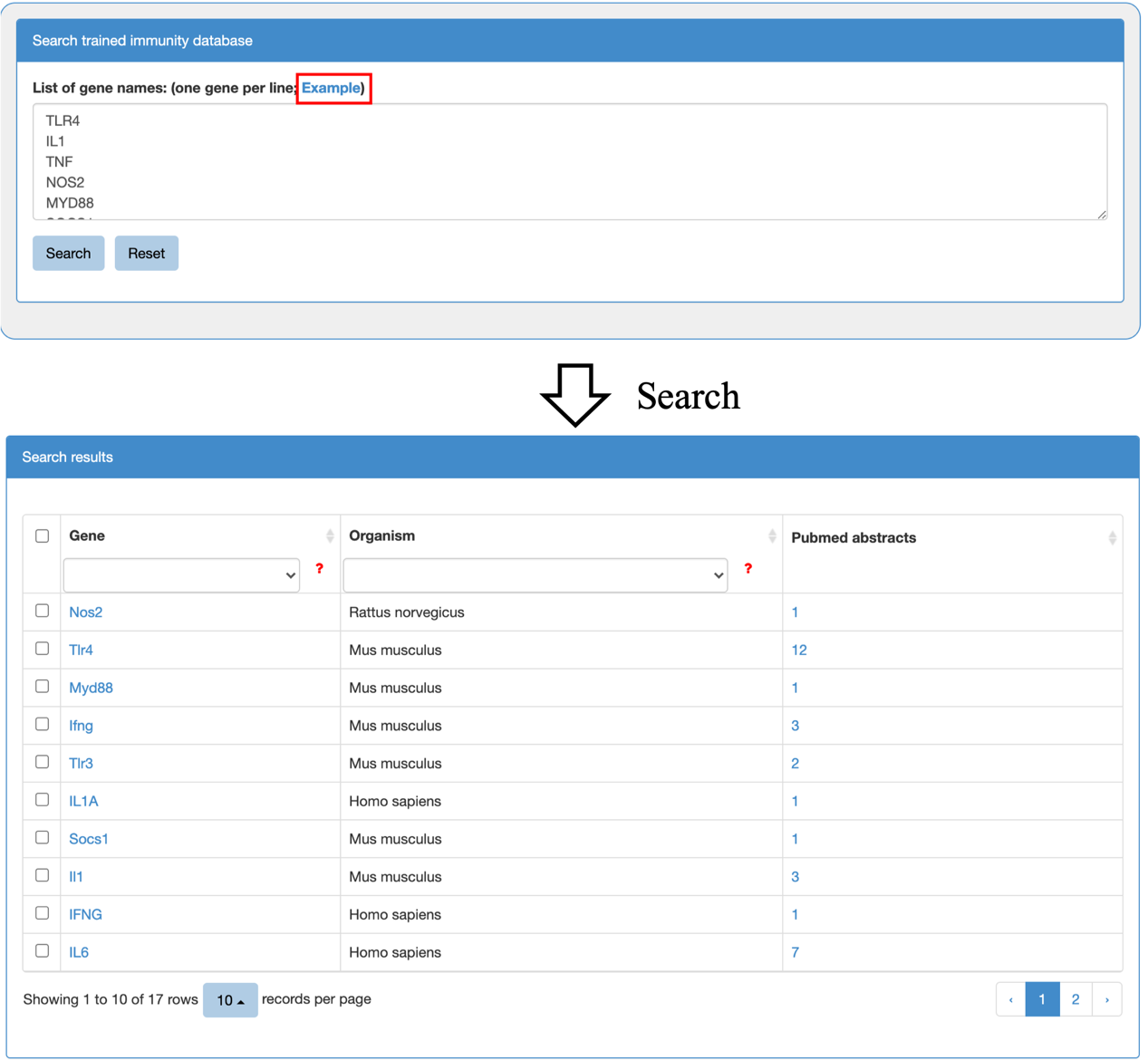
# 1. What is the TIDB?

The TIDB refers to “**T**rained **I**mmunity **D**ata**B**ase”. It contains a comprehensive list set of well-annotated trained immunity-related genes with their correspondent evidences in literature, derived by text mining and manual curation. Moreover, **TIDB** also provides three modules to analyze genes of interest, including Reactome pathway over-representation analysis, GO enrichment analysis and PPI network reconstruction.

# 2. How can I search trained immunity-related genes and their supporting literature evidence in the database?

You can input a single gene name/symbol or a list of gene names/symbols (one gene per line) to launch the search process in the database. For a quick trial, you can just click the **example** displayed above the input box to view a sample list of gene symbols.

After clicking the “Search” button, the search engine will run and return the queried results as a three columns table, gene symbol, organism and the supporting literature evidence.



