGSLGREQVPGAREPLMASCHRASRVSYDMVPGSHLYASTGDLLEKDDLFPHLDDILOHVNGLQEVDDNSKDVPELOTH DTLVGEPGLSTFPNQITLDFEGNSVSEGRTPPSDVERDVTSLNESEPPGVDRDRDSGVGASLTPNRRLRWNSFQLSHQRPAPASPHISSQATASQL SLLQRFSLRLTAIMEKHSMSNKHGWTWSVPKFMKRMKVDPYKDKAVFGVPLIVHVQRTGGPLPQIQQALRYLRSNCLDQVGLFRKSGVKSRIHALRO MNEFPEWNYEDQAYOVADWVKQFFRDLPEPLFTNKLSETFLLHYQVYSKEQRQVQAAITLLADENREVLOTLLCFLDVNVNLEENQMPMHLA VCLAPSLFRLNLLKKESSPRVIOKRYATGKPDQKDLNENLAAGDLAHPHMECORLFEVPHELVADSRNSVYEAETHYPTLLEELGTQLEESGATFHTYL NHLTQGLQKEAKENKFKQWTCSSDITDILAFKXVGDGDLPLKLMASVEVEAPPVLLRVLRENHLKQEDFVQWVETLDROTEIYQVLLNSHAPPS RDFVLRWKTDLPKMCTLVSLVHEEAQLLGGVRAVWMSOYLIEPCGSGKSLRTHICRIDLKGHPWYSKGFGLCAAEVARIRNSFQPLAEG PETKI

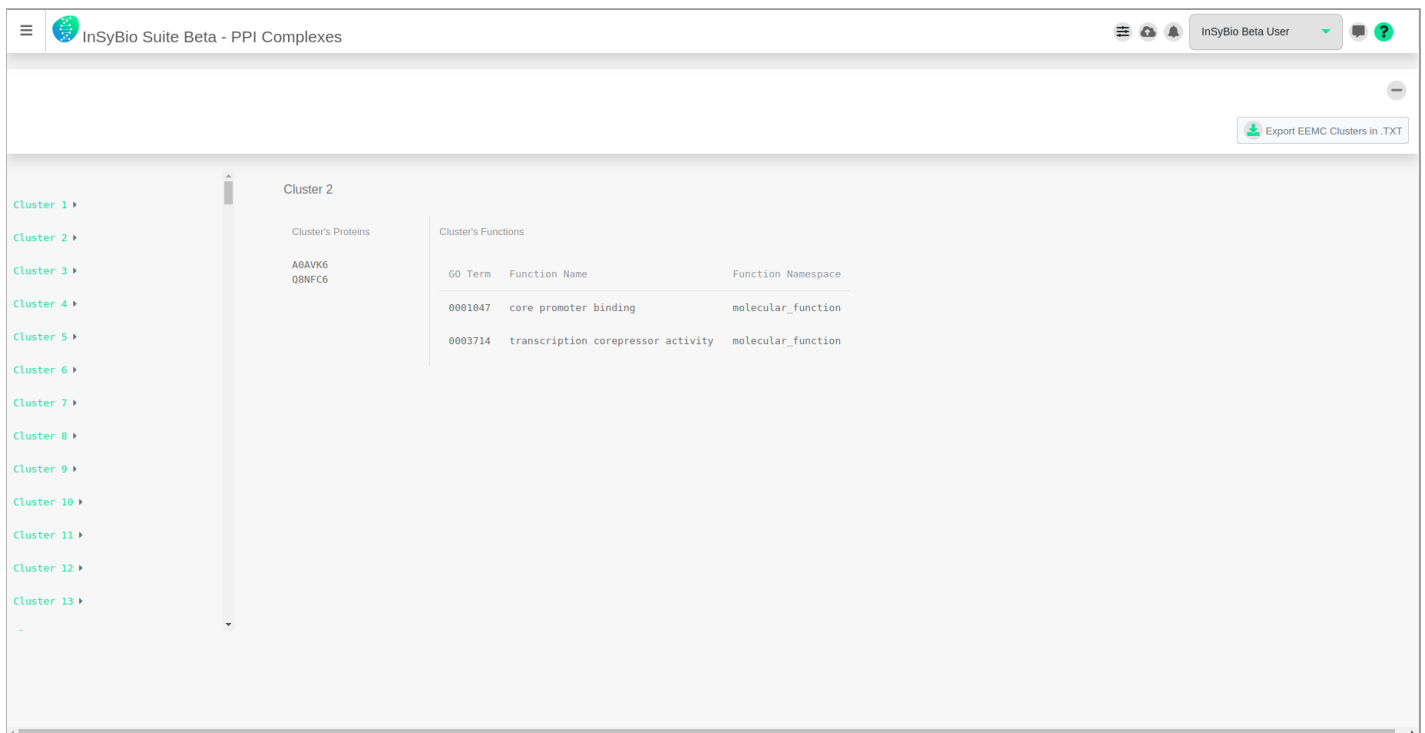
The result is a table containing information about the protein of interest.

