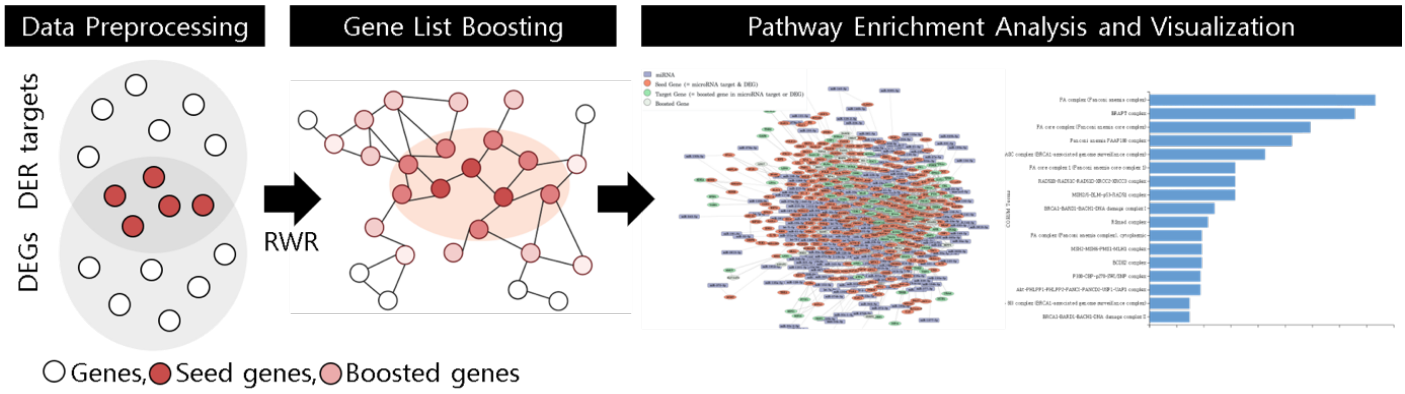
MIMR tutorial

(2024.10)



Important Notes:

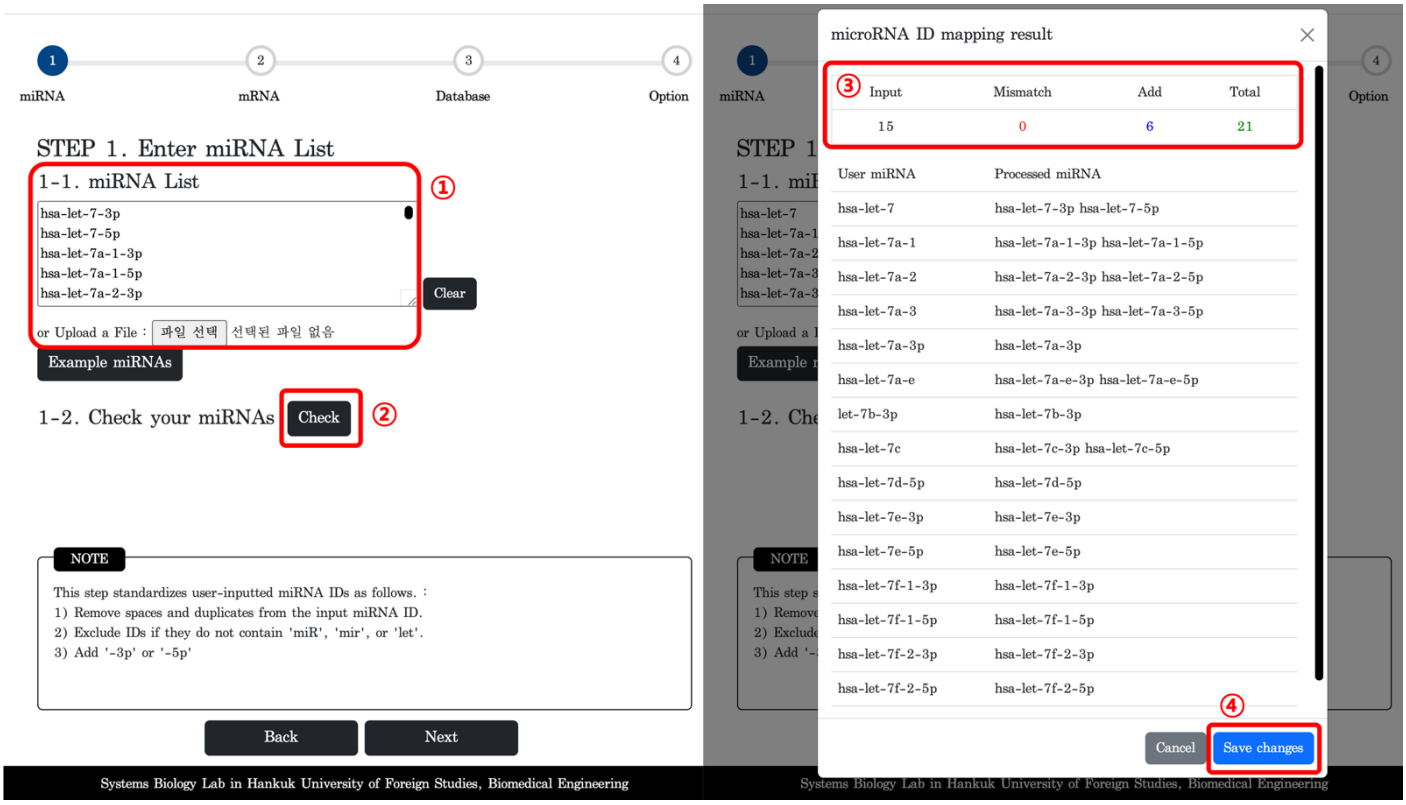
1. MIMR only supports human data.