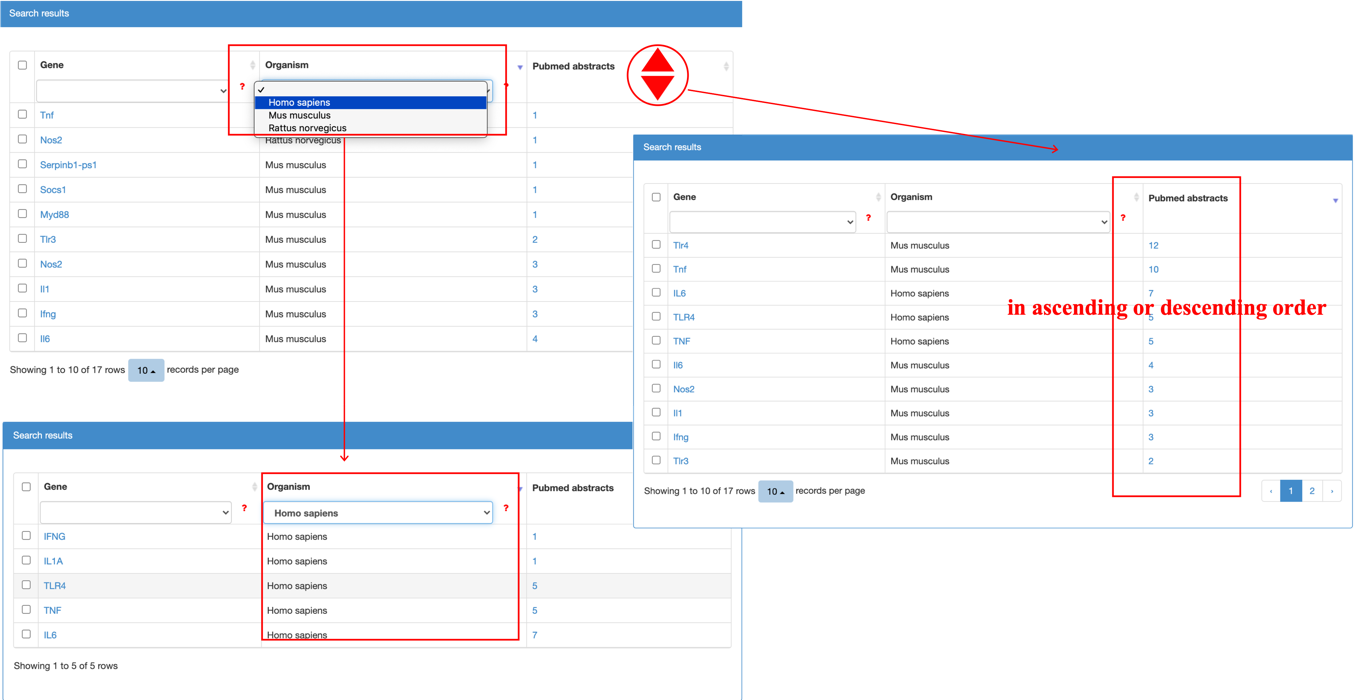
(search.png)

By clicking the hyperlink of the gene symbol in the “Gene column”, you can see the basic information of the gene and the cross references to external databases (i.e. NCBI Entrez Gene, Ensemble).



(gene\_information.png)

The result table will be resorted in ascending/descending order by the column by clicking the small triangle in the head of each column. And you can select a term in the drop list displayed below the column header to filter the listed genes or organisms from the results.



(result\_filter.png)

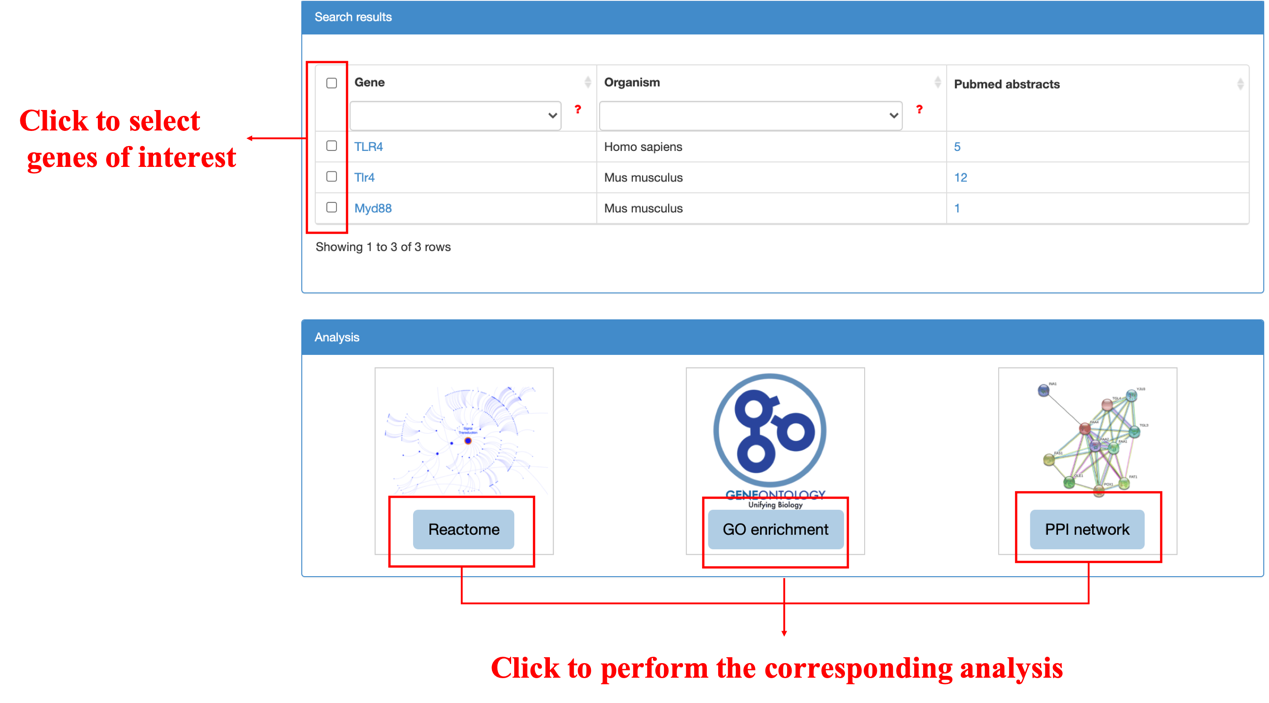
When you click the hyperlink of the number of the evidence abstracts in the “Pubmed abstracts” column, a table containing gene name, gene symbol, organism, PubMed ID, evidence and manual validation will be displayed. Then, you can click on the “Evidence” column to see the original abstract in which the keywords, such as gene names, are highlighted.

