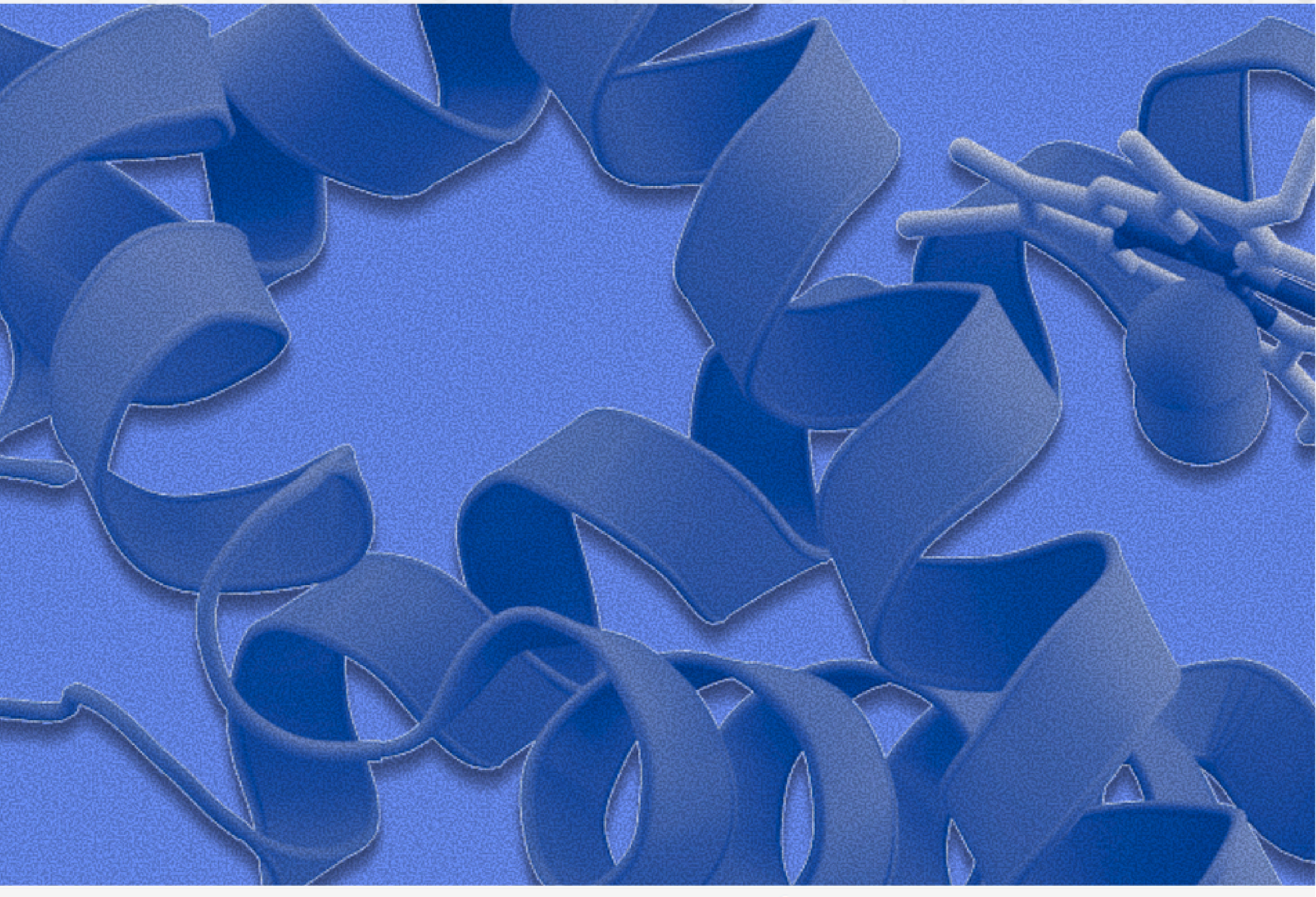
IPB University
— Bogor Indonesia —

iPrimeTarget

System User Guide:
Integrative Protein Network Analysis for Precision Medicine Research