2. The analysis of miRNA and mRNA can be performed independently, allowing you to analyze either miRNA or mRNA alone.

Content

1. Analysis tutorial

1.1 Input and Transform miRNA Data



1. Enter miRNA List: Input the list of miRNAs to be analyzed in the text box or upload a file. Supported delimiters include commas (,), line breaks (Enter), and spaces. If you're uploading a file, click the 'Choose File' button to select it.

- Note: Only IDs starting with 'hsa-', 'miR', 'mir', or 'let' are accepted, and duplicate miRNAs will be automatically removed. If miRNAs are not specified as '3p' or '5p', both versions will be included.

2. Click the Check Button: Press the 'Check' button to convert the miRNA IDs into the standard format.

3. Review Conversion Results: The conversion results will be displayed in a new window, where you can check the following information:

- Input: Number of miRNAs entered

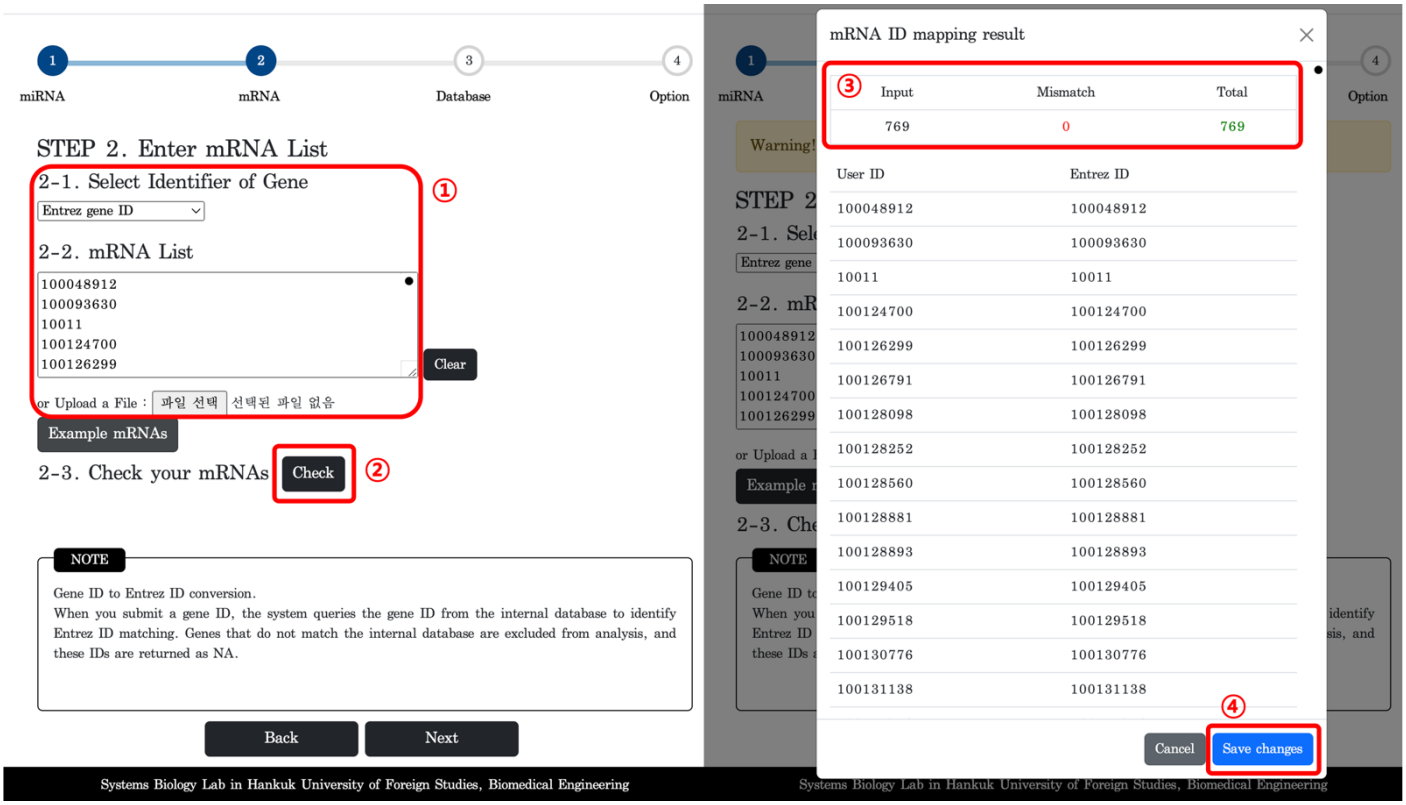
- Mismatch: Number of miRNAs that failed to convert