# 

(pubmed\_highlight.png)

# 3. How can I perform downstream analysis on searched genes of interest?

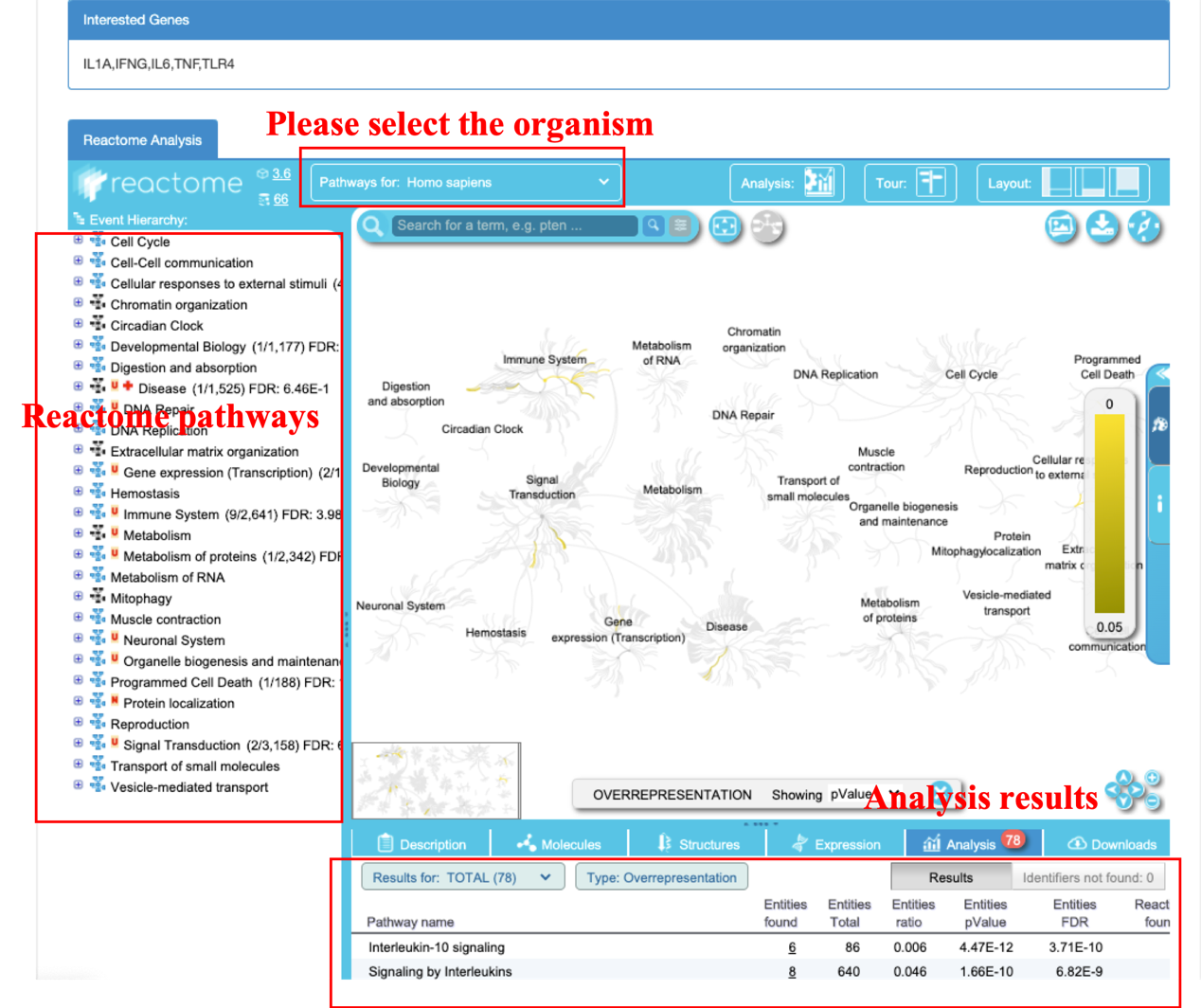
Once the search process is completed. You can select the genes of interest in the search results to perform three downstream analysis, including Reactome pathway over-representation analysis, GO enrichment analysis and PPI network reconstruction. These three analysis modules displayed below the search results, and you can click the three buttons to perform the corresponding analysis. **Please note to select the genes of the same organism**.