BDI
BMHS Diklat Indonesia

RSUD
CIBINONG

IRSI RESEARCH CENTRE

User Manual – Website iPrimeTarget

Address: https://iprimetarget-1.onrender.com/

Description

iPrimeTarget is an integrated protein network analysis platform for precision medicine research. This application helps researchers:

1. **Retrieve disease-based gene/protein** lists or manual inputs.
2. **Automatic PPI data collection** via the **STRING database**.
3. **Community Detection & Centrality Analysis** (including the Greedy Modularity Disassembly method).
4. **Network Visualization & Enrichment Analysis** to generate reproducible insights.
5. **Perform Community Consideration Centrality (CCC)**.

This platform runs on modern browsers (Chrome, Firefox, Edge) without additional installation and uses data from UniProt and STRING as its biological information base.

1. Landing Page

On the landing page, users can see:

- **General information** about iPrimeTarget.
- **About Us** to see information about the application developer.
- There are two main buttons:
  - **User Manual**, which contains instructions on how to use iPrimeTarget.
  - **Start Analysis** to enter the protein analysis process.

Integrative Protein Network Analysis for Precision Medicine Research

Computational and network-based protein selection for identifying gene and protein interactions more efficiently and accurately.

2. Enter the Analysis Module

From the home page, click **Start Analysis**. You will be directed to the Protein Analysis page with the left side menu:

- Analysis Home / Overview

- Data Gen
- Protein–Protein Interaction Network
- Centrality Analysis
- Community Detection
- Enrichment Analysis
- Community Consideration Centrality

All result tabs will display the message “Analysis Required” until the first analysis is successfully run.

3. Entering Input

On the input form, users will be asked to enter the following parameters:

Category	Parameter	Description
Select Input Method	Users select the input method using the disease name or using the gene data list.	
	Disease Name	Disease keywords (e.g., diabetes mellitus). The system will search for proteins related to this disease.
	Manual Gene List	List of specific genes/proteins to be analyzed. Users must write one gene/protein per line.
Basic Parameters	Parameters required by the system to perform analysis according to user requirements.	
	Organism (NCBI Taxonomy ID)	ID Species according to NCBI Taxonomy ID. The default is 9606: Homo sapiens (human). If users want to analyze other species, they can input the species' NCBI Taxonomy ID. Examples: • Mus Musculus (mouse): 10090 • Arabidopsis thaliana: 3702 • Saccharomyces cerevisiae: 4932 Or can be searched at: https://www.ncbi.nlm.nih.gov/Taxonomy/Browser/wwwtax.cgi

Category	Parameter	Description
Basic Parameters	Disease Search Query Method	Method used to retrieve protein data from UniProt • Non CC_Disease: all proteins/genes associated with the input disease name will be analyzed • CC_Disease: only proteins/genes associated with disease names found in the UniProt database
	Community Detection Methods	Methods or algorithms used in the community detection stage: • DGM (Disassembly Greedy Modularity) • Louvain • Greedy Modularity • Label Propagation • Spinglass • Leiden
	Top N Community Consideration Centrality (CCC)	Measures how important a community is in connecting and influencing other communities in the network: • Top N CCC-10: The ten most influential communities based on the highest CCC values. • Top N CCC-20: The twenty most influential communities for a broader picture.
Advanced Parameters	Additional (optional) parameters for more detailed analysis customization.	
	STRING Database Parameters	Required Score: • 0.15 (Low confidence) • 0.40 (Medium confidence) • 0.70 (High confidence) • 0.90 (Highest confidence)
	Community Detection Parameters	Configuring the community detection algorithm in a network: • DGM P Parameter: This parameter controls the strength or sensitivity of the algorithm in