- Add: Number of miRNAs added during conversion

- Total: Total number of miRNAs after conversion

4. Save Conversion Results: Click the 'Save changes' button to save the converted data. Clicking 'Cancel' or 'X' will close the window without saving.

1.2 Input and Transform mRNA Data



1. Enter mRNA List: Select the identifiers for the mRNA you wish to analyze. Supported identifiers include Gene Symbol, Entrez Gene ID (default), Ensembl Gene ID, Ensembl Transcript ID, and RefSeq. Then, input the mRNA list or upload a file according to the selected identifier format, using commas (,), line breaks (Enter), or spaces as delimiters.

- **Note:** If using Gene Symbol, please ensure you use standardized symbols.

2. Click the Check Button: Press the 'Check' button to convert the entered mRNA.

3. Review Conversion Results: The conversion results will be displayed in a new window. IDs that fail to convert will be marked as 'NA' and will not be included in the analysis.

4. Save Conversion Results: Click the 'Save changes' button to save the converted data. Clicking 'Cancel' or 'X' will close the window without saving.

1.3 Select miRNA-Target Database

Users can choose a miRNA-target database from the following options:

Option1. Validated DB

These databases provide experimentally validated miRNA-target interaction data and are suitable for researchers who wish to analyze only highly reliable interactions.

- **DIANA-TarBase:** This database contains 422,893 experimentally validated miRNA-mRNA interactions and provides the most comprehensive collection of validated interactions. It is suitable for researchers who want to base their analysis on experimentally confirmed interactions.
- **miRTarBase:** The second largest database, curated from approximately 11,000 research papers, providing 380,640 interactions. It is ideal for researchers who want to analyze publicly available, highly reliable interactions.
- **Union of DIANA-TarBase and miRTarBase:** This option includes all interactions provided by both databases, making it suitable for researchers who want to maximize their analysis of reliable data by combining the strengths of both databases.
- **Intersection of DIANA-TarBase and miRTarBase:** This option includes only those interactions validated by both databases, allowing for a more conservative analysis of highly reliable interactions. It is recommended for researchers who prioritize reliability while maintaining a balanced amount of data.

Option2. Predicted DB

These databases are suitable for researchers who want to explore large numbers of predicted interactions and investigate potential new miRNA-target interactions. Predicted data may contain false positives, so caution is advised.

- **miRDB:** This prediction-based database integrates data from TargetScan, PicTar, miRanda, and others using an SVM algorithm and provides 3,375,741 interactions. Although it provides the largest set of predicted interactions, there may be a significant number of false positives that require careful interpretation.
- **TargetScan:** Based on sequence comparison between miRNAs and mRNAs, the TargetScan algorithm predicts 10,165,094 interactions. It is useful for researchers trying to predict novel miRNA-mRNA interactions.
- **Intersection of miRDB and TargetScan:** This option includes only those interactions commonly predicted by both databases, balancing prediction reliability with broad coverage of novel interactions.

Option3. All DB

- **Intersection of DIANA-TarBase, miRTarBase, miRDB and TargetScan:** This option includes only interactions found in all four databases, offering the most conservative analysis of highly reliable interactions. It comprises 3,669 intersections involving 365 miRNAs and 1,098 genes.

Note: Users should select the appropriate database according to their research goals, data reliability, and scope of analysis.

1.4 Set Algorithm Parameters

1. **Restart Probability: Set the algorithm's restart probability (default: 0.9).**
   Restart probability determines the likelihood of returning to the starting point during a random walk in the RWR (Random Walk with Restart) algorithm. A higher value makes the algorithm focus more on genes near the seed gene (defined as the intersection of the user's miRNA target genes and mRNA).

   - **High values (0.9 ~ 0.7):** The algorithm prioritizes genes directly connected to the seed gene. This is useful when the researcher wants to explore genes closely related to the seed gene.
   - **Low values (0.7 ~ 0.5):** The algorithm allows broader propagation across the network, exploring genes farther from the seed gene. This is suitable for researchers investigating more distant connections in the network.