For each protein, the following information is presented:

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

PPI complexes

You can view a list of all the clusters computed from the algorithm EEMC. EEMC (Evolutionary Enhanced Markov Clustering) is a novel fully unsupervised methodology that InSyBio team has proposed for the prediction of protein complexes from weighted PPI graphs. It is a hybrid combination of an adaptive evolutionary algorithm and a state-of-the-art clustering algorithm named Enhanced Markov Clustering. When applied to new human datasets EEMC's performance was encouraging in the prediction of protein complexes which consist of proteins with high functional similarity. From EEMC's analysis, we suggest new potentially true Human protein complexes which should be further validated using experimental techniques.



The screenshot shows the InSyBio Suite Beta interface for PPI Complexes. The top navigation bar includes the InSyBio logo, the title 'InSyBio Suite Beta - PPI Complexes', and user information 'InSyBio Beta User'. A button 'Export EEMC Clusters in .TXT' is visible in the top right. The main content area is divided into a sidebar on the left with a list of clusters (Cluster 1 to Cluster 13) and a main panel for 'Cluster 2'. The main panel is split into two sections: 'Cluster's Proteins' and 'Cluster's Functions'. The 'Cluster's Proteins' section lists A8AVK6 and Q8NFC6. The 'Cluster's Functions' section contains a table with the following data:

GO Term	Function Name	Function Namespace
0001047	core promoter binding	molecular_function
0003714	transcription corepressor activity	molecular_function

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 Q9COH2,0005254
2,Q8NFC6 A0AVK6,0001047 0003714
3,A0AVT1 Q7Z6D5,0019780 0016874
...
```

PPI functions

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 screenshot displays the InSyBio Suite Beta - PPI Functions interface. The 'GO Term (7-digit code)' is set to 0000987. The 'Complexes' tab is selected, showing the following details:

- GO term:** 0000987
- Function Name:** core promoter proximal region sequence-specific DNA binding
- Function Namespace:** molecular_function

The 'Complexes List' shows the following clusters:

- Cluster 468
- Cluster 721
- Cluster 1116
- Cluster 7687

The 'Cluster's Proteins' section lists the proteins F8MAJ5 and 096C01. The 'Cluster's Functions' section displays a table of related GO terms and their namespaces:

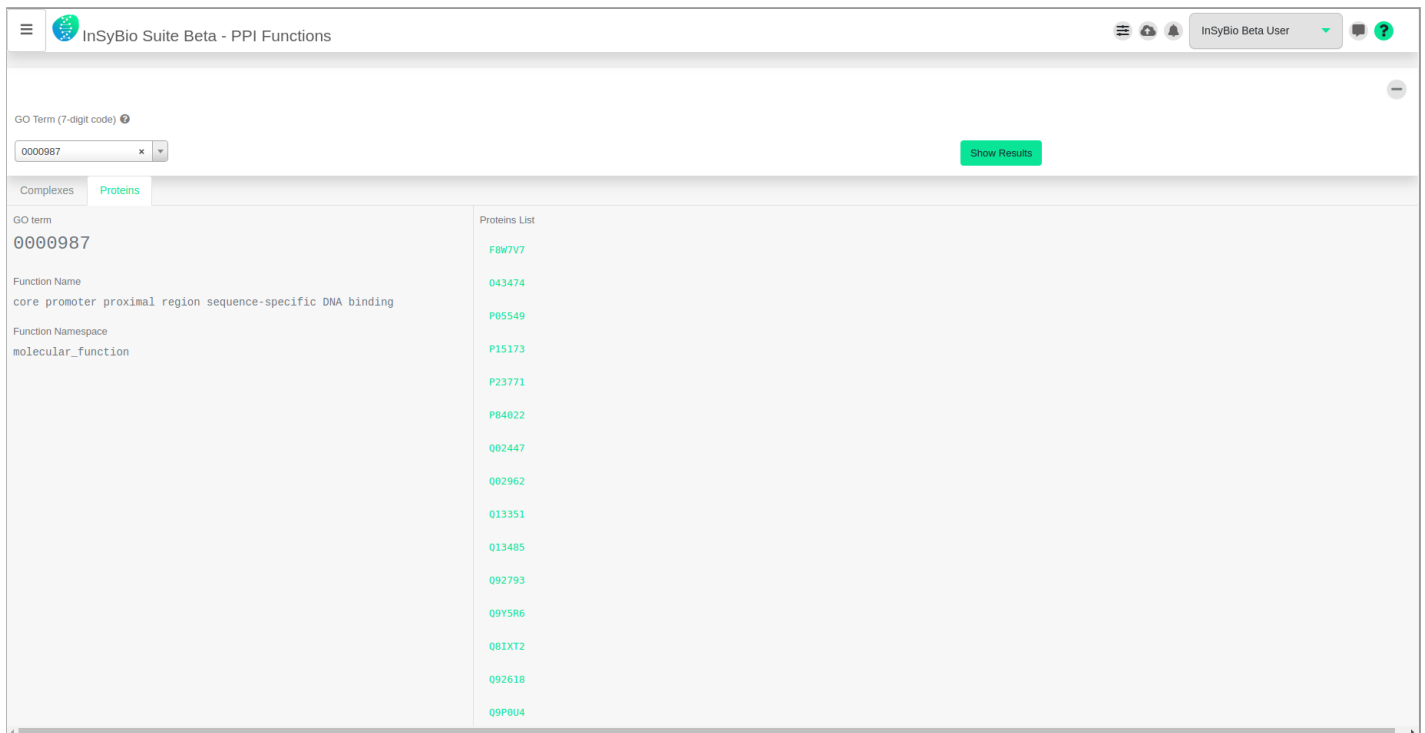
GO Term	Function Name	Function Namespace
0000977	RNA polymerase II regulatory region sequence-specific DNA binding	molecular_function
0000987	core promoter proximal region sequence-specific DNA binding	molecular_function