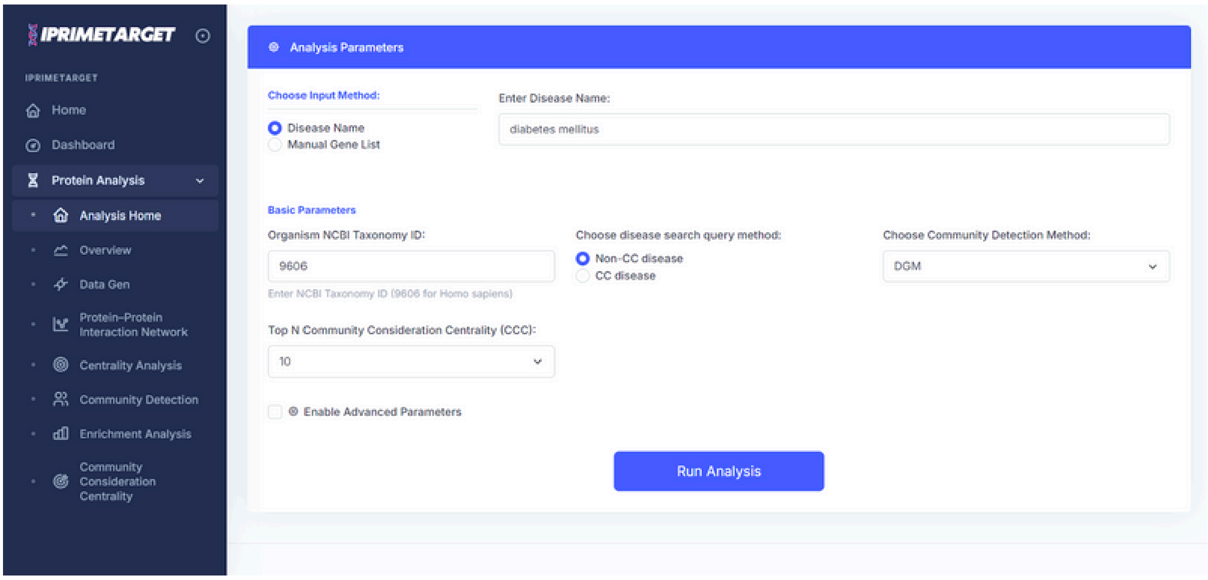
Category	Parameter	Description
Advance Parameters	Community Detection Parameters	detecting communities influences how tightly the resulting communities are clustered. • DGM Random Seed: Sets the random value used to start the community detection process, ensuring consistent results when the algorithm is run repeatedly.
	Clique Analysis Parameters	Analyzing cluster structures in clique-based networks: • Top N Cliques: Determining the number of top cliques (groups of interconnected nodes) to be analyzed based on their size or importance. • Approximation Algorithm: An algorithm used to calculate cliques in a network more efficiently, sacrificing a little accuracy for processing speed.
	CCC Analysis Parameters	Analyzing the influence of communities in networks based on CCC: • CCC Alpha Parameter (0.01 – 1.0): Controls the sensitivity level in CCC calculations, affecting how much influence is given to connectivity between communities in the analysis. • The default value of 0.1 is suitable as a starting point; it can be increased/decreased as needed for exploration.
	Overlap Community Parameters	Analyzing overlap between communities in a network. • Overlap Detection Method: A method used to detect and identify communities that have overlapping members or connections. ◦ Core_expansion ◦ Umstmo ◦ Walkscan • Generate overlap visualization: An option to generate a visualization showing how these communities overlap in the network.

After all parameters have been successfully filled in, click **Run Analysis** to begin.

4. Input Example

For example, enter:

- Input Method: Disease Name
- Disease keyword: Diabetes Mellitus
- Organism: Homo sapiens (9606)
- Search method: Non-CC disease
- Community detection method: DGM
- Top N Community Consideration Centrality (CCC): 10
- Without activating additional parameters



Analysis Parameters

Choose Input Method:
● Disease Name
○ Manual Gene List

Enter Disease Name: diabetes mellitus

Basic Parameters

Organism NCBI Taxonomy ID: 9606
Enter NCBI Taxonomy ID (9606 for Homo sapiens)