2. **Boosting Probability: Set the boosting probability (default: 0.1).**
   Boosting probability determines the threshold for genes from the RWR results. It determines how many genes are added to the network to check for direct or indirect interactions with the seed gene and is used in pathway and ontology enrichment analysis.

   - **Default value (0.1):** A good starting point, selecting a balanced range of genes connected directly or indirectly to the seed gene. It provides detailed pathway analysis results.
   - **Low value (below 0.1):** Focuses on genes directly connected to the seed gene, reducing the analysis scope but minimizing false positives and improving precision.
   - **High value (above 0.1):** Includes more genes indirectly linked to the seed gene. This widens the analysis but may increase false positives, so caution is needed when interpreting results. Useful for visualizing broader networks.

3. **STRING Score Cut-off: Set the cut-off value for STRING Score (default: 0.01).**
   The RWR algorithm uses protein-protein interaction (PPI) data from the STRING database, where each interaction is assigned a confidence score (STRING score). Higher scores indicate more reliable interactions. The cut-off value determines the size and reliability of the network for RWR analysis.

   - **High cut-off value (default 0.01):** Only highly reliable interactions (STRING score ≥ 945) are included. This is ideal for researchers needing high-confidence networks.
   - **Low cut-off value (0.05, 0.1):** Includes broader networks with lower reliability, using cut-offs of 0.05 (STRING score ≥ 662) and 0.1 (STRING score ≥ 491). This is suitable for researchers exploring new or potential gene interactions.

Note: Users should adjust these parameters based on their research goals. Restart probability, boosting probability, and STRING score cut-off help guide the extent and reliability of network exploration, so choosing values that fit your objectives will enhance your analysis.

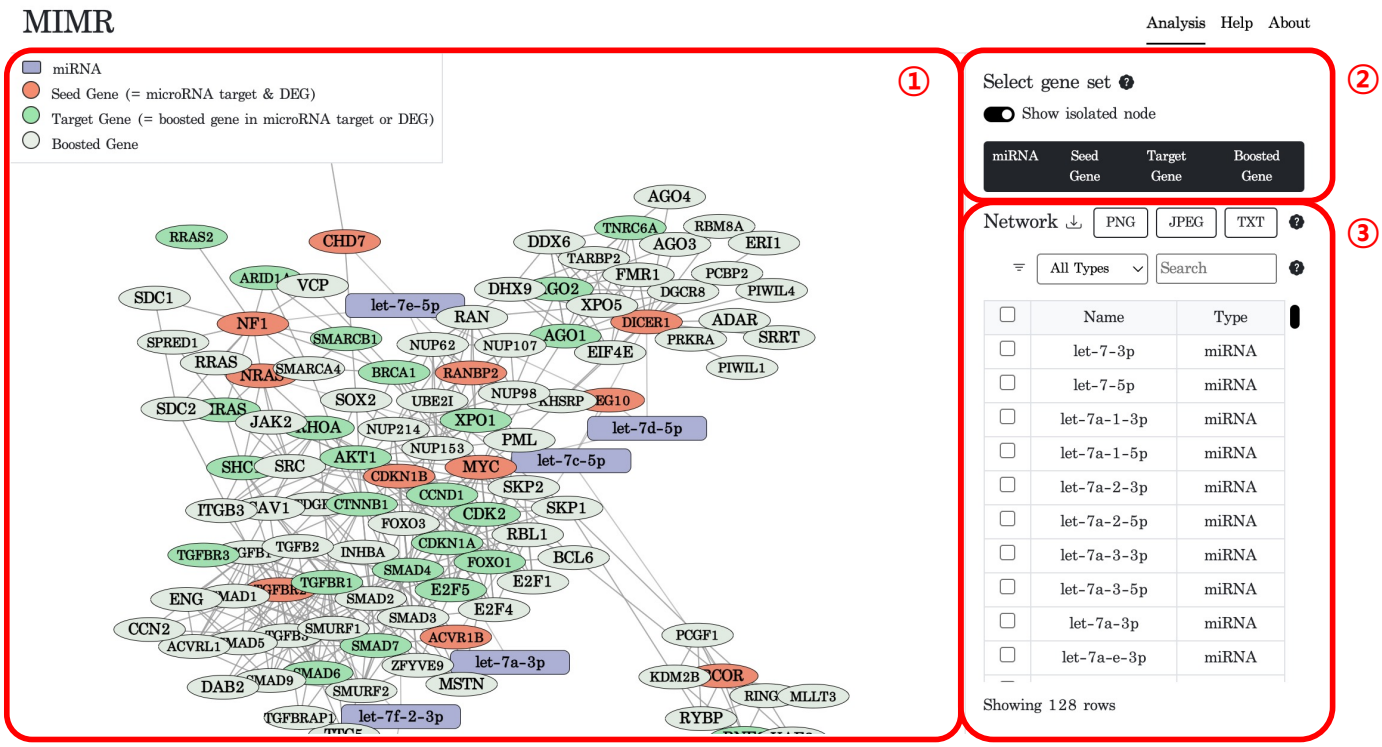
1.5 Execute and Check Results

1. Click the Submit Button: Once you've completed the settings, press the Submit button to run the analysis.

2. Wait for Results: Depending on the data size, the analysis may take up to 5 minutes to complete. You can copy the URL to check the results later.

