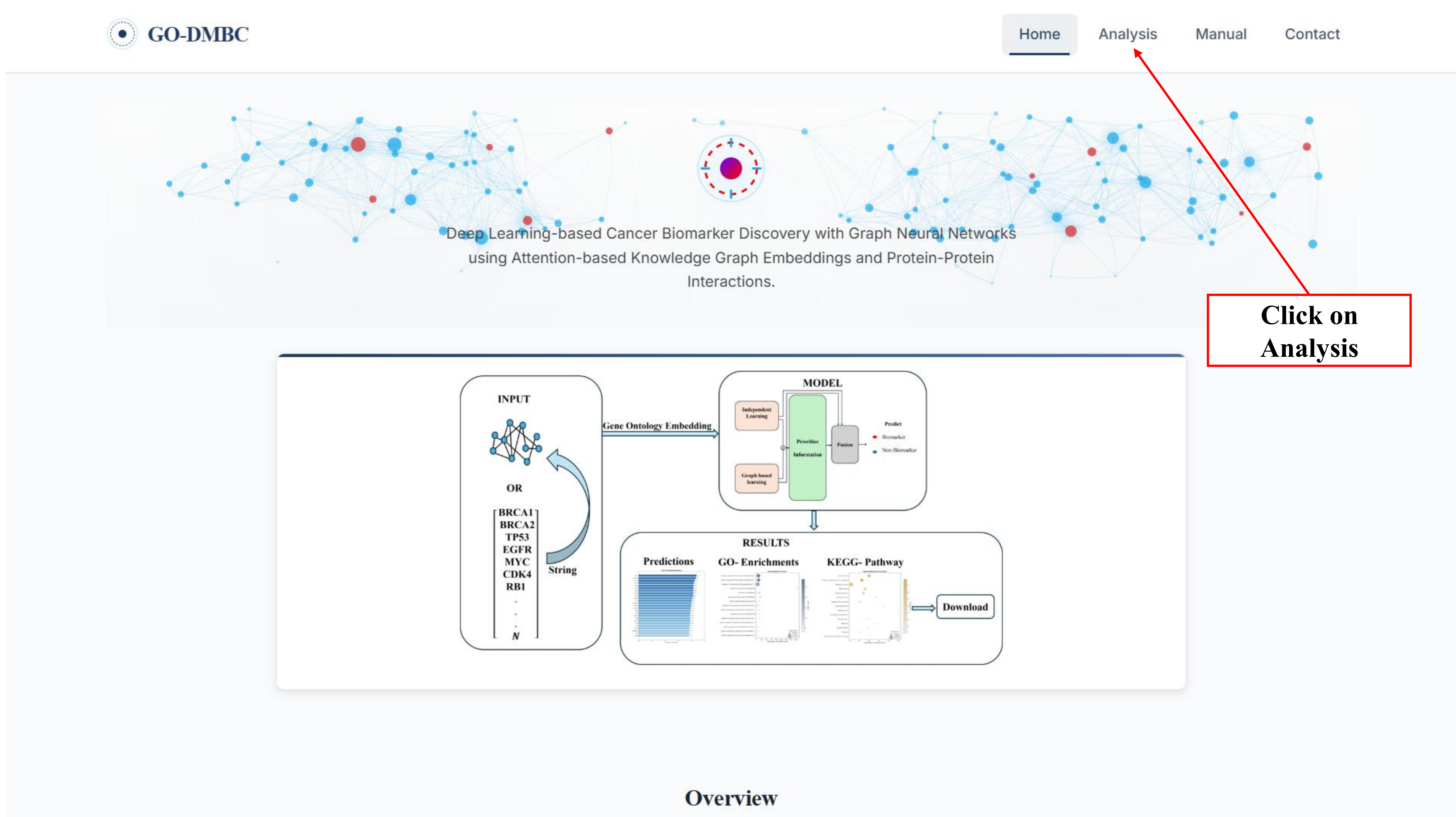
GO-DMBC MANUAL

Homepage

The header on the homepage contains :

- Home
- Analysis
- Manual
- Contact

To perform the Analysis, click on the Analysis.



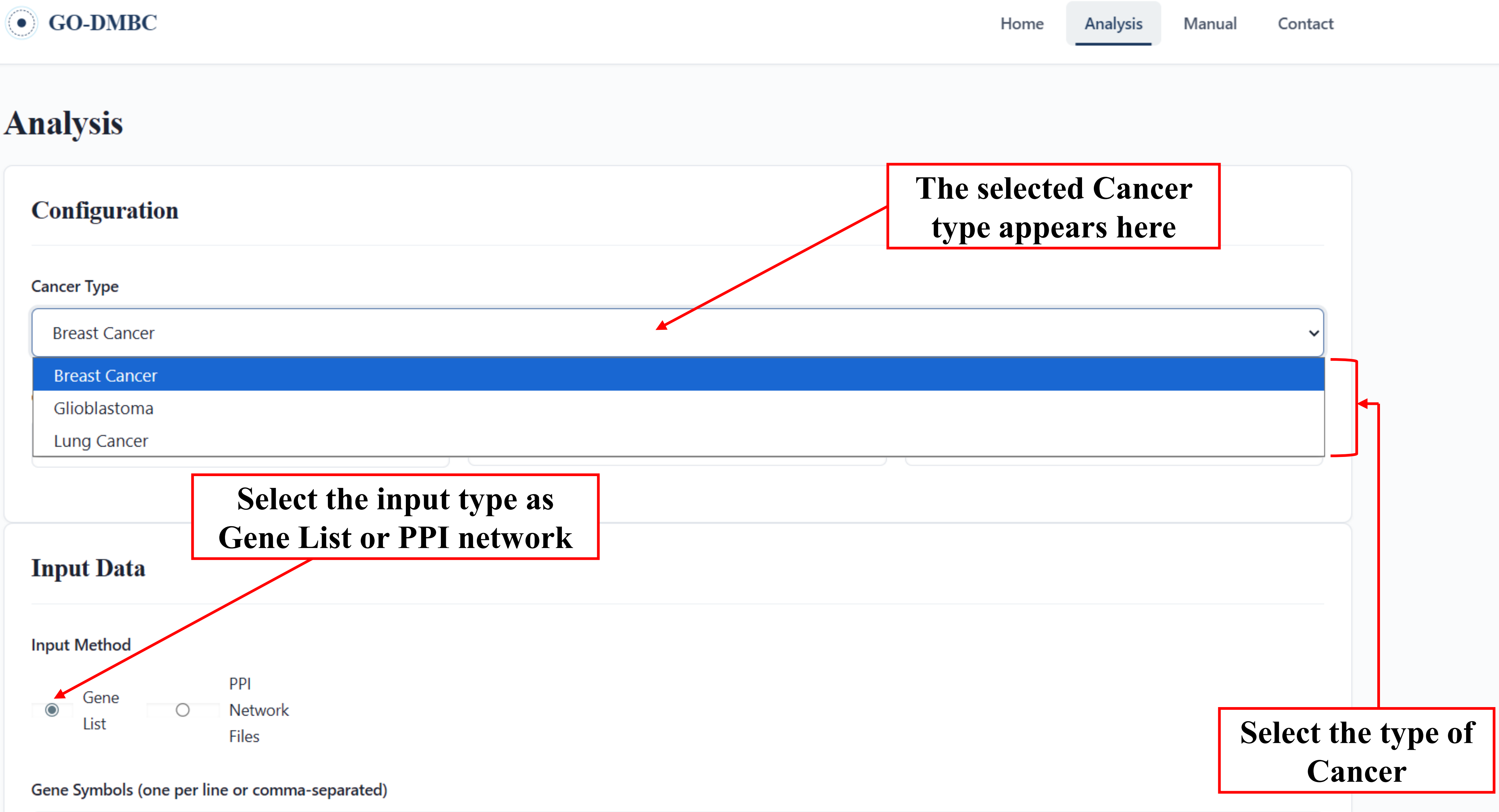
Deep Learning-based Cancer Biomarker Discovery with Graph Neural Networks using Attention-based Knowledge Graph Embeddings and Protein-Protein Interactions.