Choose disease search query method:
● Non-CC disease
○ CC disease

Choose Community Detection Method: DGM

Top N Community Consideration Centrality (CCC): 10

☐ Enable Advanced Parameters

5. Run Analysis

1. Ensure all required parameters are filled in:
    a. Input method
    b. Disease name or gene list
    c. Organism ID
2. Set advanced parameters as needed. (Optional)
3. Click the Run Analysis button

5.1 Progress Bar

Once run, the Analysis Progress section will display the status:

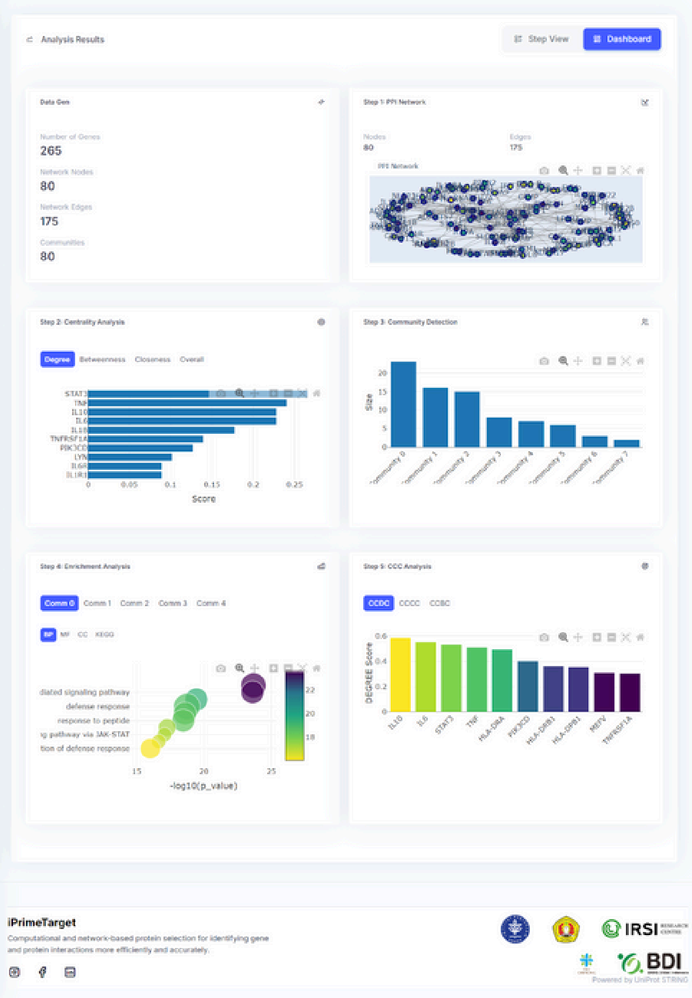
- Initializing analysis parameters
- Fetching disease/gene data
- Building protein-protein interaction network

Wait until the process is complete. If a failure occurs, the Analysis Error section will display a message that can be used for diagnosis (e.g., disease not found or connection to external service failed).

6. Dashboard

The dashboard is the main page of the application that displays a comprehensive summary of the analysis results. Through the dashboard, users can see an overview of all the stages that have been processed while accessing detailed results for each step of the analysis, namely:

- **Data Gen** that has been input as the basis for analysis.
- **Step 1: PPI Network**
- **Step 2: Centrality Analysis**
- **Step 3: Community Detection**
- **Step 4: Enrichment Analysis**
- **Step 5: CCC Analysis**