0001846	core promoter sequence-specific DNA binding	molecular_function
0001077	transcriptional activator activity, RNA polymerase II core promoter proximal region sequence-specific binding	molecular_function
0003705	transcription factor activity, RNA polymerase II distal enhancer sequence-specific binding	molecular_function
0042826	histone deacetylase binding	molecular_function
0046983	protein dimerization activity	molecular_function
0071837	HMG box domain binding	molecular_function

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 screenshot displays the InSyBio Suite Beta - PPI Functions interface. At the top, there is a search bar for the GO Term (7-digit code) with the value 0000987 entered. A green 'Show Results' button is located to the right of the search bar. Below the search bar, there are two tabs: 'Complexes' and 'Proteins', with 'Proteins' selected. The main content area is divided into two columns. The left column shows the GO term 0000987, its function name 'core promoter proximal region sequence-specific DNA binding', and its function namespace 'molecular_function'. The right column, titled 'Proteins List', contains a vertical list of protein IDs: FBW7V7, 043474, P05549, P15173, P23771, P84022, Q02447, Q02962, Q13351, Q13485, Q92793, Q9Y5R6, Q81XT2, Q92618, and Q9P0U4.

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.

PPI datasets

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

InSyBio Suite Beta - PPI Dataset Creation Tool

Number of positive interactions: All positive PPIs in HPRD:
 Number of negative interactions: Feature normalization to the interval [0-1]:

Actions:

Uniprot Id 1	Uniprot Id 2	Interaction Score	GO Function	GO Component	GO Process	Sequence Similarity	Co-localization	Homology Yeast	Domain Domain Interaction	HPRD	APID_verified	Co-expressions		
P38858	Q960L0	-0.69411	0.0045977	0.092388	0.0375	Not similar	Probable localized on the same or neighbor celllural compartment	No	0.029164	No	0	⊖		
GDS531	GDS534	GDS596	GDS651	GDS806	GDS807	GDS843	GDS987	GDS1885	GDS2855	GDS1482	GDS181	GDS1888	GDS841	GDS3257
0.80245	0.85146	0.56638	0.67886	0.31599	0.27074	0.06652	0.38644	0.26068	0.35597	0.43682	0.88319	-0.7132	0.58059	-0.23966
P62277	P63892	-0.61444	0	0	0	Not similar	Not localized on the same celllural compartment	No	0.028885	No	0	⊕		
P8CG48	Q5JWF2	-0.59187	0	0	1	Not similar	Not localized on the same celllural compartment	No	0.044216	No	0	⊕		
D1MER9	E7E566	-0.58602	0.0079141	0.041426	0.0038688	7e-17	Not localized on the same celllural compartment	No	0.012593	No	0	⊕		

Showing 1 to 50 of 200 entries

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 through the Actions Button. The dataset created can also be saved in InSyBio Data Store for future use.