Overview

Click on Analysis

The Analysis Page opens where you can:

- Select the cancer type
- Choose the Input Data Type



Analysis

Configuration

Cancer Type

Breast Cancer

Input Data

Input Method

● Gene List ○ PPI Network Files

Gene Symbols (one per line or comma-separated)

The selected Cancer type appears here

Select the input type as Gene List or PPI network

Select the type of Cancer

If you choose Gene List:

- Enter the gene list in the and run the analysis as shown.

Input Data

Input Method

● Gene List ○ PPI Network Files

Gene Symbols (one per line or comma-separated)

```
ABAT
ABCA6
ABCA9
ACAA2
ACACB
ACADL
ACADS
ACKR1
ACKR3
ACKR4
```

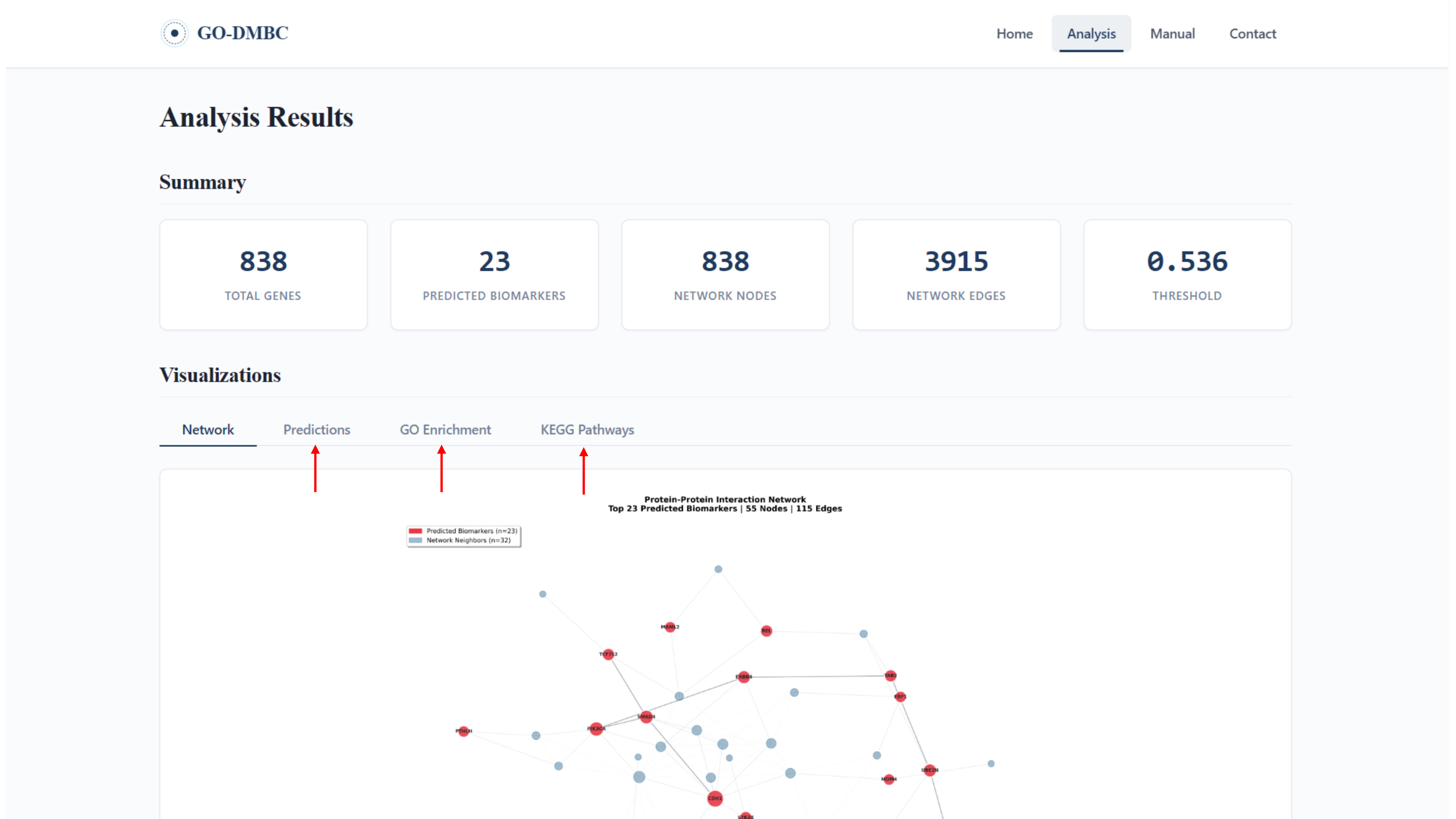
Enter Gene List here

Click on Run Analysis

Wait, and after some time, the analysis results are displayed

The results consist of :

- Summary: Consisting of the total number of genes you submitted, the total number of biomarkers predicted, the number of network nodes and edges (genes with string and GO mapping), and the prediction threshold.
- Visualization: The network is visualized, and the predictions are accompanied by GO enrichment and KEGG pathway analysis with plots.



GO-DMBC

Home Analysis Manual Contact

Analysis Results

Summary

TOTAL GENES	PREDICTED BIOMARKERS	NETWORK NODES	NETWORK EDGES	THRESHOLD
838	23	838	3915	0.536

Visualizations

Network Predictions GO Enrichment KEGG Pathways

Prediction Probability Distribution

The probability distribution plot shows the model-assigned biomarker probabilities for all input genes.
Most genes receive low probabilities, reflecting the strong class imbalance inherent in biomarker discovery, while a smaller subset exceeds the selected prediction threshold.
The vertical dashed line indicates the confidence threshold used to classify genes as predicted biomarkers.

This visualization allows users to:

- Inspect score separation between predicted biomarkers and non-biomarkers
- Adjust confidence thresholds based on downstream requirements
- Assess overall model calibration and prediction sharpness

Top Biomarkers

The Top Biomarkers bar plot ranks the highest-confidence predicted biomarkers based on their prediction probabilities.
Each bar corresponds to a gene, ordered from highest to lowest probability.

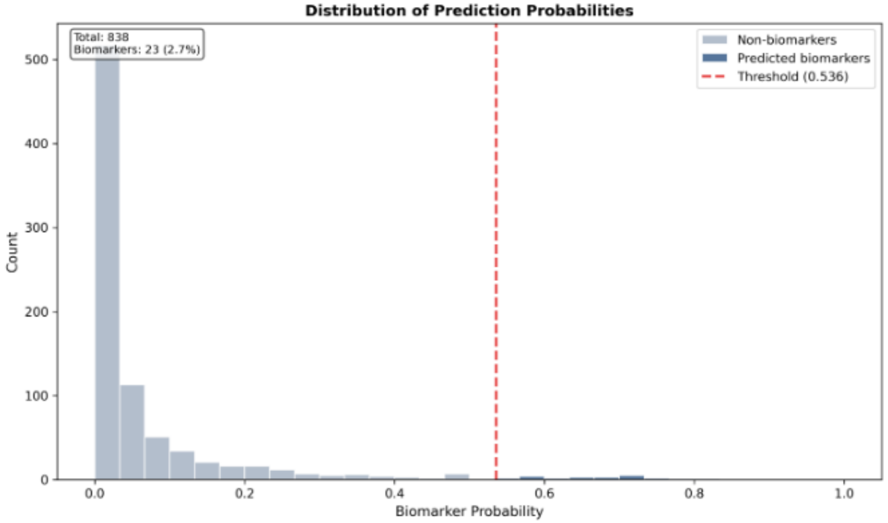
This plot highlights:

- High-confidence candidate biomarkers
- Relative ranking among predictions
- Immediate candidates for experimental validation
- The same information is also provided in a tabular format listing Gene/Protein symbol based on prediction probability.