Analysis Results

Data Gen

Number of Genes
265

Network Nodes
80

Network Edges
175

Communities
80

Step 1: PPI Network

Nodes
80

Edges
175

iPrimeTarget

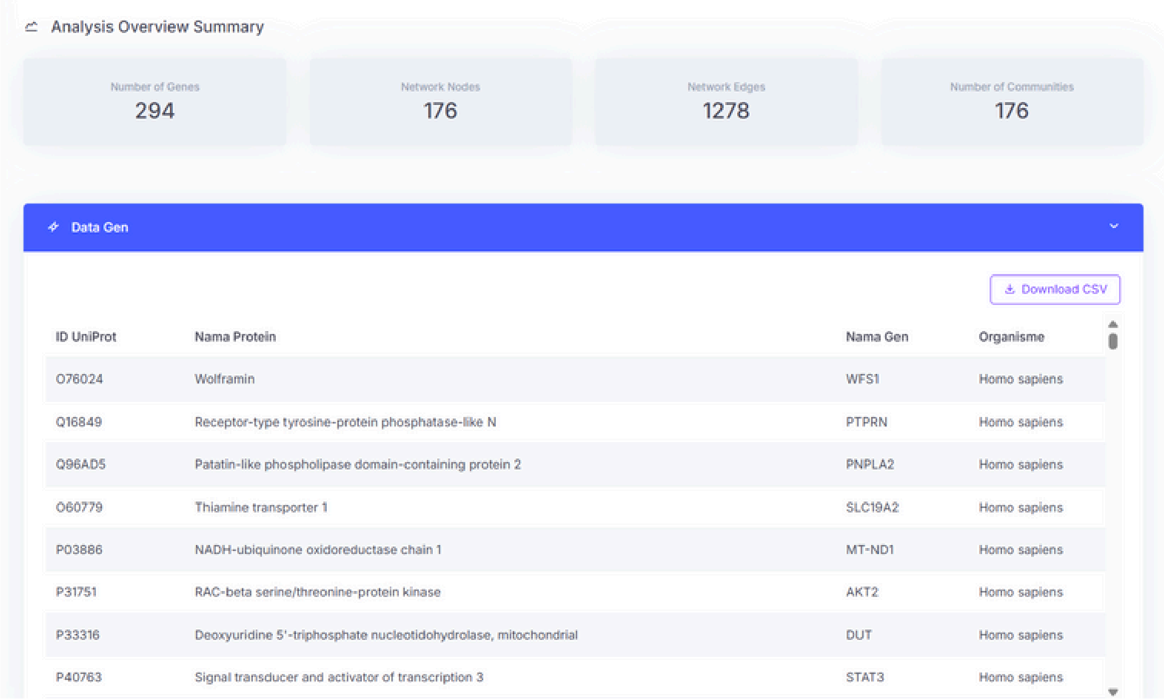
Computational and network-based protein selection for identifying gene and protein interactions more efficiently and accurately.

7. Analysis Results Navigation

- **Overview & Data Gen (UniProt Result)**

  This page displays:

  - A **list of proteins** from UniProt related to the disease/gene list entered.
  - The total number of proteins successfully found (Number of Genes, Network Nodes, Network Edges, Number of Communities)



