**User Manual & Case studies**

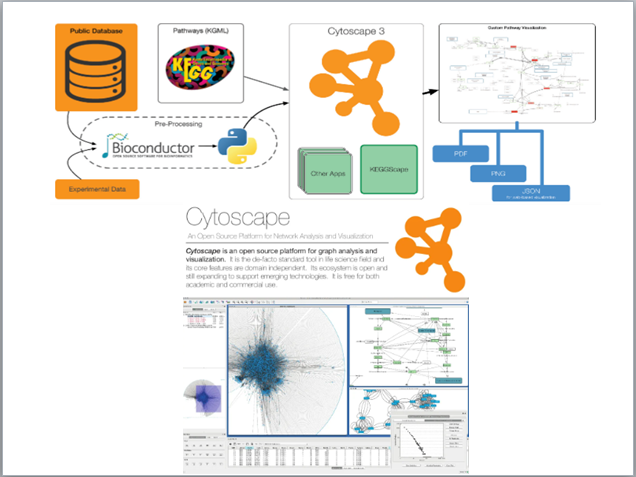


Table of Contents

[User Manual & Case studies 1](#_Toc109806001)

[I. Setup 5](#_Toc109806002)

[1. Install Cytoscape 5](#_Toc109806003)

[2. Install C-Biomarker.net 9](#_Toc109806004)

[3. Install KEGGscape for reading network files from the KEGG database 12](#_Toc109806005)

[II. Overview of C-Biomarker.net 13](#_Toc109806006)

[III. Case study: Identification of biomarker genes of 17 cancer types from KEGG pathways 13](#_Toc109806007)

[Step 1: Load networks 14](#_Toc109806008)

[Step 2: Find biomarker nodes from the loading network. 18](#_Toc109806009)

[Step 3: Rank candidate biomarker genes 19](#_Toc109806010)

[Step 4: Search evidence from PubMed 22](#_Toc109806011)

[Step 5: Biological function analysis 26](#_Toc109806012)

[IV. Case study: Identification of biomarker genes from a large human signaling network 40](#_Toc109806013)

[Step 1: Load networks 40](#_Toc109806014)

[Step 2: Find biomarker nodes from the loading network. 43](#_Toc109806015)

[Step 3: Rank candidate biomarker genes 44](#_Toc109806016)

[Step 4: Search evidence from PubMed 45](#_Toc109806017)

[Step 5: Biological function analysis 46](#_Toc109806018)

[V. Reference 47](#_Toc109806019)

**List of figures**

[Figure 1:Cytoscape additional library installation interface 5](#_Toc109806124)

[Figure 2: Cytoscape installation interface 5](#_Toc109806125)

[Figure 3: License interface 6](#_Toc109806126)

[Figure 4: Interface to choose the path of the installation 6](#_Toc109806127)

[Figure 5: Inferface create a desktop icon 7](#_Toc109806128)

[Figure 6: Interface for creating file associations 7](#_Toc109806129)

[Figure 7: Interface shares user information 8](#_Toc109806130)

[Figure 8: Installation end interface 8](#_Toc109806131)

[Figure 9: Cytoscape software interface 9](#_Toc109806132)

[Figure 10: App Manager interface 9](#_Toc109806133)

[Figure 11: App Installed interface 10](#_Toc109806134)

[Figure 12: Biomarker Nodes tab of C-Biomarker.net application 11](#_Toc109806135)

[Figure 13: Extension tab of C-Biomarker.net application 11](#_Toc109806136)

[Figure 14: Location of KEGGscape in Cytoscape 12](#_Toc109806137)

[Figure 15: Download KGML 13](#_Toc109806138)

[Figure 16: File XML downloaded from KEGG 14](#_Toc109806139)

[Figure 17: Interface to choose file or folder for converting KEGG to text 14](#_Toc109806140)

[Figure 18: Interface of folder to save converted file 15](#_Toc109806141)

[Figure 19: Results screen when running the algorithm 15](#_Toc109806142)

[Figure 20: Result after successful conversion 16](#_Toc109806143)