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.

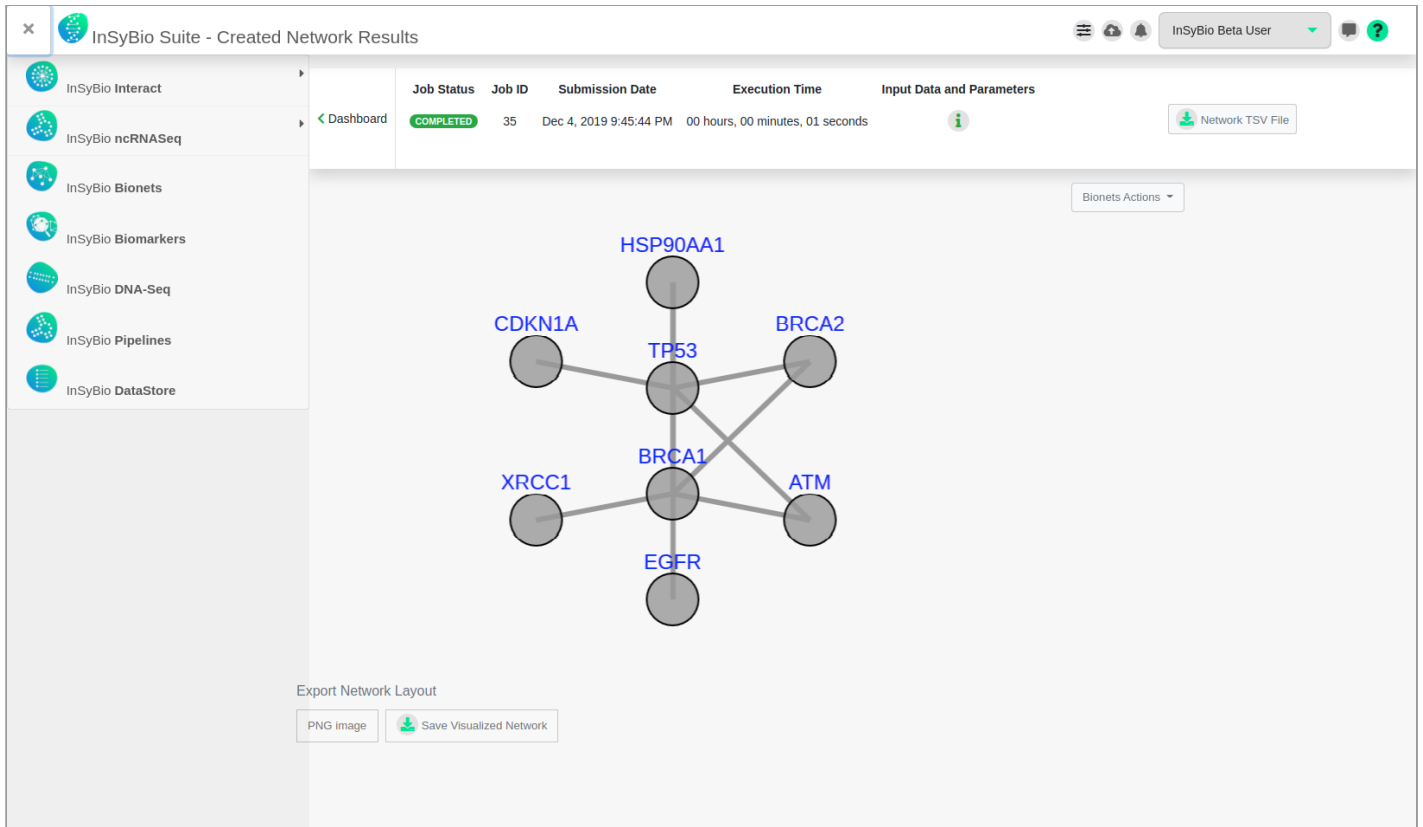
Biological Networks Creation

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.