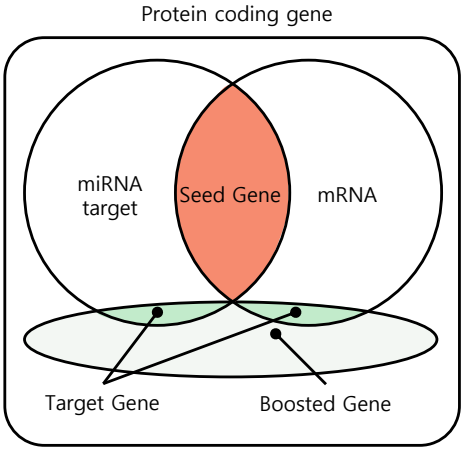
2. Result Tutorial: Network and Enrichment Analysis Visualization

2.1 Network



1. Network Exploration Features

- If you need to reset the network to its original state, you will need to refresh the page.
- Users can freely move the network on the screen using the mouse and zoom in or out using the mouse wheel.
- You can click and drag the nodes to place them in your preferred location.
- The network includes various node types such as miRNA, miRNA target, Seed Gene, Target Gene, and Boosted Gene.

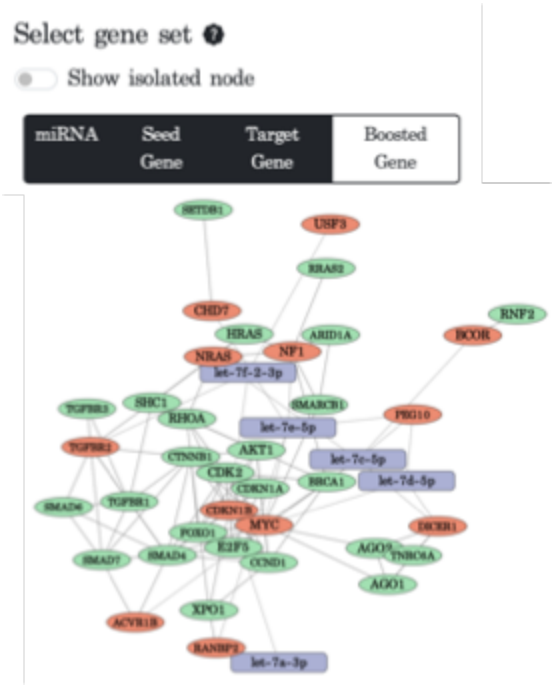


- **Seed Gene** represents genes that are common between miRNA targets and mRNA.
- Genes that are expanded based on RWR (Random Walk with Restart) and classified as either miRNA targets or mRNAs are labeled as **Target Genes**, while genes that do not belong to these categories are labeled as **Boosted Genes.**

2. Network Node Display Options:

- Initially, all miRNA and PPI (protein-protein interaction) related gene nodes are displayed.
- Gene nodes without interactions may not be displayed if they do not interact with any input genes.
- You can select or hide specific node types using the options in the top-right corner.
- If there are any nodes without interaction edges, they are displayed as Isolated Nodes, and you have the option to show or hide these nodes as needed.

Example 1:



The image shows **Isolated Nodes** hidden, with only interactions between **miRNA, Seed Genes, and Target Genes** visible. This functionality helps to reduce network complexity and allows for comparison of the network before and after boosting.