(analysis\_modules.png)

**(1) Reactome pathway over-representation analysis**

**TIDB** integrates the pathway over-representation analysis via Reactome web service. Once you select the genes of interest and click the “Reactome” button, it will then perform Reactome over-representation analysis and pop up a new page to show the results.



(Reactome\_result.png)

Analysis results are shown in the Analysis tab, within the Details Panel. All Reactome pathways are shown, in blocks of 20 pathways, ranked by the p-value obtained from over-representation analysis. If multiple pathways have the same p-value, they are ranked by the number of identifiers in the query that match the pathway. The number of molecules matched/total number of molecules and FDR values are added to the right side of pathway names in the Hierarchy Panel. The names of reactions that match at least one identifier in the query, representing positive pathway topology analysis hits, are boxed in orange.

In the Analysis tab, clicking on the name of a pathway will select it in the Hierarchy, which if neccessary will expand hidden hierarchical levels to show the pathway, while the name becomes highlighted in dark blue.

The columns in analysis details represent:

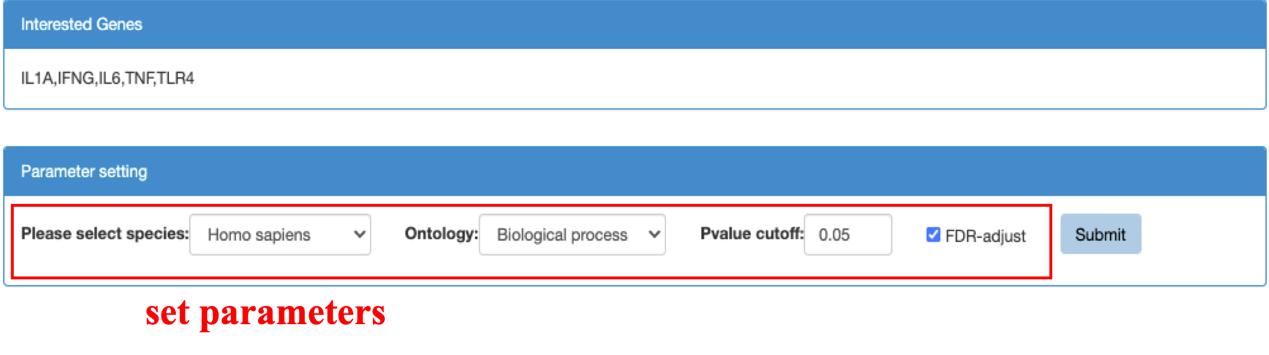
1. Pathway name: Click the name to open the pathway.
2. Entities found: the number of curated molecules of the type selected with Results Type that are common between the submitted data set and the pathway named in column 1. Click on this number to display the matched submitted identifiers and their mapping to Reactome molecules.
3. Entities total: The total number of curated molecules of the type selected with Results Type within the pathway named in column 1.
4. Interactors found (if this option was selected). The number of interactor molecules of the type selected with Results Type that are common between the submitted data set and the pathway named in column 1. Click on this number to display the matched submitted identifiers and their mapping to Reactome molecules.
5. Interactors total (if this option was selected): The total number of interactor molecules of the type selected with Results Type within the pathway named in column 1.
6. Entities ratio: Put simply, the proportion of Reactome pathway molecules represented by this pathway. Calculated as the ratio of entities from this pathway that are molecules of the type selected with Results Type Vs. all entities of the type selected with Results Type.
7. Entities pvalue: The result of the statistical test for over-representation, for molecules of the results type selected.
8. Entities FDR: False discovery rate. Corrected over-representation probability.
9. Reactions found: The number of reactions in the pathway that are represented by at least one molecule in the submitted data set, for the molecule type selected with Results Type.
10. Reactions Total: The number of reactions in the pathway that contain molecules of the type selected with Results Type.
11. Reactions ratio: Put simply, the proportion of Reactome reactions represented by this pathway. Calculated as the ratio of reactions from this pathway that contain molecules of the type selected with Results Type Vs. all Reactome reactions that contain molecules of the type selected with Results Type.
12. Species Name.

For more details see the [tutorial of analysis tools of Reactome](https://reactome.org/userguide/analysis) (超链接).