Analysis Overview Summary

Number of Genes 294

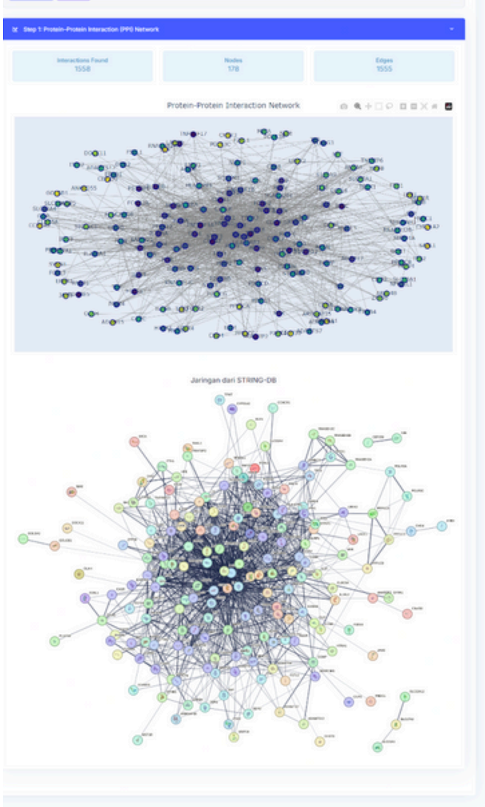
Network Nodes 176

Network Edges 1278

Number of Communities 176

ID UniProt	Nama Protein	Nama Gen	Organisme
O76024	Wolframin	WFS1	Homo sapiens
Q16849	Receptor-type tyrosine-protein phosphatase-like N	PTPRN	Homo sapiens
Q96AD5	Patatin-like phospholipase domain-containing protein 2	PNPLA2	Homo sapiens
O60779	Thiamine transporter 1	SLC19A2	Homo sapiens
P03886	NADH-ubiquinone oxidoreductase chain 1	MT-ND1	Homo sapiens
P31751	RAC-beta serine/threonine-protein kinase	AKT2	Homo sapiens
P33316	Deoxyuridine 5'-triphosphate nucleotidohydrolase, mitochondrial	DUT	Homo sapiens
P40763	Signal transducer and activator of transcription 3	STAT3	Homo sapiens

- **Step 1:** Protein-Protein Interaction (PPI) NetworkThe PPI Network page displays a PPI graph constructed from STRING.Users can zoom in, zoom out, download images, or pan the view for interactive exploration.



Step 1: Protein-Protein Interaction (PPI) Network

Interactions Found 1558

Nodes 178

Edges 1555

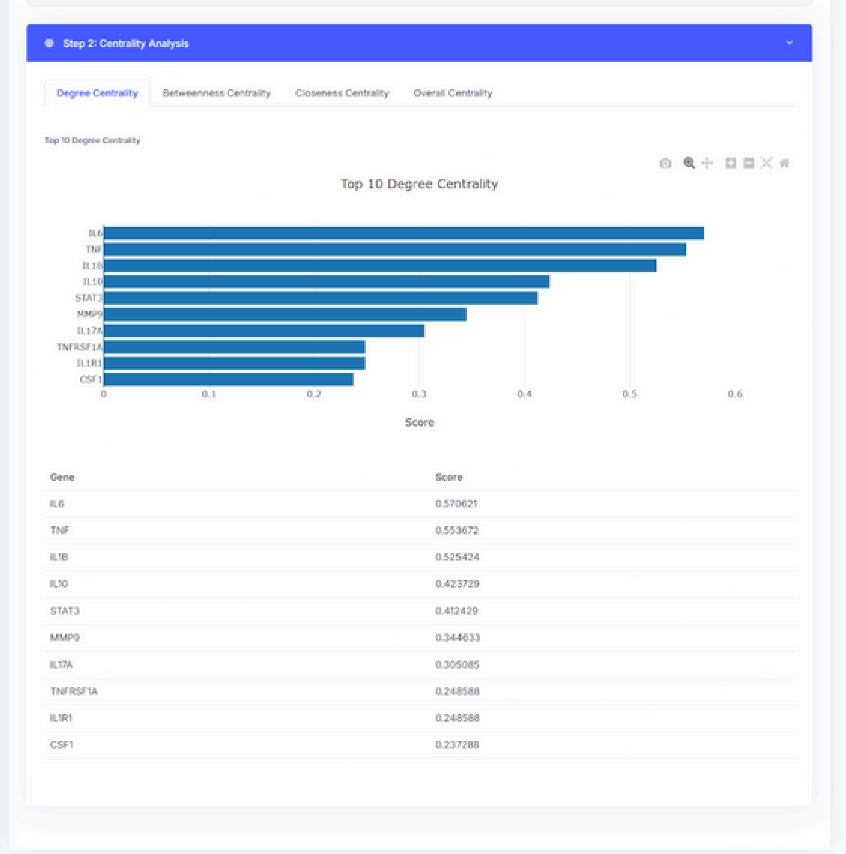
Contains PPI network visualization:

  - Interactive graph where nodes = proteins and edges = interactions.
  - Zoom/pan, node highlighting, and community display features.
  - Basic node information when hovered over (ID, degree).