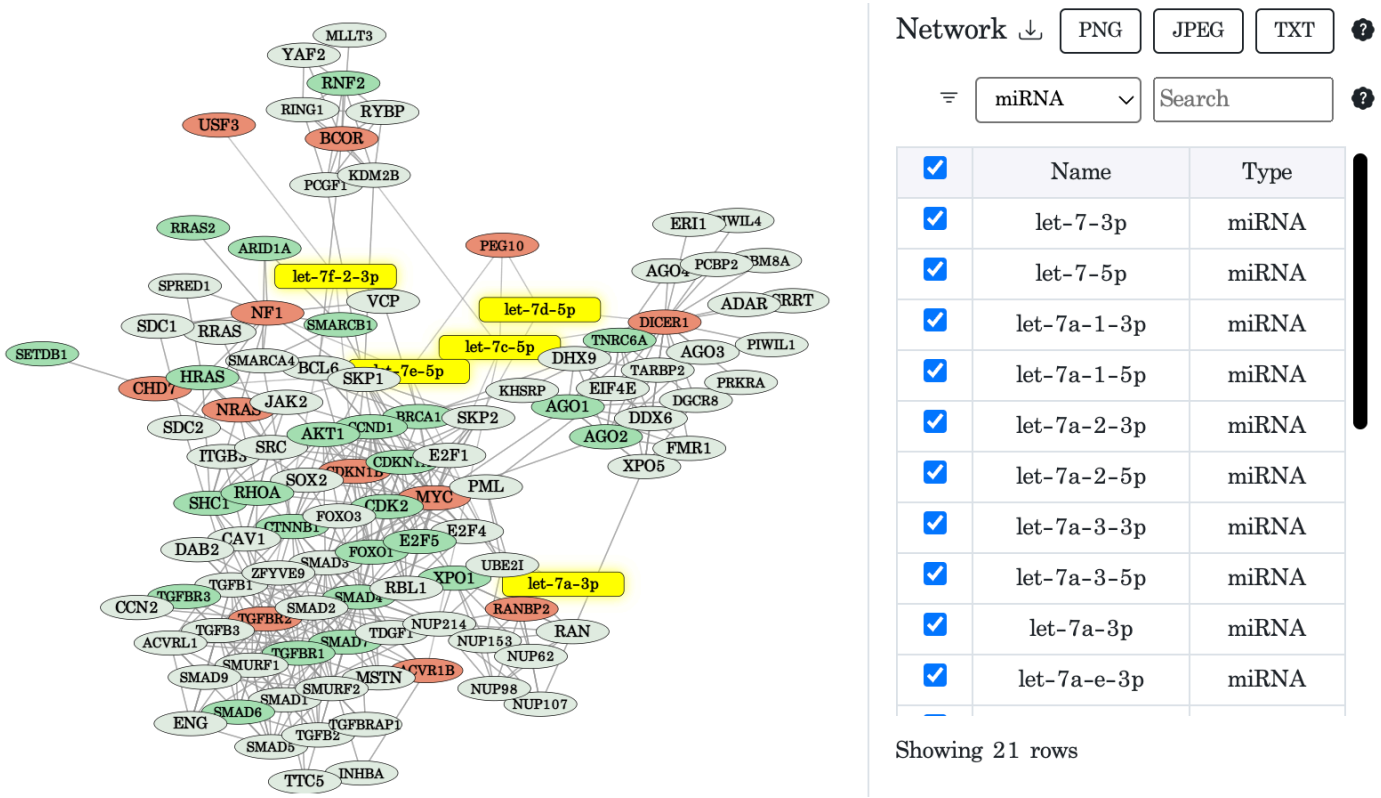
3. Image and Data Download:

- The currently displayed network can be downloaded as a PNG or JPEG image file.
- Additionally, node.txt and edge.txt files can be downloaded, which are compatible with Cytoscape. Users can further edit the network in Cytoscape with their desired styles.

4. Node Search and Highlight:

- You can search for a node by name using the search box on the right and click the node to highlight it within the network.
- The filtering feature allows you to display only specific node types through search.

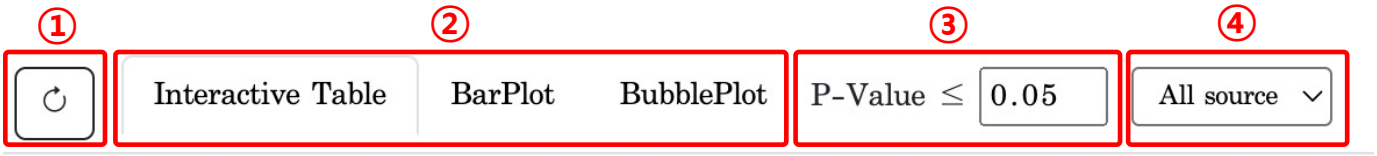
Example 2:



The example shows the selection of all miRNA types via the checkboxes. This feature helps you easily locate desired nodes within complex networks.

Enrichment Analysis Visualization Results

Below are three common visualization formats used to display the results of Enrichment Analysis performed using gProfiler.



1. Refresh Button: Clicking the refresh icon reloads the results on the current screen. If the visualization results are not displayed correctly or you need to reflect updated information, click this button to reload the results.

2. Visualization Type Selection:

- Interactive Table: Displays Enrichment results in a table format, allowing users to check detailed information and explore the results interactively.

- BarPlot: Visualizes the results in a bar plot, making it easy to compare the p-value or enrichment scores of each term.