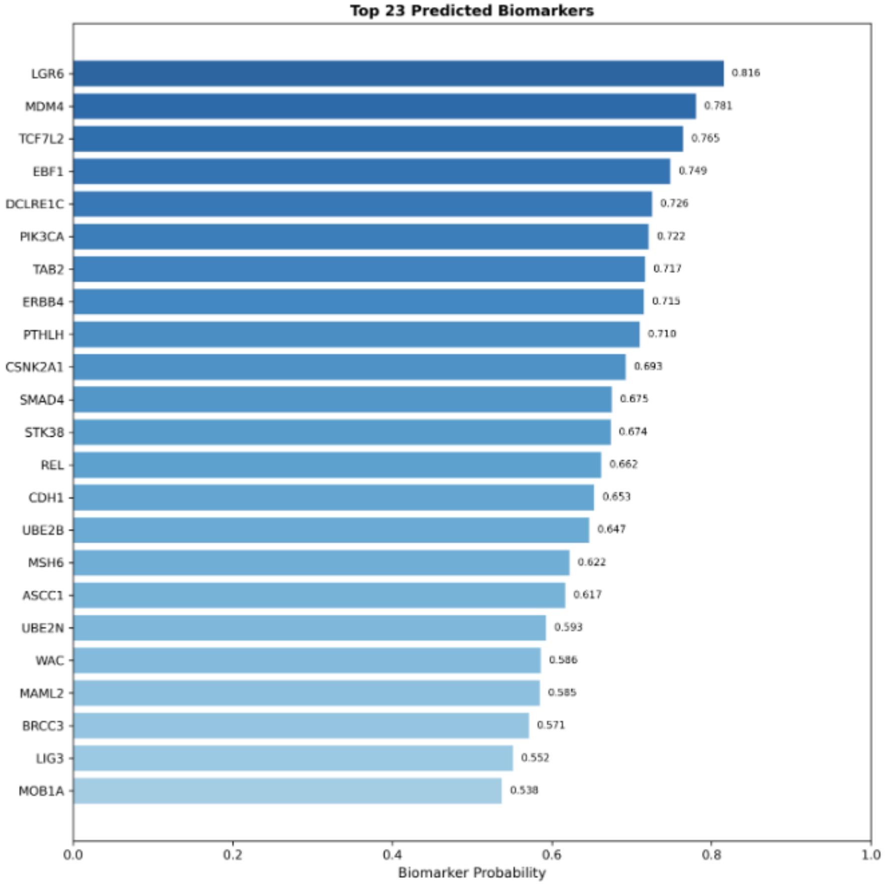
Visualizations

Network Predictions GO Enrichment KEGG Pathways

Probability Distribution



Top Biomarkers



Gene Ontology (GO) Enrichment Analysis

GO enrichment analysis is performed on the set of predicted biomarkers to identify overrepresented biological themes.
Results are grouped into the three standard GO categories:

a. Biological Process (BP)

- Enriched biological processes reflect pathways related to DNA repair, transcriptional regulation, signaling, and cellular maintenance, providing insight into disease-relevant mechanisms.

b. Molecular Function (MF)

- Molecular function terms highlight activities such as kinase activity, ubiquitin-related functions, and DNA-binding roles, indicating functional characteristics of the predicted biomarkers.

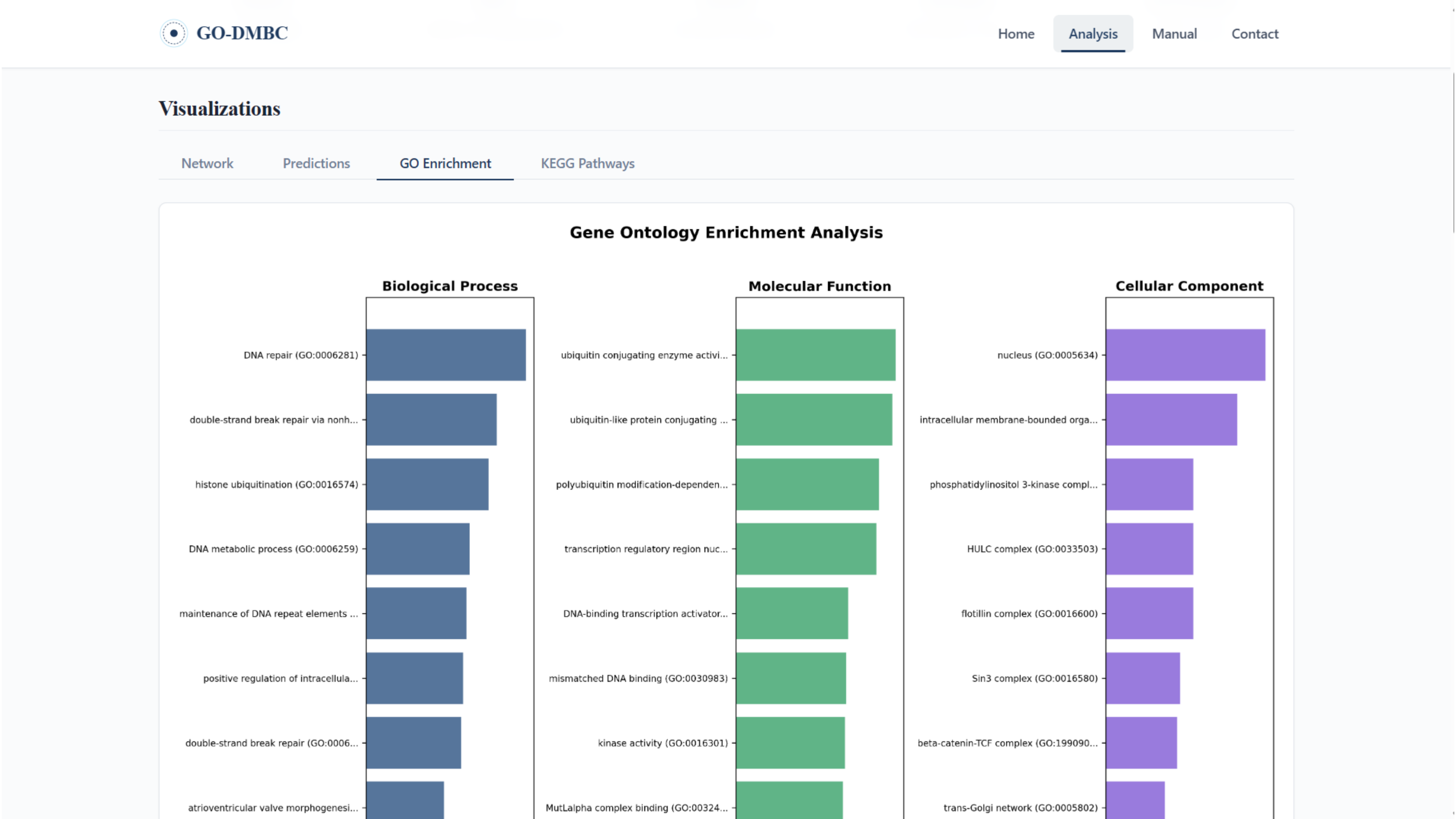
c. Cellular Component (CC)

- Cellular component enrichment reveals the subcellular localization of predicted biomarkers, including nuclear complexes, membrane-bound organelles, and signaling assemblies.