The screenshot shows the 'Create Network' interface in the InSyBio Suite Beta. The page title is 'InSyBio Suite Beta - Network Creation'. The main heading is 'Create Network', with a sub-heading: 'Creation of biological networks from your set of biomarkers based on the protein-protein interactions knowledge base of InSyBio Interact.' Below this, there are tabs for 'Created Networks', 'New Network Creation' (which is active), 'Running and Pending Processes', and 'Errors'. A sub-heading reads 'Create a new network from a list of biomarkers'. Underneath, there is an 'Input mode' section with two radio buttons: 'Type or Copy-Paste biomarkers' (selected) and 'Upload a file with biomarkers'. The form includes a 'New network's title:' field with a text input containing 'Title...'. Below that is a 'List of symbols' field with a text area containing 'ATM,TP53,BRCA1...'. At the bottom left, there is a green button labeled 'Start Network Creation'.

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.

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:

InSyBio Suite Beta - Enrichment Analysis

Perform Enrichment Analysis

Performed Enrichment Analyses **New Enrichment Analysis** Running and Pending Enrichment Analyses Errors at Enrichment Analyses

Perform Enrichment Analysis

GO term Enrichment Analysis from your set of biomarkers based on the protein-go term correlation knowledge base of InSyBio Interact.

Input mode

Type or Copy-Paste biomarkers Upload a file with biomarkers Upload custom input files

Use any known identifier for denoting your biomarkers: Uniprot IDs, gene symbols, RefSeq_id and so on. **Mixed identifiers are not supported!**

Pvalue threshold ? : 0.05

New enrichment analysis title : Title...

List of symbols ? : ATM,TP53,BRCA1,...

Start Enrichment Analysis

- Either Type or Copy-Paste biomarkers or

InSyBio Suite - Enrichment Analysis

Perform Enrichment Analysis

Performed Enrichment Analyses **New Enrichment Analysis** Running and Pending Enrichment Analyses Errors at Enrichment Analyses

Perform Enrichment Analysis

GO term Enrichment Analysis from your set of biomarkers based on the protein-go term correlation knowledge base of InSyBio Interact.

Input mode

Type or Copy-Paste biomarkers Upload a file with biomarkers Upload custom input files

Use any known identifier for denoting your biomarkers: Uniprot IDs, gene symbols, RefSeq_id and so on. **Mixed identifiers are not supported!**

Pvalue threshold ? :

New enrichment analysis title :

Select File Biomarker Input ? :

Start Enrichment Analysis

- Upload a file with biomarkers or

InSyBio Suite - Enrichment Analysis

Perform Enrichment Analysis

Performed Enrichment Analyses **New Enrichment Analysis** Running and Pending Enrichment Analyses Errors at Enrichment Analyses

Perform Enrichment Analysis

GO term Enrichment Analysis from your set of biomarkers based on the protein-go term correlation knowledge base of InSyBio Interact.

Input mode

Type or Copy-Paste biomarkers Upload a file with biomarkers Upload custom input files

Use any known identifier for denoting your biomarkers: Uniprot IDs, gene symbols, RefSeq_id and so on. **Mixed identifiers are not supported!**

Pvalue threshold ? : 0.05

New enrichment analysis title : Title...

Select File Biomarker Input ? : Drop file here to upload or Click to select file

Custom Enrichment Files

Select file (Go Terms) ? : Drop file here to upload or Click to select file

Select file (Annotations) ? : Drop file here to upload or Click to select file

Select file (Go Terms Full) ? : Drop file here to upload or Click to select file

Select file (Go Types) ? : Drop file here to upload or Click to select file

Select file (Goa Pruned) ? : Drop file here to upload or Click to select file

Start Enrichment Analysis

- 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.