[Figure 21: Data file loading interface 16](#_Toc109806144)

[Figure 22: Data file loading interface 17](#_Toc109806145)

[Figure 23: Column format interface 17](#_Toc109806146)

[Figure 24: Screen after successful import 18](#_Toc109806147)

[Figure 25: Interface of saving the result 18](#_Toc109806148)

[Figure 26: Result interface when running the algorithm 19](#_Toc109806149)

[Figure 27: Convert from gene ID to gene symbol 20](#_Toc109806150)

[Figure 28: Data file loading interface 22](#_Toc109806151)

[Figure 29:Search Biological function analysis 26](#_Toc109806152)

[Figure 30: Data column format used in C-Biomarker.net 41](#_Toc109806153)

[Figure 31: Data file loading interface 42](#_Toc109806154)

[Figure 32: Network data preview interface 42](#_Toc109806155)

[Figure 33: Column format interface 43](#_Toc109806156)

[Figure 34: Screen after successful import 43](#_Toc109806157)

[Figure 35: Interface to save the result 44](#_Toc109806158)

[Figure 36: Result interface when running the algorithm 44](#_Toc109806159)

[Figure 37: Search evidence from PubMed 45](#_Toc109806160)

[Figure 38: Search Biological function analysis 46](#_Toc109806161)

**List of tables**

[Table 1: Top 3 peaks with the biggest Biomarker results 22](#_Toc109806169)

[Table 2: Top 3 nodes with the biggest Biomarker results 45](#_Toc109806170)

1. **Setup**
2. **Install Cytoscape**

**Step 1**: Check if the computer has java JDK installed, open cmd, and command java –version, if the computer has java installed, the name of the java version will be displayed.

If java is not installed, download and install it :

* https://java.com/en/download/help/windows\_manual\_download.html.

**Step 2**: Download Cytoscape software from the website:

* [https://Cytoscape.org/](https://cytoscape.org/)

**Step 3**: After downloading the software, open the software and install it, when installing, Cytoscape will display a message to download additional supported libraries for it, the interface isin Fig.1:

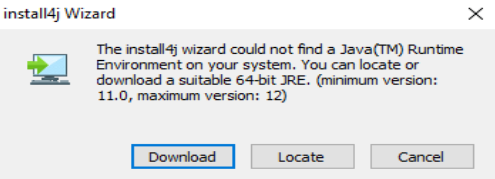


Figure 1:Cytoscape additional library installation interface

Now click Download and wait for the process to complete.

After downloading the supported library, the Cytoscape installation interface will appear as follows:



Figure 2: Cytoscape installation interface

**Step 4**: Click next, and the licensing interface appears as follows:

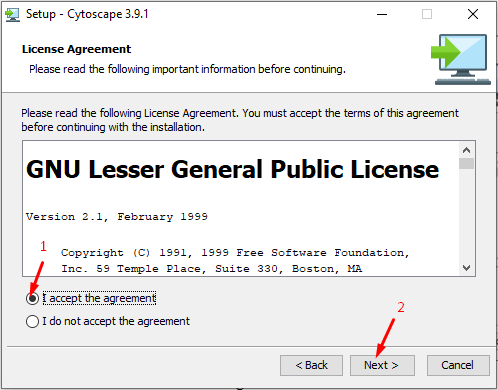


Figure 3: License interface

**Step 5**: Select “I accept the agreement” and click next, the interface selects the installation path, then click Next.

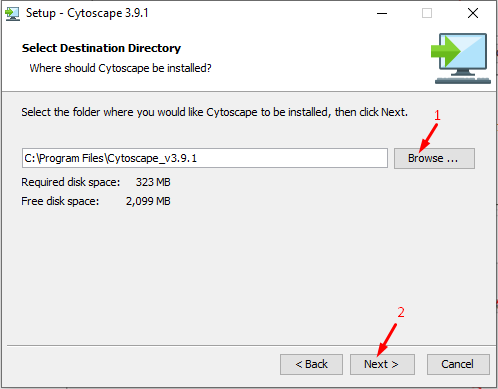


Figure 4: Interface to choose the path of the installation

**Step 6**: The desktop shortcut creation interface appears, if you want to create a software shortcut on the desktop, check the checkbox, otherwise uncheck, then click next.