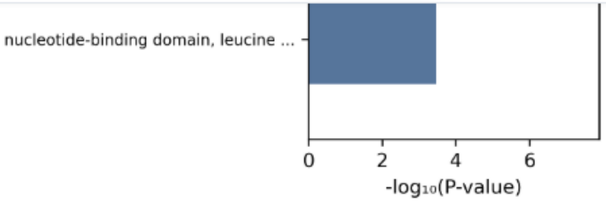
For each GO category, results are visualized using:

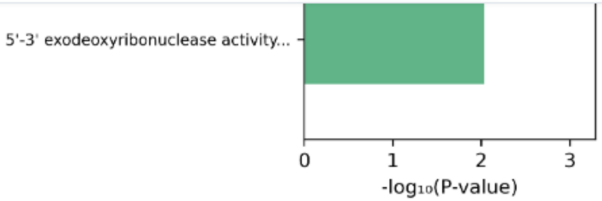
- Bar plots showing statistical significance ($-\log_{10}$ p-value)
- Bubble plots, where bubble size represents gene count and color intensity reflects enrichment significance

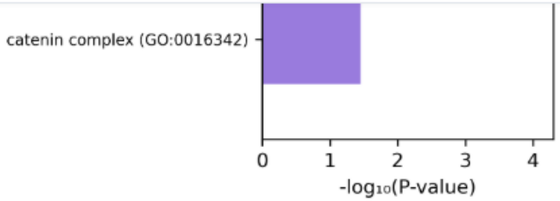
These visualizations enable both statistical and biological interpretation of enrichment results.



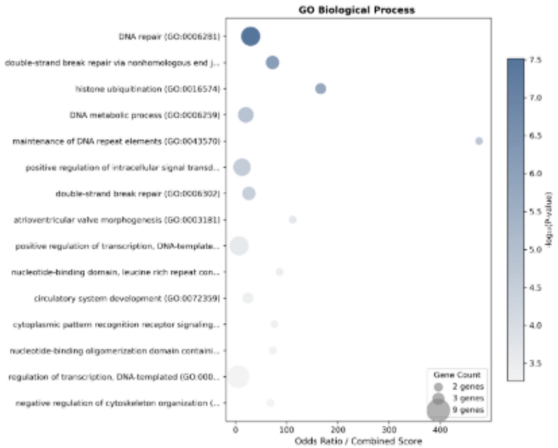
Visualizations



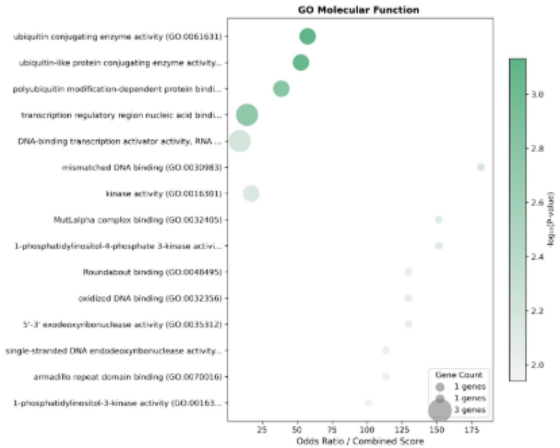




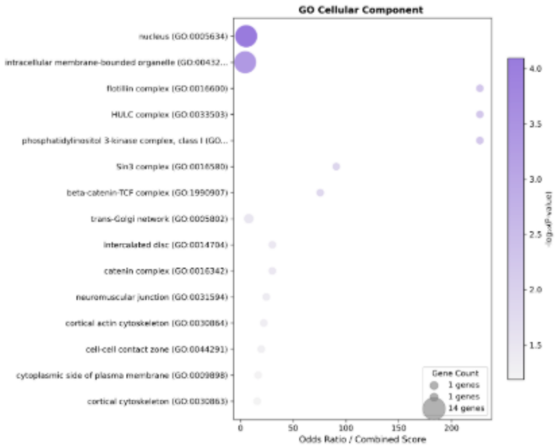
GO Biological Process



GO Molecular Function



GO Cellular Component



Predicted Biomarkers

RANK	PROTEIN	PROBABILITY
1	**LGR6**	0.8157
2	**MDM4**	0.7809
3	**TCF7L2**	0.7646

KEGG Pathway Enrichment Analysis

KEGG pathway analysis identifies signaling and disease-related pathways enriched among the predicted biomarkers.

The KEGG bar plot displays pathways significantly enriched, ranked by statistical significance, including cancer-associated pathways and key signaling cascades.

The accompanying KEGG bubble plot provides an integrated view:

- X-axis: odds ratio or combined enrichment score
- Bubble size: number of genes involved
- Color scale: $-\log_{10}$ adjusted p-value

Together, these plots allow users to:

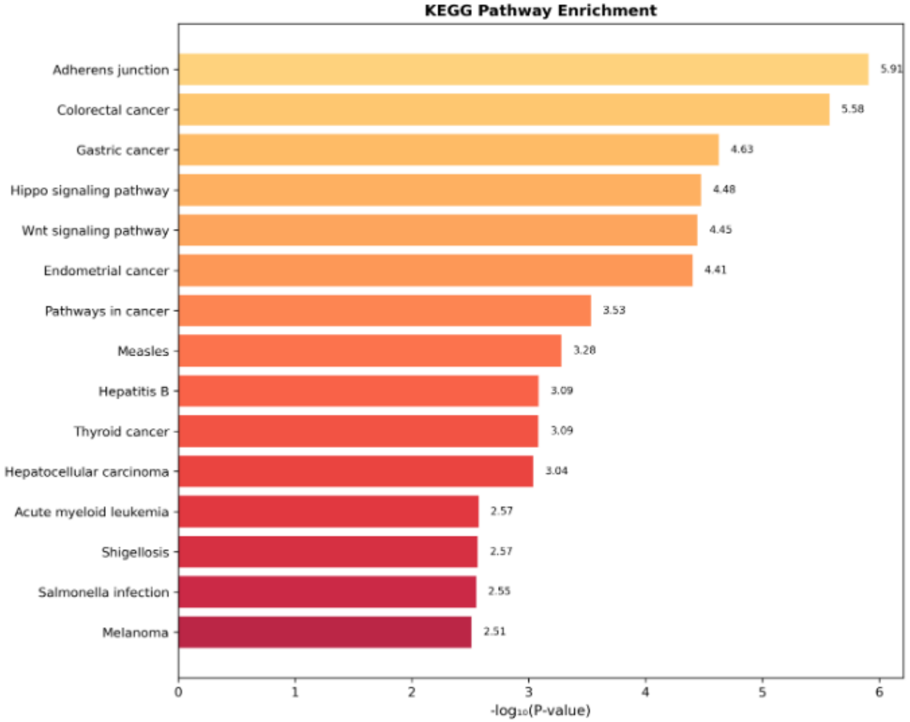
- Identify dominant disease pathways
- Compare pathway relevance at a glance
- Link biomarker predictions to known cancer biology

A detailed KEGG results table is also provided, including p-values, adjusted p-values, enrichment scores, and contributing genes.