- **Step 2: Centrality Analysis:** The results of the centrality analysis on the entire PPI network include:

Parameter	**Description**
Top Degree	Protein with the highest number of connections (degree).
Top Betweenness	Proteins most often serve as “bridges” between nodes.
Top Closeness	The protein with the shortest average distance to other proteins.
Top Overall (PCA)	Combined ranking based on Principal Component Analysis.

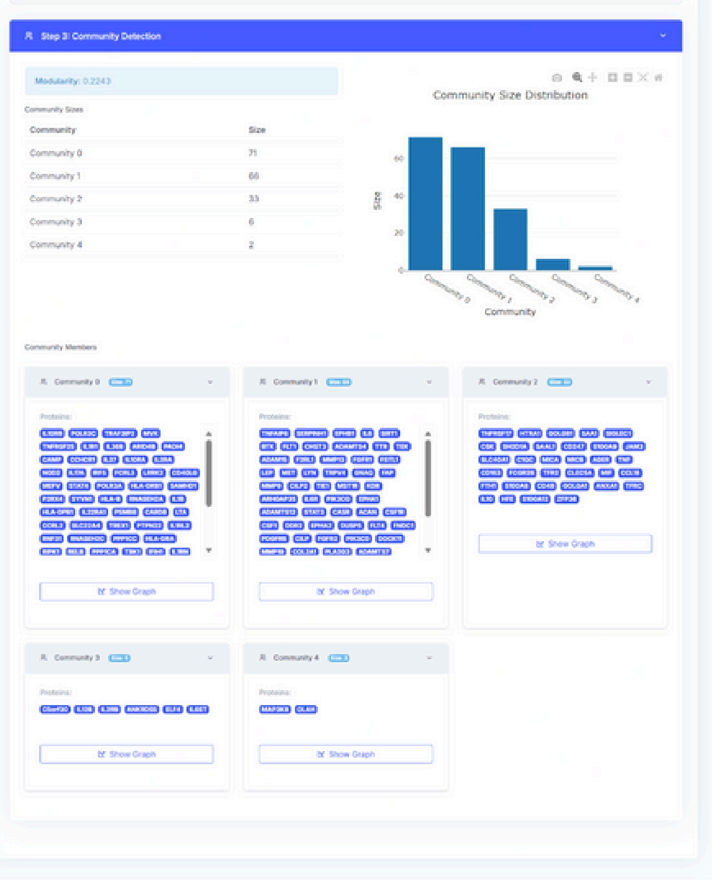
This section helps identify the proteins that most influence the disease.



Gene	Score
IL6	0.570621
TNF	0.553672
IL1B	0.525424
IL10	0.423729
STAT3	0.412429
MMP9	0.344633
IL17A	0.305085
TNFRSF1A	0.248588
IL1R1	0.248588
CSF1	0.237288

- **Step 3: Community Detection,** this page displays the formation of communities (clusters) and cliques in the network according to the algorithm specified at the beginning of the process with the following output:
  - **Modularity**
  - **List of communities** formed and the size of their members.
  - **A bar graph** showing the distribution of community sizes.

- **Community members** (number and list of proteins per community). Users can select Show Graph to display a visualization of the graph per community, complete with coloring based on community members. In the graph, each point (node) represents a protein or gene, while the lines (edges) indicate the existence of interactions or biological relationships between the two. Each node can be clicked to display a list of other nodes connected to it in the network. Users can zoom in, zoom out, download images, or shift the graph display for interactive exploration.