**(2) GO enrichment analysis**

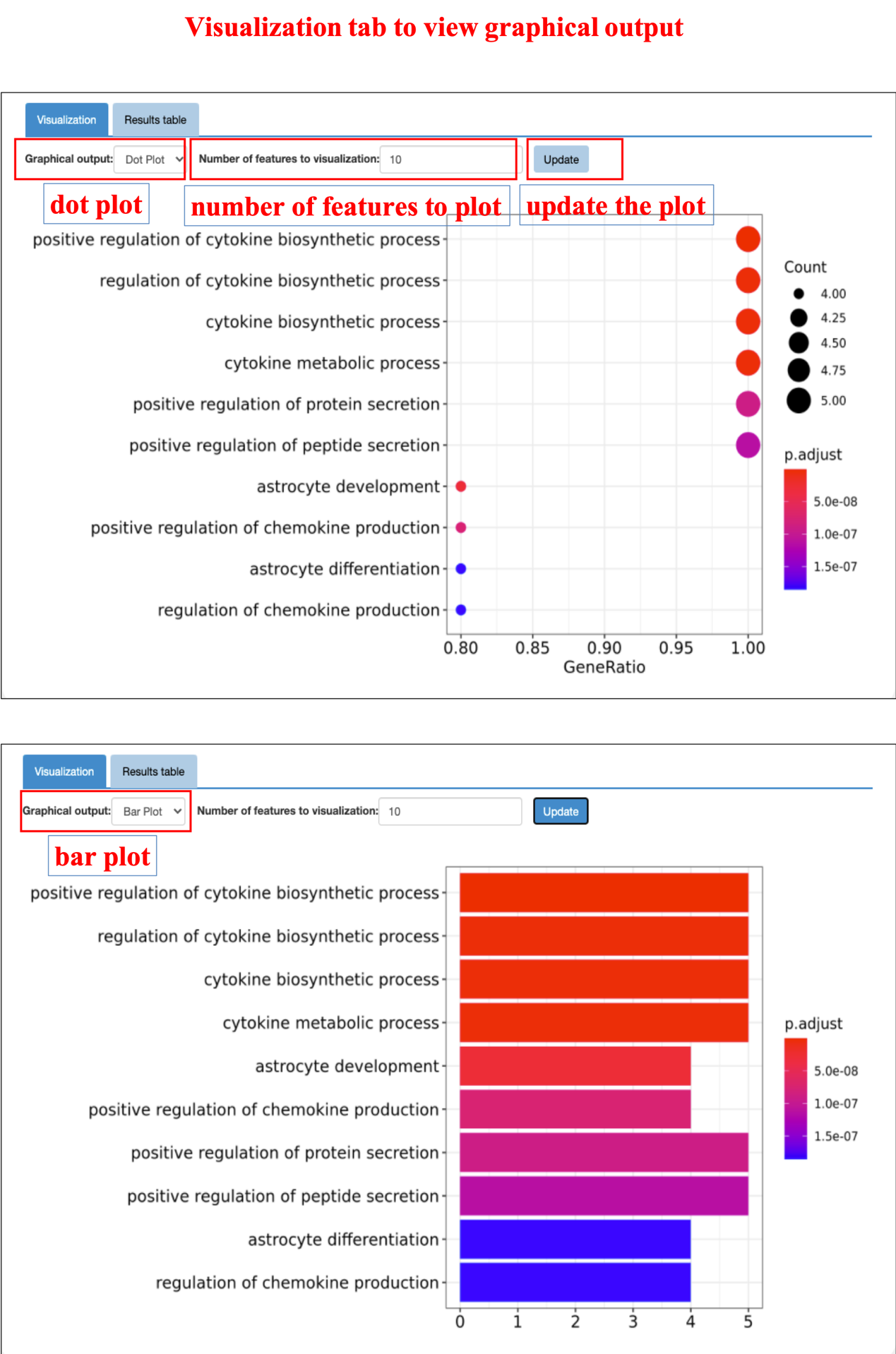
TIDB utilities R package [clusterProfiler](http://www.bioconductor.org/packages/release/bioc/html/clusterProfiler.html) (超链接) to run perform enrichment analysis. Once you select the genes of interest and click the “GO enrichment” button, it will pop up a new page for performing GO enrichment analysis. You have to set the following parameters first:

1. the species name.
2. which biological domain to choose, one of biological process, molecular function, cellular component or all.
3. pvalue cutoff on enrichment tests to report
4. whether adjust the pvalue using fdr method or not



(GO\_para.png)

Then click the “submit” button to start the analysis. Once the analysis finished, the results are displayed on the results panel. There are two tabs to show the results, Visualization tab to view the graphical output and Results table tab to view the analysis details. Two types of visualization (dot plot and bar plot) were supported, and you can specify the number of features to show in the plot.



(GO\_vis.png)

The Result table tab shows the details of GO enrichment analysis, and you can download the table by clicking “Export details as csv”. The columns in analysis details represent:

1. GO ID: GO ID, click to open the Gene Ontology

2. GO description: the description of Gene Ontology

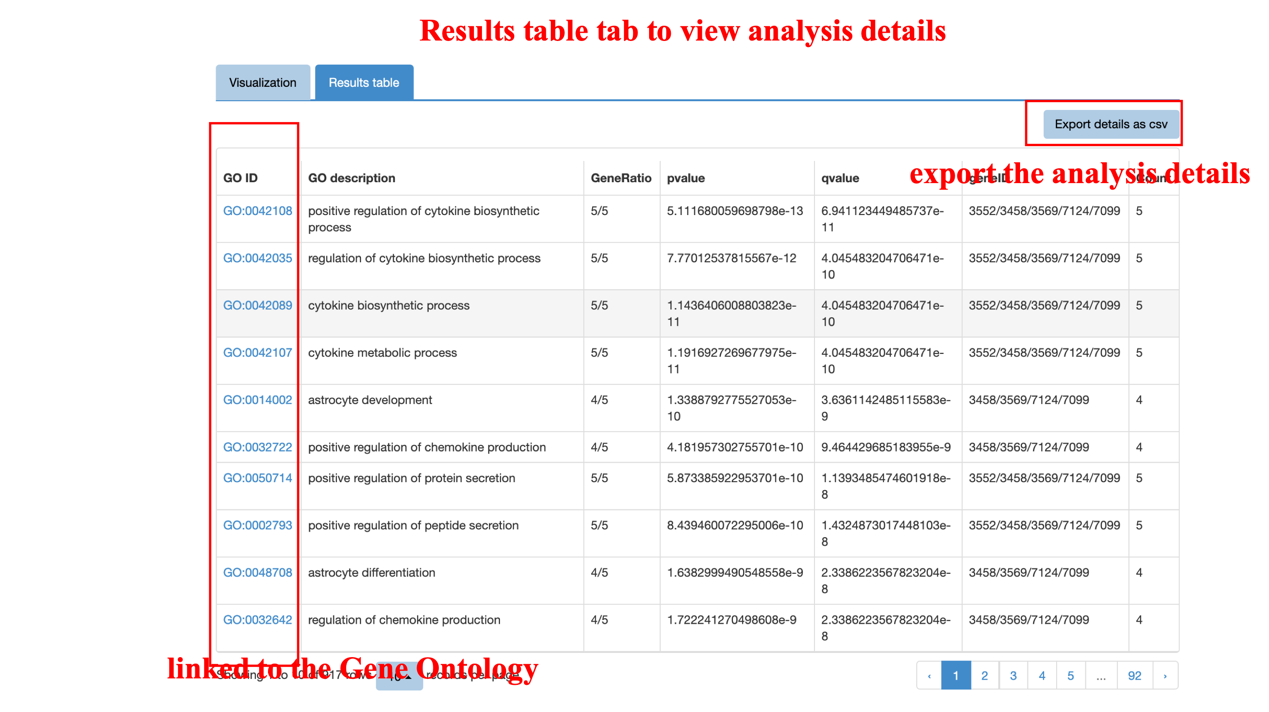
3. GeneRatio: (number of genes among the interest genes mapping to the GO term)/(the total number of genes mapping the GO term)

4. pvalue

5. qvalue

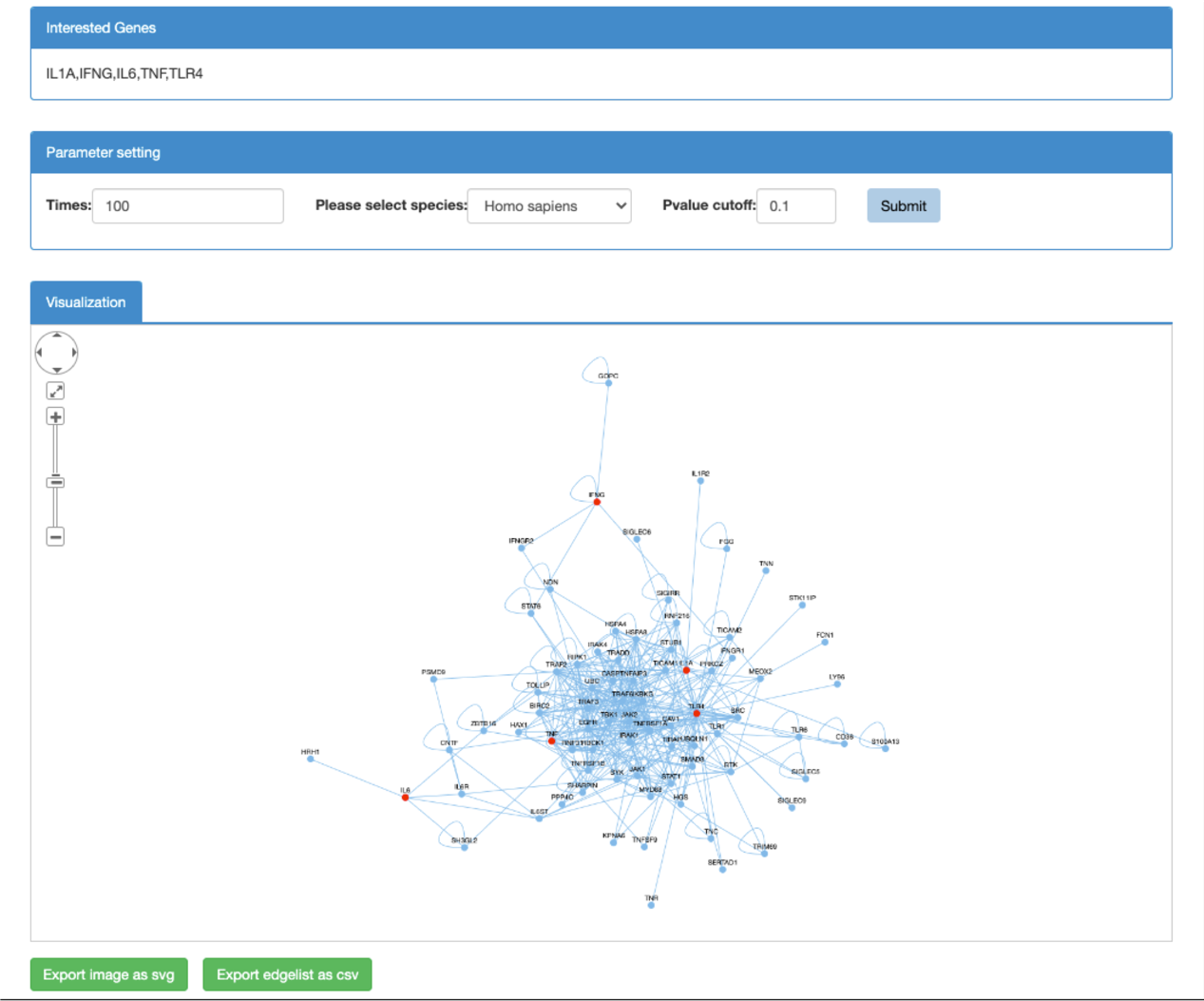
6. GeneID: Entrez gene ID among the interest genes mapped to the GO term