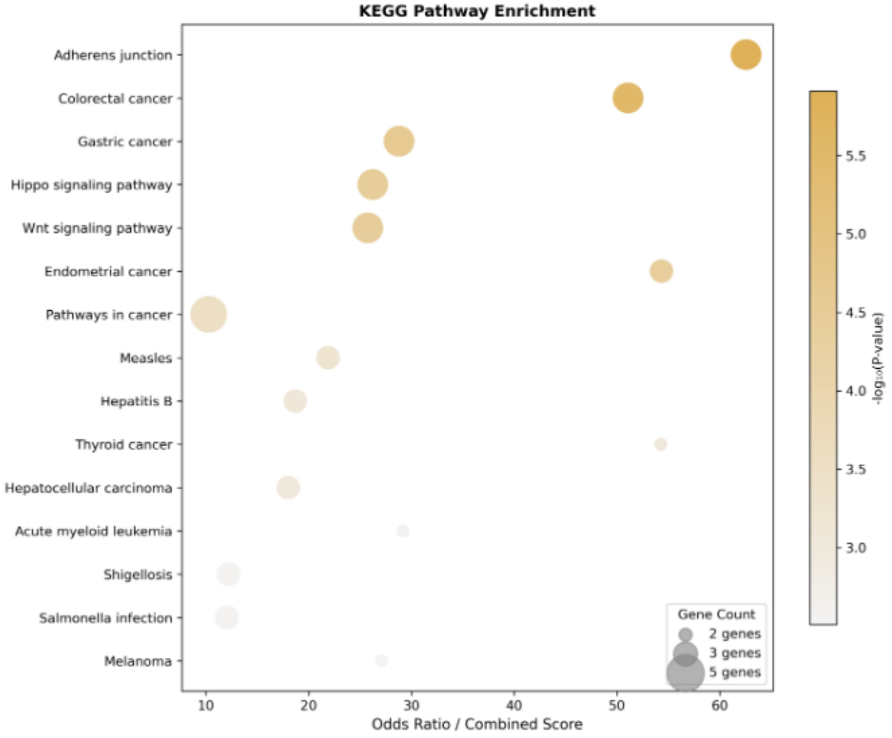
Visualizations

Network Predictions GO Enrichment **KEGG Pathways**

KEGG Pathway Enrichment



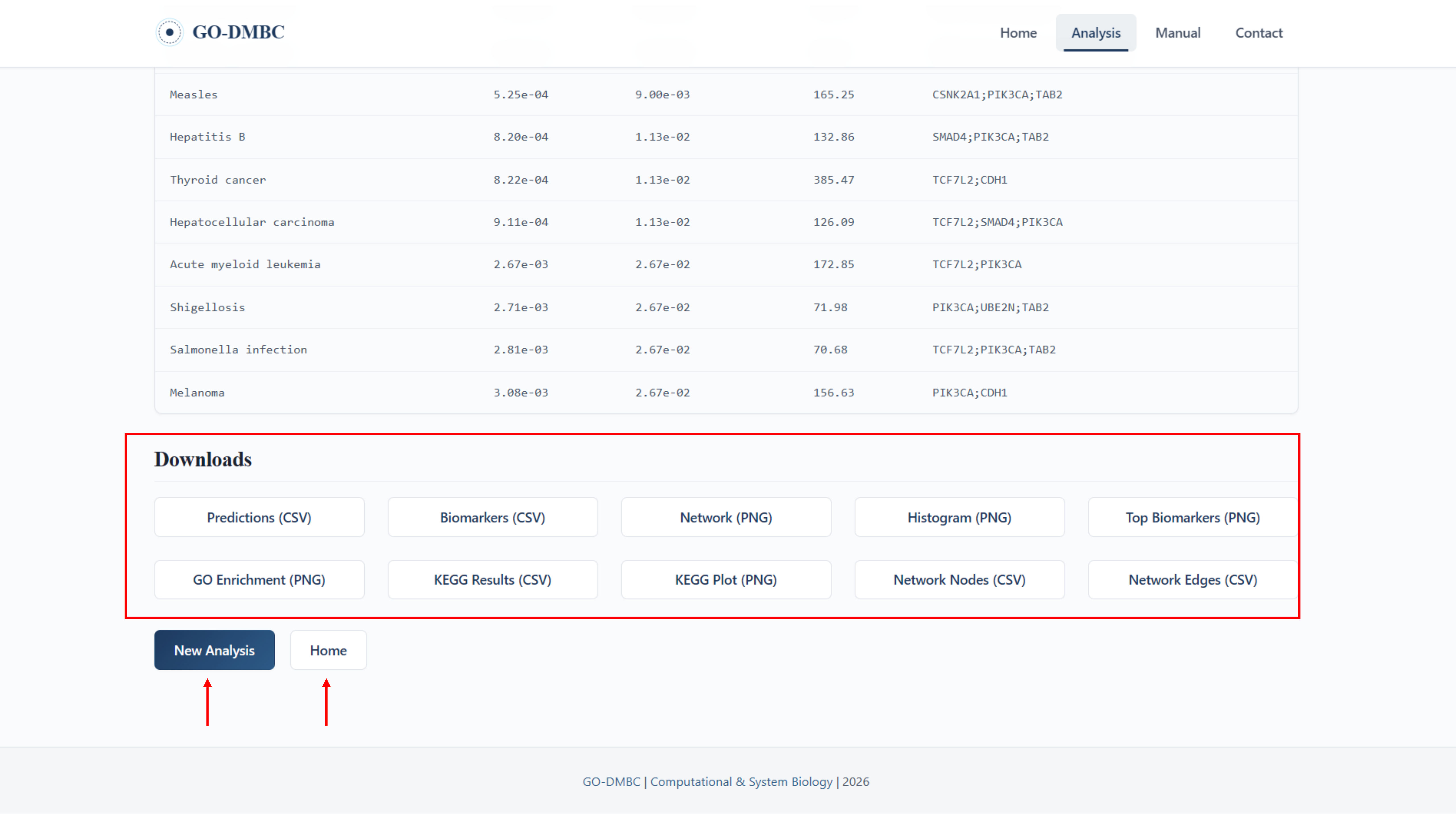
KEGG Bubble Plot



Downloadable Outputs

- All major results can be downloaded for offline analysis, including:
- Prediction scores (CSV)
- Predicted biomarkers list (CSV)
- Network visualization images (PNG)
- GO enrichment plots (PNG)
- KEGG enrichment plots and tables (PNG / CSV)
- Network node and edge files (CSV)

These outputs ensure reproducibility and facilitate integration with downstream bioinformatics pipelines.



Measles	5.25e-04	9.00e-03	165.25	CSNK2A1;PIK3CA;TAB2
Hepatitis B	8.20e-04	1.13e-02	132.86	SMAD4;PIK3CA;TAB2
Thyroid cancer	8.22e-04	1.13e-02	385.47	TCF7L2;CDH1
Hepatocellular carcinoma	9.11e-04	1.13e-02	126.09	TCF7L2;SMAD4;PIK3CA
Acute myeloid leukemia	2.67e-03	2.67e-02	172.85	TCF7L2;PIK3CA
Shigellosis	2.71e-03	2.67e-02	71.98	PIK3CA;UBE2N;TAB2
Salmonella infection	2.81e-03	2.67e-02	70.68	TCF7L2;PIK3CA;TAB2
Melanoma	3.08e-03	2.67e-02	156.63	PIK3CA;CDH1

Downloads

GO-DMBC | Computational & System Biology | 2026

PPI Network Input

GO-DMBC supports protein–protein interaction (PPI) network-based analysis, enabling users to predict biomarkers directly on custom interaction networks.