- BubblePlot: Visualizes terms as bubbles, where their size and color represent the importance and p-value, allowing users to quickly understand the relative significance of each term.

3. p-value Filter Setting:

- The p-value filter is set to ≤ 0.05 by default, allowing users to filter only statistically significant terms. You can change this value to reflect your desired significance threshold.

4. Data Source Selection:

- Users can select specific databases or sources from the **All source** dropdown to filter results. This allows you to focus on results from specific databases or data sources.

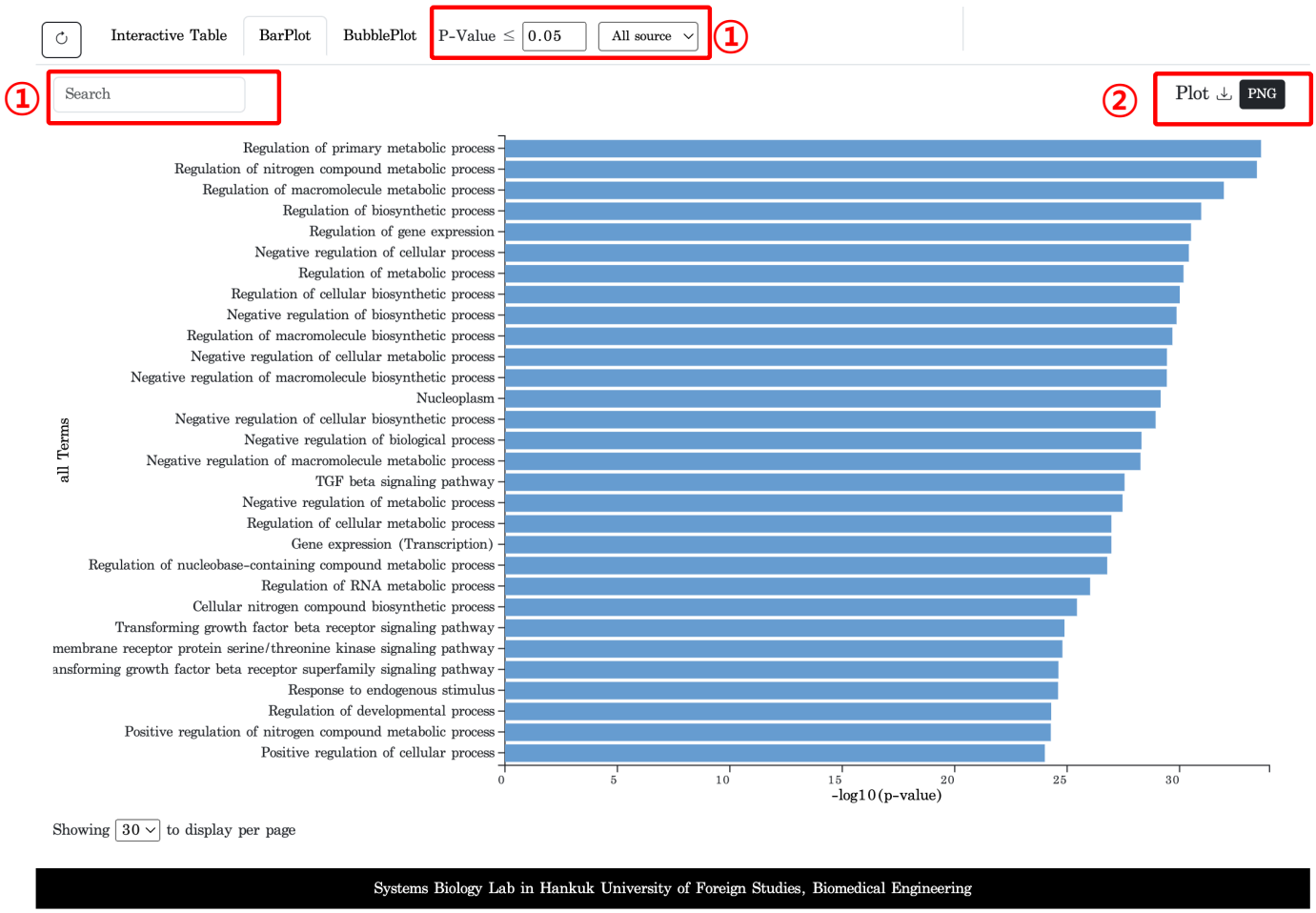
2.2 Table Format Results

1. Users can view pathway enrichment results in a table format and toggle between the three result types using a button.

   - Highlighted rows are newly significant, revealed only after boosting, and not seen with seed genes alone.
   - Target genes are miRNA targets or mRNAs among boosted genes.

2. The Seed Genes, Boosted Target Genes and Boosting results can each be downloaded in TSV file format. We obtain the full results, set the p-value down to 0.05 and select 'All' for the source.