7. Count: number of genes among the interest genes mapped to the GO term



(GO\_table.png)

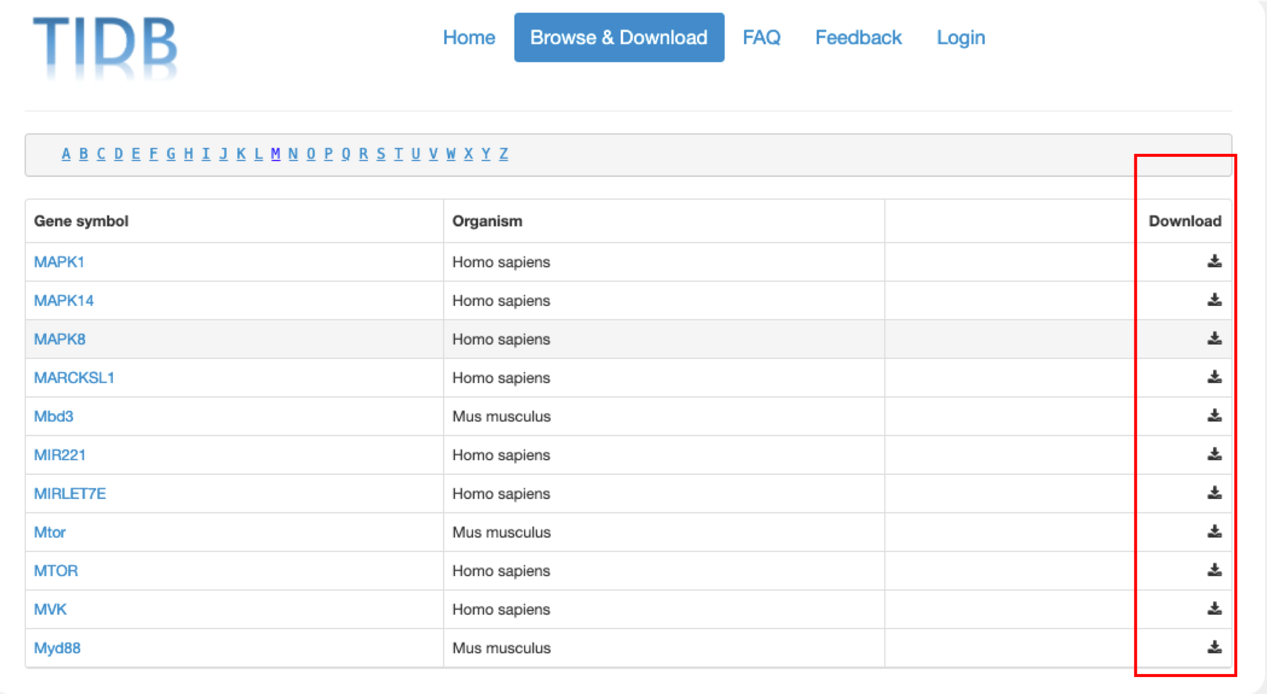
**(3) PPI network reconstruction**

****

**(PPI\_reconstruction.png)**

# 4. How can I browse and download the data?

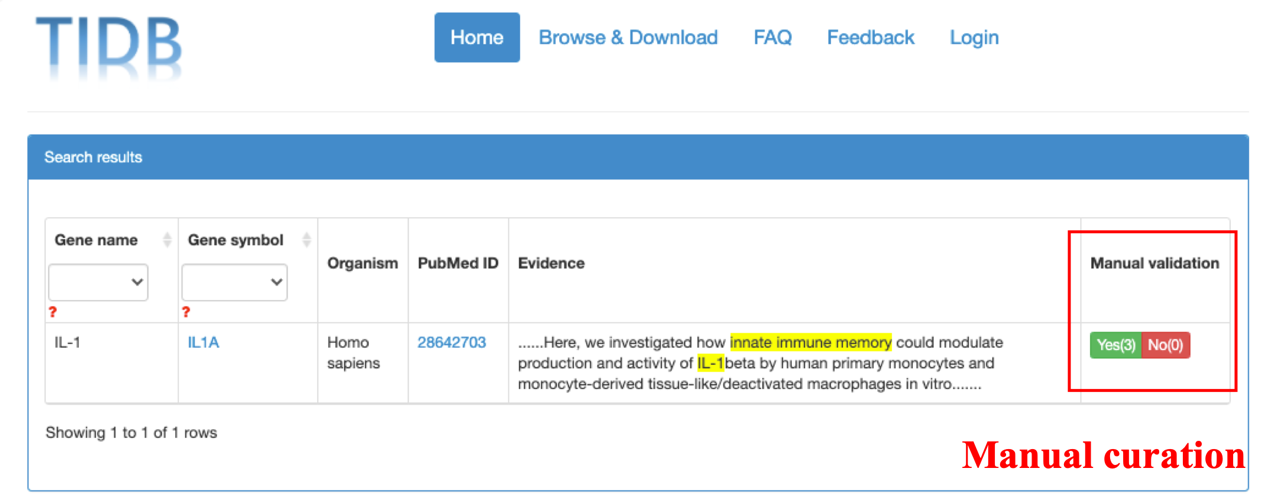
The complete data in **TIDB** can be browsed by clicking the ‘Browse & Download’ button in the navigation bar. All the information for trained immunity-related genes and their supporting literature evidences can be downloaded for further analysis.



(data\_download.png)

# 5. How can I guarantee data quality in the database?

To address this issue, we have taken three rounds of strict manual curation for the data collection in TIDB. First, all sentences with trained immunity-related gene names were checked and selected by two experienced researchers independently; second, these selected sentences were submitted to an internal review, in which all gene names were manually reviewed and approved one by one by a reviewer panel consisting of three experts; third, we asked all co-authors to randomly check the data from our website make sure that all genes loaded into our database are correct. All the evidence sentences that have been manually