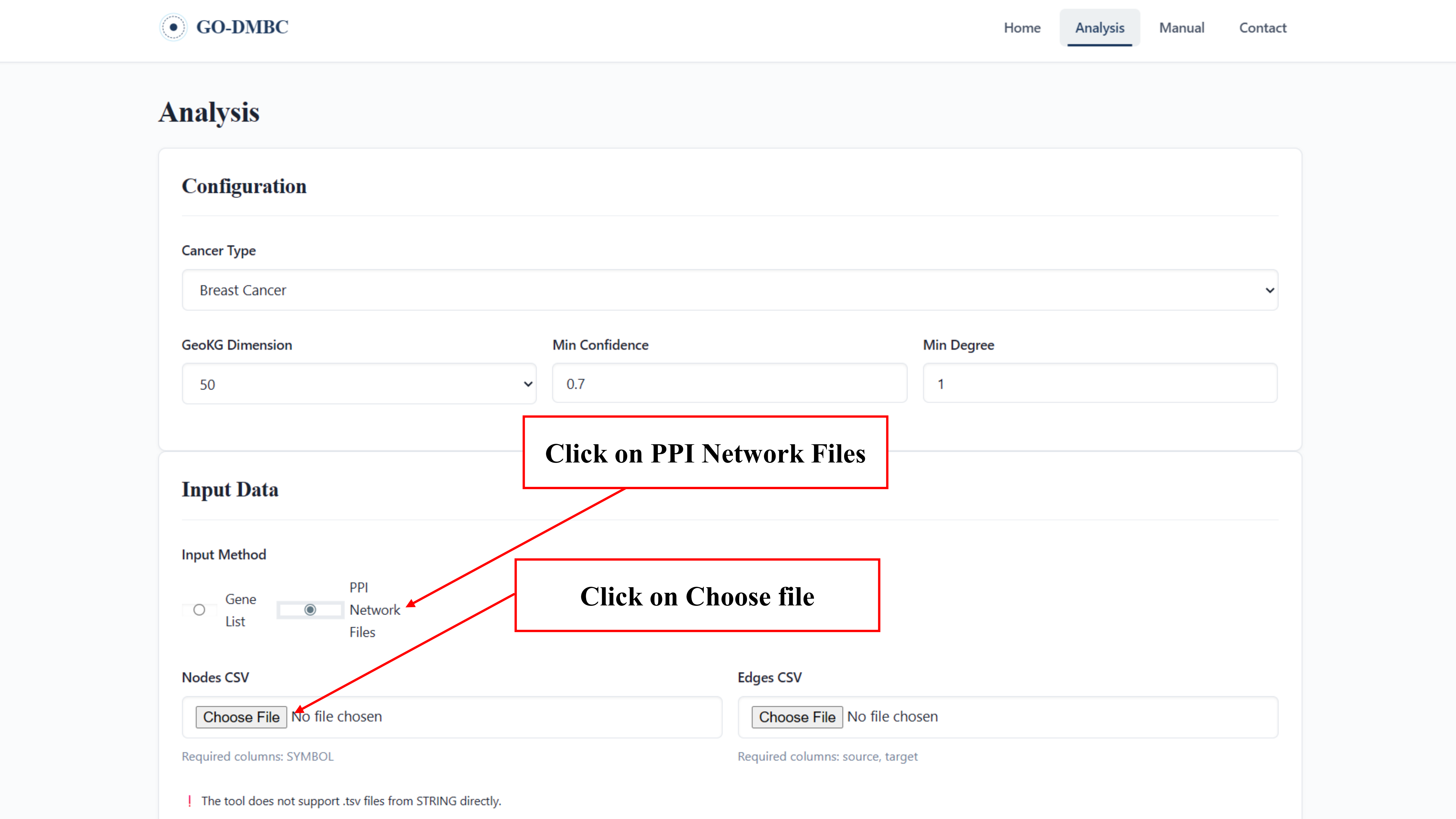
Users must select PPI Network Files as the input method and upload two CSV files:

1. Nodes File (CSV)

- This file defines the set of genes/proteins in the network.

2. Edges File (CSV)

- This file defines interactions between genes/proteins.



Analysis

Configuration

Cancer Type: Breast Cancer

GeoKG Dimension: 50

Min Confidence: 0.7

Min Degree: 1

Click on PPI Network Files

Input Data

Input Method

○ Gene List
● PPI Network Files

Click on Choose file

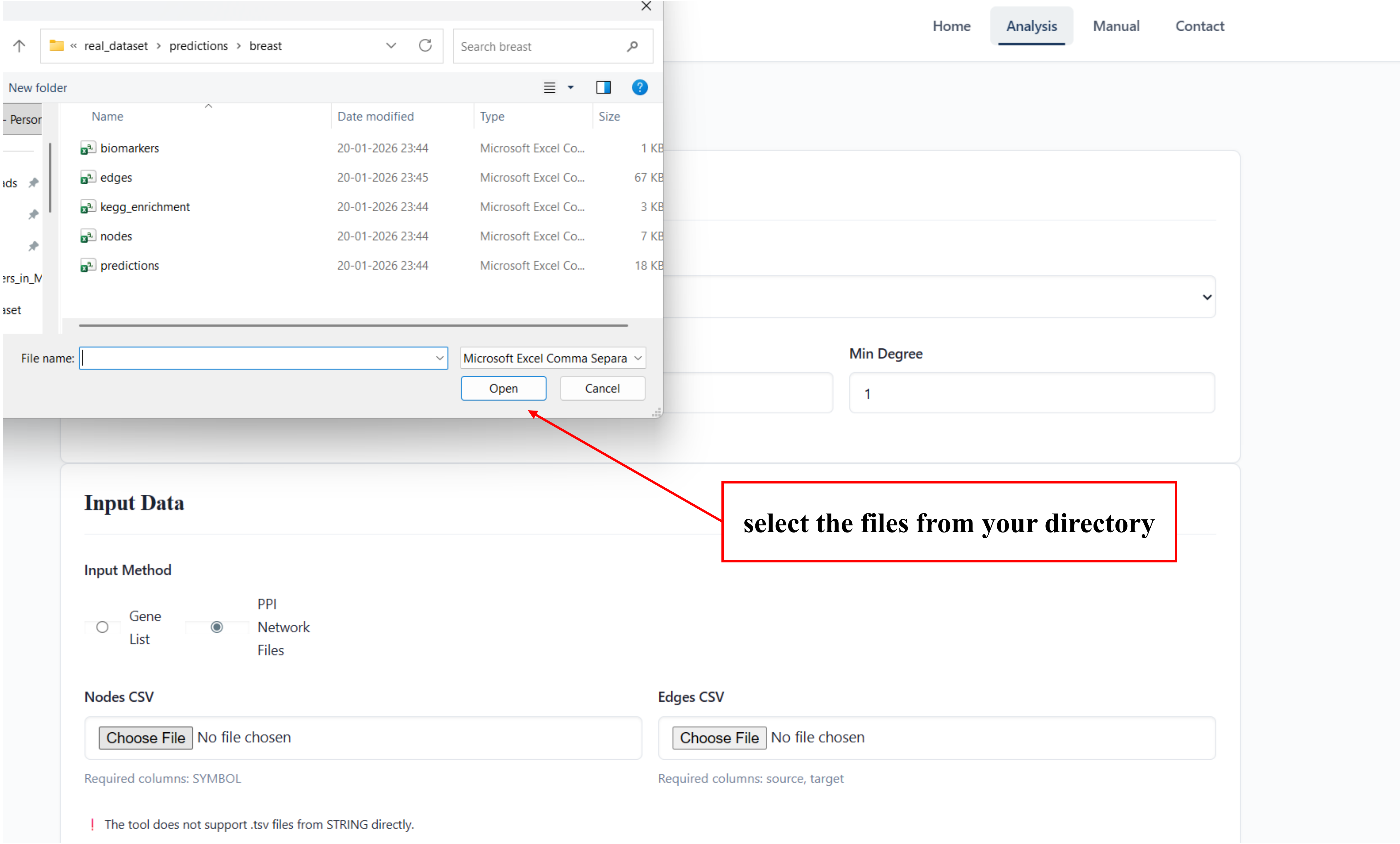
Nodes CSV: No file chosen

Required columns: SYMBOL

Edges CSV: No file chosen

Required columns: source, target

! The tool does not support .tsv files from STRING directly.