InSyBio Suite - Enrichment Analysis

Perform Enrichment Analysis

Performed Enrichment Analyses | New Enrichment Analysis | Running and Pending Enrichment Analyses | Errors at Enrichment Analyses

List of Performed Enrichment Analyses

ID	Title	Biomarkers	Symbols DB	Submitted	Start Processing	End Processing	Action
28	test_symbols_enrichment	ATM, TP53, BRCA1, ...	GENESYMBOLS	11/13/19 12:09 PM	11/13/19 12:10 PM	11/13/19 12:10 PM	View Results
25	enrichment_test	ATM, TP53, BRCA1...	GENESYMBOLS	10/11/19 1:32 PM	10/11/19 1:32 PM	10/11/19 1:32 PM	View Results
22	1_test	ATM, TP53...	GENESYMBOLS	10/4/19 2:23 PM	10/8/19 6:54 AM	10/8/19 6:54 AM	View Results
21	1_test	ATM, TP53...	GENESYMBOLS	10/4/19 2:23 PM	10/8/19 6:54 AM	10/8/19 6:54 AM	View Results

Previous 1 Next

Show 10 entries

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.

InSyBio Suite - Enrichment Analysis Results

Job Status: **COMPLETED** | Job ID: 28 | Submission Date: Nov 13, 2019 12:09:28 PM | Execution Time: 00 hours, 00 minutes, 01 seconds | Input Data and Parameters

[Export Results](#)

GO Term	GO Term's Type	GO Term's Name	GO Term's Specificity	Enrichment Score	Associated Uniprot ids	Associated Input Ids
GO:000122	biological:process	negative regulation of transcription from RNA polymerase II promoter		9	0.0352905428365	P04637 TP53
GO:000729	biological:process	DNA double-strand break processing		8	0.00584764666219	P38398, Q13315 BRCA1, ATM
GO:000731	biological:process	DNA synthesis involved in DNA repair		7	0.030471091652	P38398, Q13315 BRCA1, ATM
GO:000732	biological:process	strand displacement		6	0.0188808869879	P38398, Q13315 BRCA1, ATM
GO:0002331	biological:process	pre-B cell allelic exclusion		13	0.0249351691907	Q13315 ATM
GO:0003677	molecular:function	DNA binding		5	0.0291025373332	P04637, P38398, Q13315 TP53, BRCA1, ATM
GO:0003700	molecular:function	transcription factor activity, sequence-specific DNA binding		3	0.0284145684002	P04637 TP53
GO:0004677	molecular:function	DNA-dependent protein kinase activity		8	0.032967327857	Q13315 ATM
GO:0005515	molecular:function	protein binding		3	3.47534911816e-18	P04637, P38398, Q13315 TP53, BRCA1, ATM
GO:0005524	molecular:function	ATP binding		6	0.00218574886077	P04637, Q13315 TP53, ATM

First Previous 1 2 3 4 5 6 Next Last

Show 10 entries

Showing 1 to 10 of 55 entries

Functional Annotation

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

InSyBio Suite - Functional Annotation
InSyBio Beta User

Uniprot ID of protein Show Results

	Specificity				
Molecular Function	<ul style="list-style-type: none"> enzyme activator activity? SH3 domain binding? SH3/SH2 adaptor activity? spectrin binding? MHC class II protein complex binding? phosphoprotein binding? enzyme binding? potassium channel regulator activity? histone deacetylase binding? catalytic activity? signaling adaptor activity? ion channel binding? SH2 domain binding? DNA binding? transcription corepressor activity? GKAP/Homer scaffold activity? actin binding? protein homodimerization activity? protein heterodimerization activity? protein domain specific binding? peptide hormone binding? neuropeptide binding? metal ion binding? protein binding? 	<ul style="list-style-type: none"> tryptophan 5-monoxygenase activity? amino acid binding? protein O-GlcNAc transferase activity? acetylglucosaminyltransferase activity? cysteine-type peptidase activity? ubiquitin protein ligase binding? phosphoserine binding? actin filament binding? GTP binding? ATP binding? protein kinase binding? histone deacetylase activity? G-protein coupled glutamate receptor binding? galanin receptor activity? 3-beta-hydroxy-delta5-steroid dehydrogenase activity? cholest-5-ene-3-beta,7-alpha-diol 3-beta-dehydrogenase activity? 	<ul style="list-style-type: none"> iron ion binding? protein N-acetylglucosaminyltransferase activity? phosphatidylinositol-3,4,5-trisphosphate binding? aminopeptidase activity? carboxypeptidase activity? cysteine-type endopeptidase activity? myosin tail binding? phosphatidylinositol-5-phosphate binding? phosphatidylinositol-3-phosphate binding? phosphatidylinositol-4-phosphate binding? sphingomyelin phosphodiesterase activity? 	<p>No Data</p>	<ul style="list-style-type: none"> histone acetyltransferase activity (H4-K16 specific)? histone acetyltransferase activity (H4-K8 specific)? histone acetyltransferase activity (H4-K5 specific)?