curated are regarded as validated evidence. In addition, we added a manual curation function to the evidence phrases, with which you can provide their feedback by simply clicking the "Yes" or "No" button to confirm or reject the evidence phrases. IL-1 gene of Homo sapiens is used as an example in the figure shown below:



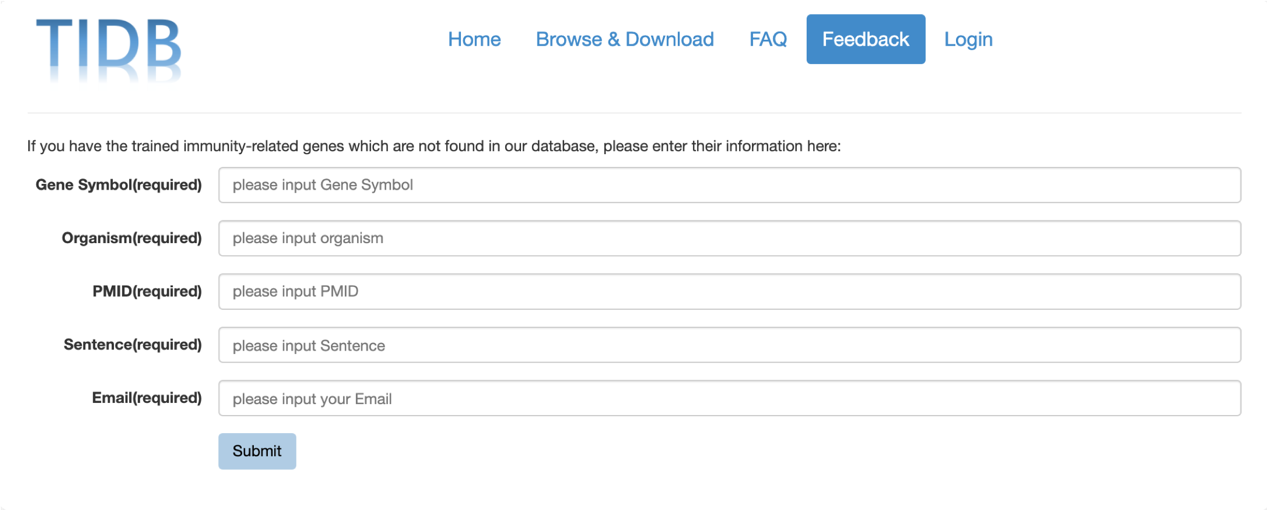
(data\_curation.png)

# 6. Can I provide feedback to the evidence of trained immunity-related genes?

Yes, as described above, we provide a manual annotation function at the end of each evidence phrase, by which you can confirm the evidence or deny it by simply clicking “Yes” or “No”. We will update our database periodically to include the feedback from the user.

In addition, we added “Feedback” feature to our website, with which you can submit new genes to our database manually; the database will be updated periodically in future.

You can also send email to us for further questions or potential collaborations.



(feedback.png)

# 7. Do I need to log in for manual curation?

Yes, the login is necessary for you to do the manual curation or download data from our database. We will monitor and validate community curation feedback before inclusion in the database.