Name	Date modified	Type	Size
biomarkers	20-01-2026 23:44	Microsoft Excel Co...	1 KB
edges	20-01-2026 23:45	Microsoft Excel Co...	67 KB
kegg_enrichment	20-01-2026 23:44	Microsoft Excel Co...	3 KB
nodes	20-01-2026 23:44	Microsoft Excel Co...	7 KB
predictions	20-01-2026 23:44	Microsoft Excel Co...	18 KB

File name:

Microsoft Excel Comma Separa

Min Degree: 1

select the files from your directory

Input Data

Input Method

○ Gene List ● PPI Network Files

Nodes CSV

Choose File No file chosen

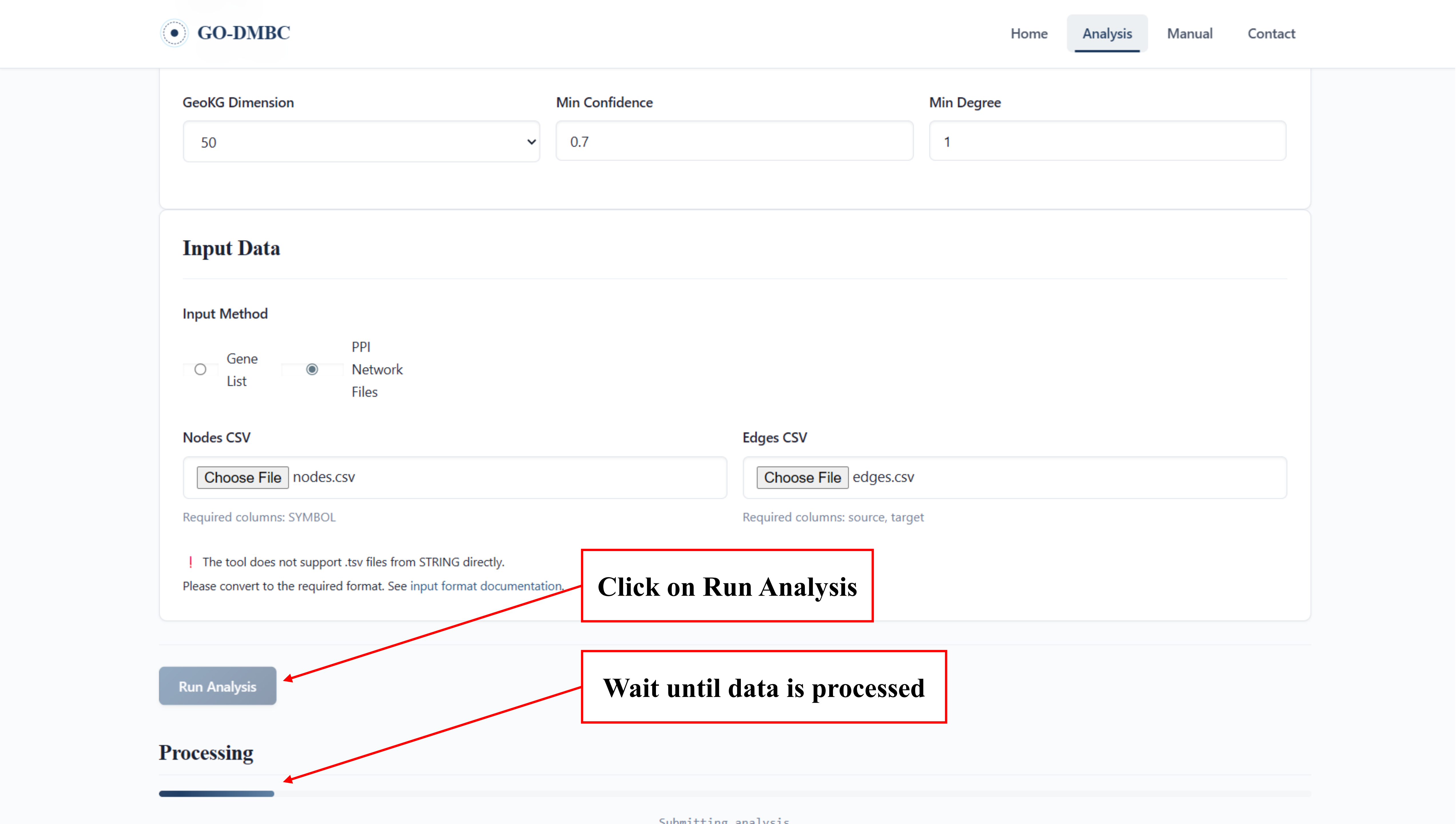
Required columns: SYMBOL

Edges CSV

Choose File No file chosen

Required columns: source, target

! The tool does not support .tsv files from STRING directly.



GeoKG Dimension: 50

Min Confidence: 0.7

Min Degree: 1

Input Data

Input Method

○ Gene List ● PPI Network Files

Nodes CSV: nodes.csv

Required columns: SYMBOL

Edges CSV: edges.csv

Required columns: source, target

! The tool does not support .tsv files from STRING directly.
Please convert to the required format. See input format documentation.