Biological Process	<ul style="list-style-type: none"> response to activity? response to nutrient levels? oxidation-reduction process? circadian rhythm? positive regulation of catalytic activity? signal transduction? response to nutrient? response to drug? response to toxic substance? proteolysis? protein complex assembly? regulation of membrane repolarization? membrane organization? viral process? intracellular signal transduction? immune response? adaptive immune response? behavioral response to cocaine? chemical synaptic transmission? cell surface receptor signaling pathway? feeding behavior? 	<ul style="list-style-type: none"> response to calcium ion? response to estrogen? regulation of insulin receptor signaling pathway? cellular response to retinoic acid? circadian regulation of gene expression? response to insulin? apoptotic process? inner ear morphogenesis? regulation of microvillus length? cellular protein complex assembly? hippocampus development? hippo signaling? regulation of heart rate by cardiac conduction? regulation of cellular response to heat? skeletal muscle contraction? positive regulation of signal transduction? B cell activation? immunoglobulin secretion? transcription, DNA-templated? regulation of G-protein coupled receptor protein signaling pathway? chemical homeostasis within a tissue? learning or memory? muscle contraction? glycosphingolipid metabolic process? 	<ul style="list-style-type: none"> response to glucocorticoid? cellular response to lithium ion? indolalkylamine biosynthetic process? aromatic amino acid family metabolic process? serotonin biosynthetic process? regulation of gluconeogenesis involved in cellular glucose homeostasis? phosphatidylinositol-mediated signaling? regulation of glycolytic process? sensory perception of light stimulus? equilibrioception? photoreceptor cell maintenance? regulation of potassium ion transmembrane transporter activity? G2/M transition of mitotic cell cycle? regulation of heart rate by hormone? protein targeting? positive regulation of abscisic acid-activated signaling pathway? regulation of cation channel activity? regulation of calcium ion import? positive regulation of calcium ion transport? regulation of transcription, DNA-templated? G-protein coupled glutamate receptor signaling pathway? phosphatidylinositol metabolic process? neuropeptide signaling pathway? phospholipase C-activating G-protein coupled 	<ul style="list-style-type: none"> positive regulation of proteolysis? regulation of Rac protein signal transduction? negative regulation of protein ubiquitination? positive regulation of granulocyte differentiation? protein O-linked glycosylation? homocysteine catabolic process? membrane repolarization during cardiac muscle cell action potential? positive regulation of protein insertion into mitochondrial membrane involved in apoptotic signaling pathway? negative regulation of cysteine-type endopeptidase activity involved in apoptotic process? neuron migration? negative regulation of transcription, DNA-templated? regulation of store-operated calcium entry? phospholipase C-activating G-protein coupled glutamate receptor signaling pathway? skeletal muscle fiber development? negative regulation of T cell activation? antigen receptor-mediated signaling pathway? negative regulation of transcription from RNA polymerase II promoter? histone deacetylation? skeletal muscle cell differentiation? sensory perception of sound? calcium-mediated signaling using intracellular calcium source? negative regulation of adenylate cyclase activity? adenylate cyclase-activating G-protein coupled receptor signaling pathway? adenylate cyclase-modulating G-protein coupled receptor signaling pathway? 	<ul style="list-style-type: none"> histone H4-K16 acetylation? histone H4-K8 acetylation? histone H4-K5 acetylation? histone H3-K4 trimethylation? positive regulation of histone H3-K27 methylation? positive regulation of histone H3-K4 methylation? protein polyubiquitination? inner ear receptor stereocilium organization? cerebral cortex development? substantia nigra development? inactivation of MAPK activity? neuron projection development? positive regulation of cytosolic calcium ion concentration?

Confidence Score

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 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).

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 3.0 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

External Publication of InSyBio Ltd - Any InSyBio information that is to be used in advertising, press releases, or promotional materials requires prior written approval from the InSyBio Ltd. A draft of the proposed document should accompany any such request. InSyBio Ltd reserves the right to deny approval of external usage for any reason.

Copyright 2022 InSyBio Ltd. Reproduction without written permission is completely forbidden.