Modularity: 0.2243

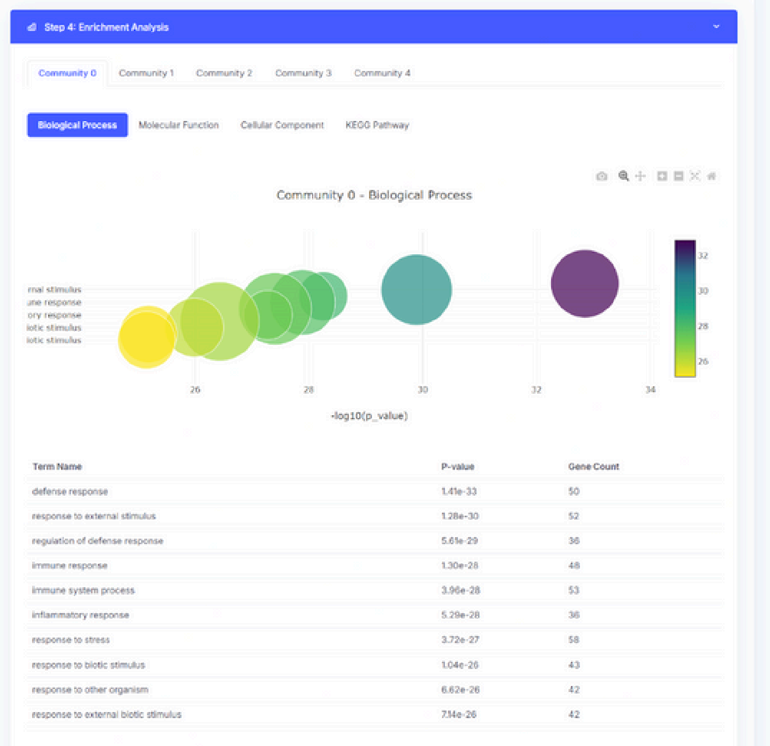
Community Sizes

Community	Size
Community 0	71
Community 1	66
Community 2	33
Community 3	6
Community 4	2

- **Step 4: Enrichment Analysis,** Enrichment Analysis is the process of analyzing enrichment in a community or entire network by identifying significant biological pathways, gene functions, or molecular processes.

  In this step, users can use databases such as GO (Gene Ontology) to understand the biological functions of each protein group with the following category options:
  - Biological Process
  - Molecular Function
  - Cellular Component
  - KEGG Pathway

  The analysis results are presented in a graph table containing:
  - **Term Name:** Name of the enriched biological process.
  - **P-value:** Statistical probability value. The smaller the value, the stronger the relationship between the gene and the term.
  - **Gene Count:** Number of genes contributing to the term.



Term Name	P-value	Gene Count
defense response	1.41e-33	50
response to external stimulus	1.28e-30	52
regulation of defense response	5.61e-29	36
immune response	1.30e-28	48
immune system process	3.96e-28	53
inflammatory response	5.29e-28	36
response to stress	3.72e-27	58
response to biotic stimulus	1.04e-26	43
response to other organism	6.62e-26	42
response to external biotic stimulus	7.14e-26	42

- **Step 5: Community Consideration Centrality (CCC),** this page displays the most central (important) genes or proteins in interaction networks based on various centrality metrics. This analysis helps identify key genes (hub genes) that play a major role in a biological community or network.

  This section contains several centrality metric tabs that can be selected:
  - **CCDC (Community Closeness Degree Centrality)**
  - **CCCC (Community Closeness Clustering Centrality)**
  - **CCBC (Community Betweenness Centrality)**

8. Usage Tips and Best Practices

1. Start with default parameters

   For new users, use the default values (Organism = 9606, required score = 0.40 or 0.70) and then explore the changes.

2. Validate the gene list

   Always check the Gen Data tab to ensure that the genes extracted from the disease are relevant to the literature you are familiar with.

3. Pay attention to network density

   If the network is too dense/heavy, increase the required score value.

4. Compare several community methods

   Try Louvain, Leiden, and other methods to see if the community structure is stable.

5. Use CCC as the final stage

   a. First identify important communities (e.g., communities with specific pathways).

   b. Use the CCC tab to select key proteins per community as priority targets.

9. Handling Common Errors

- “Analysis Required. Please run the analysis first to view this section.”
  - Run the analysis again from the Analysis Home / Overview tab.
- No genes found for a specific disease.
  - Try varying the disease name (e.g., use more common MeSH/OMIM terms).
  - Ensure a stable internet connection.
- Analysis takes a long time at the data fetching stage
  - This may be related to external service limitations. Try again a few minutes later or limit the complexity (number of genes, high STRING score).