Processing

Submitting analysis

Click on Run Analysis

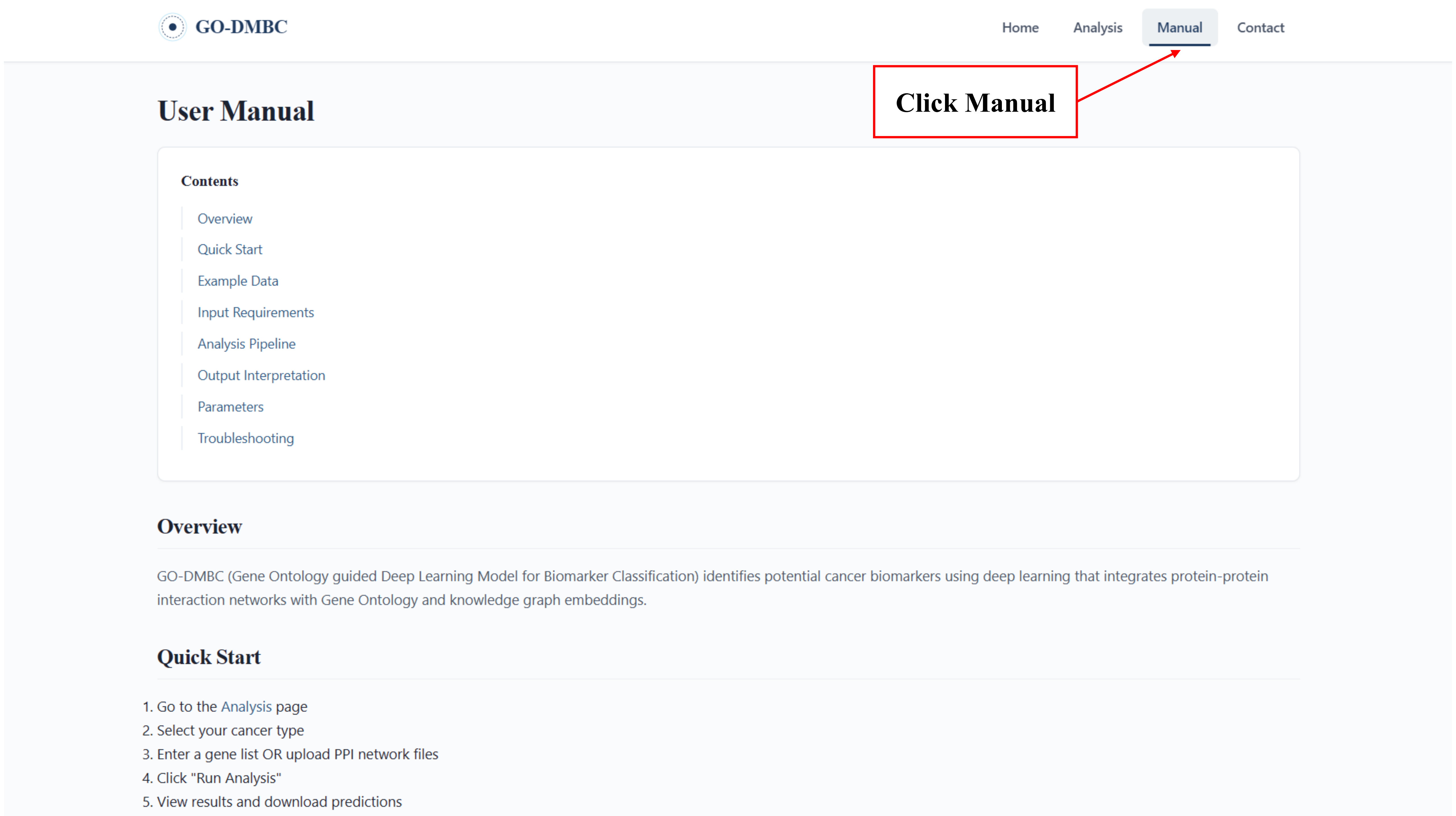
Wait until data is processed

User Manual

The User Manual provides detailed guidance on:

- Input data preparation
- Supported file formats
- Parameter selection
- Analysis workflow

For step-by-step instructions and example use cases, please refer to the relevant sections of the manual.



Click Manual

User Manual

Contents

Overview

GO-DMBC (Gene Ontology guided Deep Learning Model for Biomarker Classification) identifies potential cancer biomarkers using deep learning that integrates protein-protein interaction networks with Gene Ontology and knowledge graph embeddings.

Quick Start

1. Go to the Analysis page
2. Select your cancer type
3. Enter a gene list OR upload PPI network files
4. Click "Run Analysis"
5. View results and download predictions

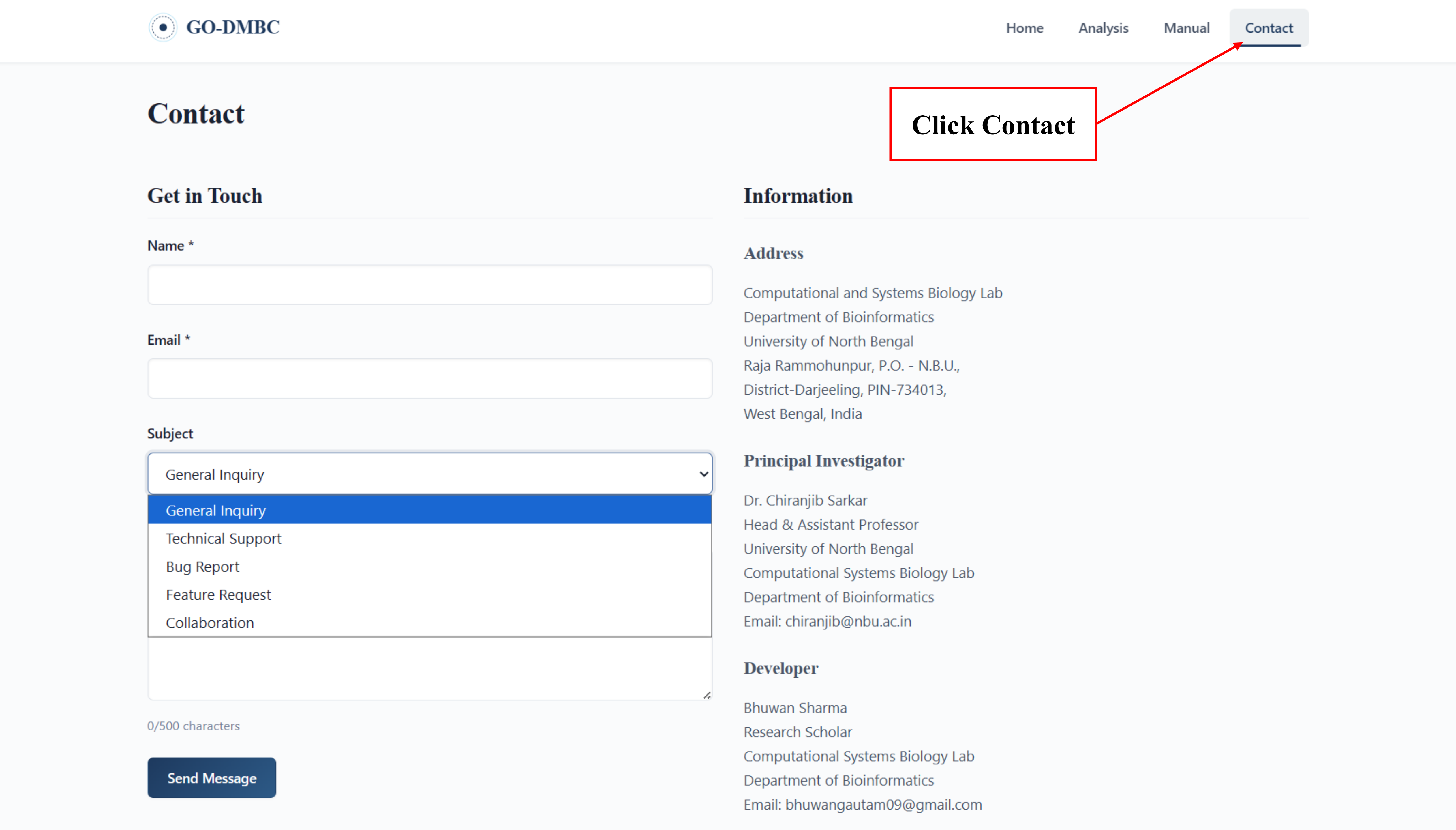
Contact & Support

For technical issues, unexpected errors, or questions related to the analysis results, please contact us.

If you experience:

- File upload failures
- Format-related errors
- Processing interruptions
- Inconsistencies in output files

Please first consult the User Manual and Troubleshooting section.
If the issue persists, feel free to reach out to the development team.



Contact

Get in Touch

Name *

Email *

Subject

General Inquiry

0/500 characters

Information

Address

Computational and Systems Biology Lab
Department of Bioinformatics
University of North Bengal
Raja Rammohunpur, P.O. - N.B.U.,
District-Darjeeling, PIN-734013,
West Bengal, India

Principal Investigator

Dr. Chiranjib Sarkar
Head & Assistant Professor
University of North Bengal
Computational Systems Biology Lab
Department of Bioinformatics
Email: chiranjib@nbu.ac.in

Developer

Bhuwan Sharma
Research Scholar
Computational Systems Biology Lab
Department of Bioinformatics
Email: bhuwangautam09@gmail.com

Discover smarter with GO-DMBC!

~Because biology is messy, your insights shouldn’t be!