2.3 Bar Plot

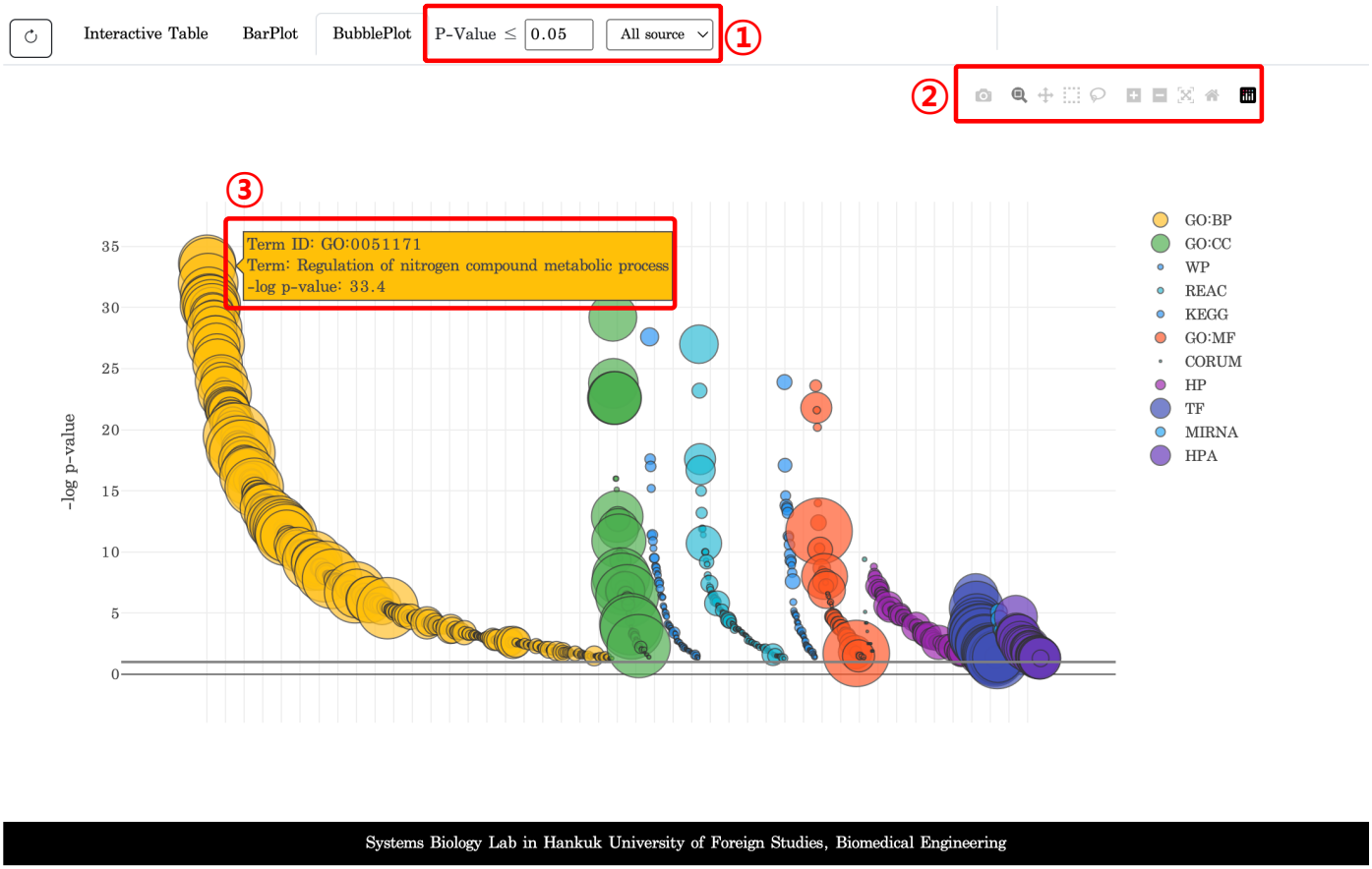


1. Only Boosting analysis results are displayed, and visualization follows the selected p-value and source conditions.

2. You can search for keyword of terms by name in the search box at the top of the bar plot and export the current graph as a PNG file.

Note: Even if the setting is to display 30 terms per page, only terms that match the selected p-value, source, and search conditions will be shown. As a result, fewer terms may be displayed if the conditions are too restrictive.

2.4 Bubble Plot



1. The Bubble Plot visualizes the relative importance of terms based on their p-value.

2. Users can zoom in, zoom out, and move the view for easier exploration. The graph can also be downloaded as a PNG file.

3. Hovering over a bubble reveals detailed information about the term, including its Term ID, name, and -log(p-value), providing a deeper understanding of the specific terms.