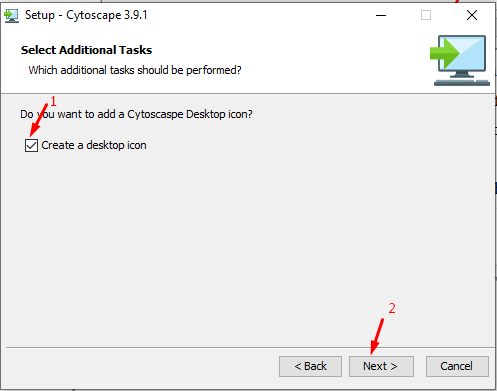


Figure 5: Inferface create a desktop icon

**Step 7**: The interface to create file associations appears, then click next.

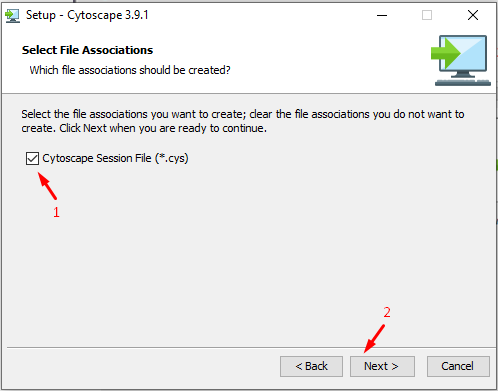


Figure 6: Interface for creating file associations

**Step 8**: The shared usage information interface appears, if you want to share user information, then check the check box, otherwise uncheck and click next. Wait for the installation process.

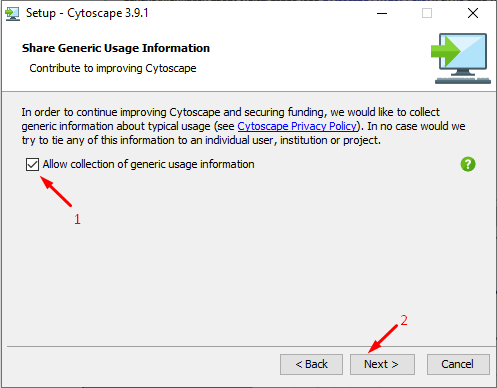


Figure 7: Interface shares user information

**Step 9**: After the installation is successful, the interface will appear as

follows:

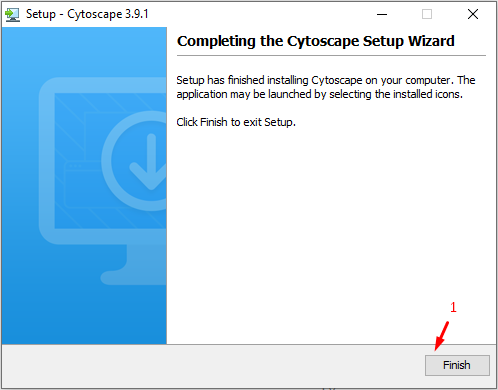


Figure 8: Installation end interface

1. **Install C-Biomarker.net**

Include 2 ways:

* **The first way: installing from a file**

After installing Cytoscape, open the software to the interface when the software has opened up as follows:

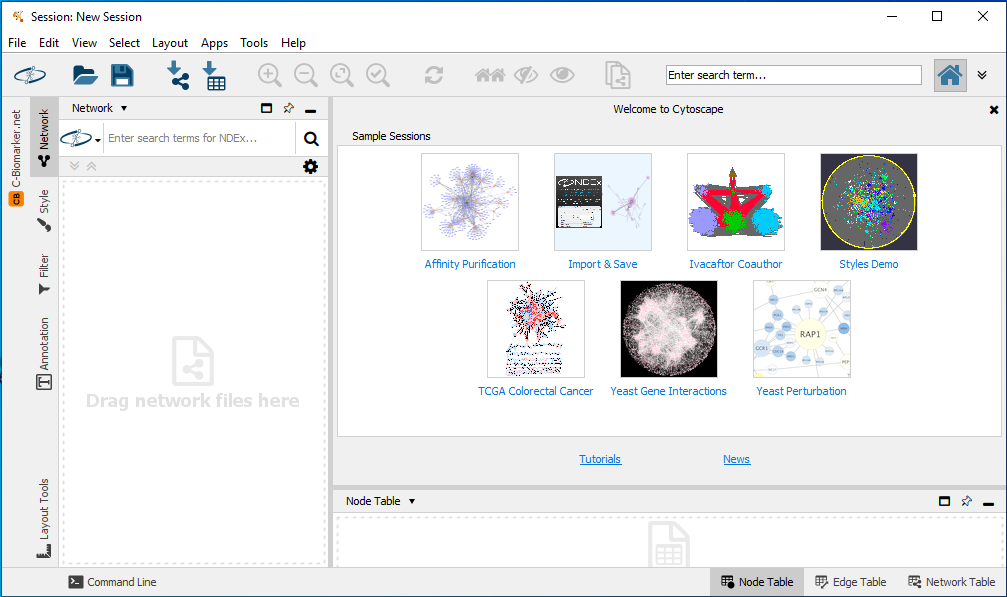


Figure 9: Cytoscape software interface

To install the app into Cytoscape, at the software menu select Apps, then select App Manager, the App Manager interface appears:

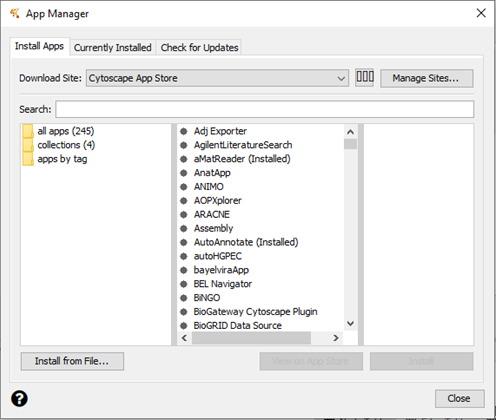


Figure 10: App Manager interface

Click “Install from file” then select the path to the plugin file and then click open. After successful installation, the Currently Installed interface will appear as follows:

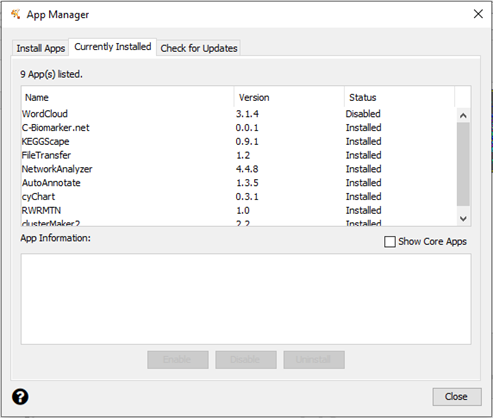


Figure 11: App Installed interface

* **The second way: installing from AppStore**

Install automatically from the Cytoscape AppStore: Choose the menu Apps → App Manager in Cytoscape. Then type C-Biomarker.net in the search box click on C-Biomarker.net in area two and then click on the Install button. The installation is successful.

* **Briefly introduce the used way of the software**

After successfully installing the application, in the vertical menu of Cytoscape, a vertical menu titled "C-Biomarker.net" will appear. Click on the vertical menu, the application interface will appear as follows:

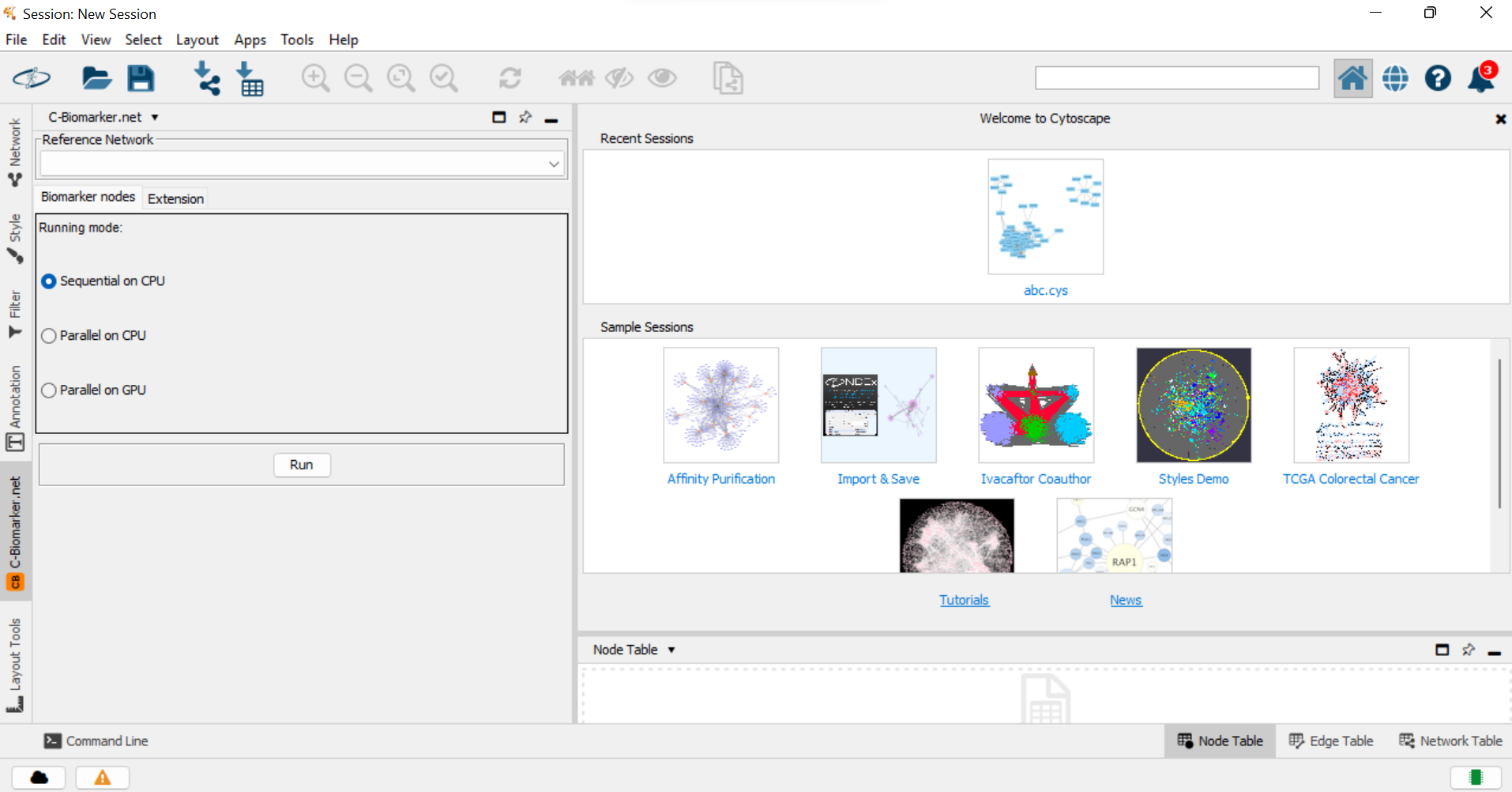


Figure 12: Biomarker Nodes tab of C-Biomarker.net application

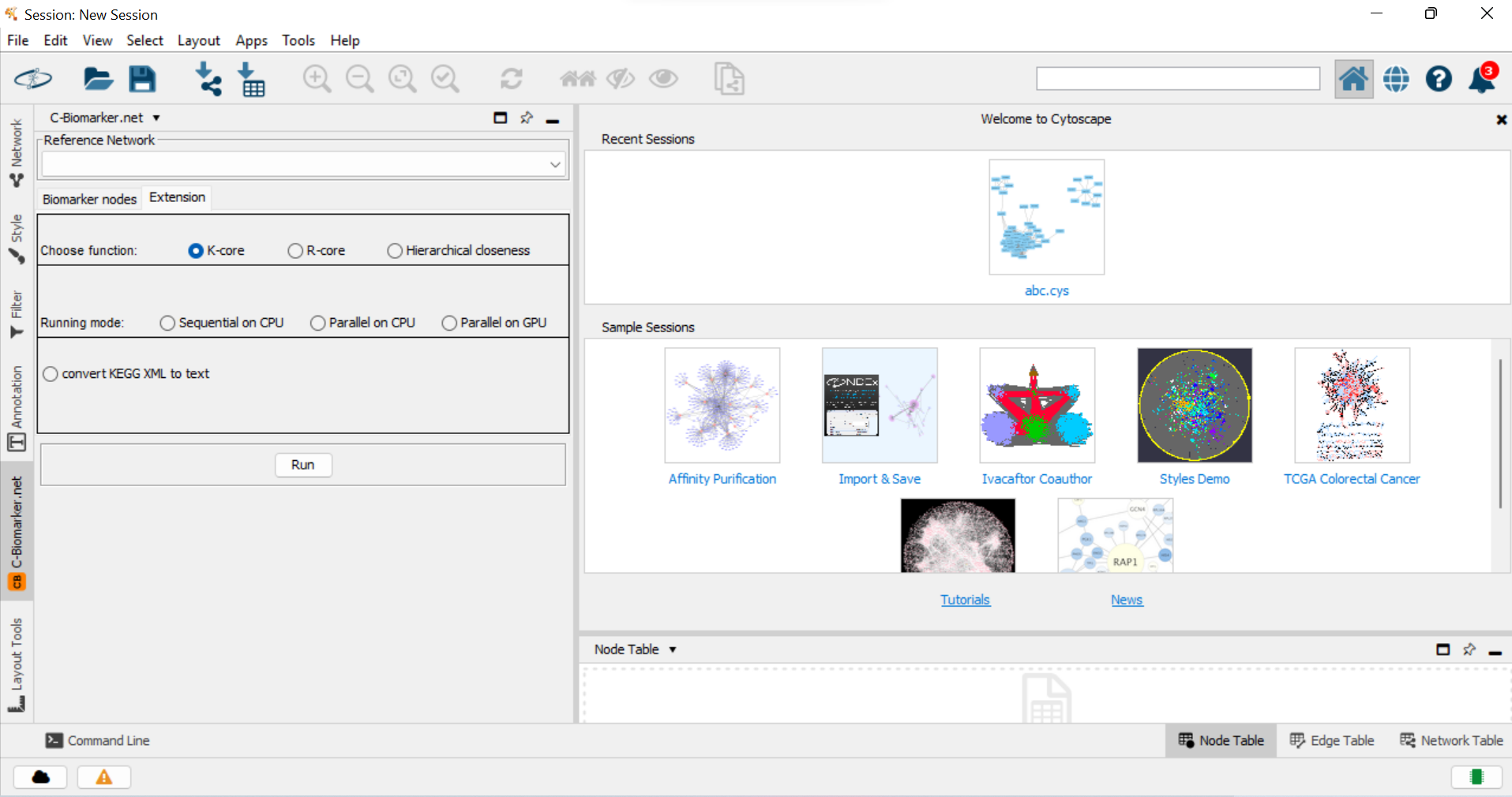


Figure 13: Extension tab of C-Biomarker.net application

The application consists of 2 tabs as follows:

* Tab Biomarker Nodes (Main function): to find biomarker nodes from a current network, please follow these steps:

1. Choose running mode by clicking on one of following radio buttons: Sequential on CPU, Parallel on CPU Parallel on GPU.

2. Click on Run button to choose the result file to save.

* Tab Extension (Extra function):

1. Finding network cores (K-core, R-core) or finding fragile node against mutation Hierarchical closeness (HC) by the steps as follows:

1.1. K-Core, R-Core, Hierarchical closeness.

1.2. Similarly, choose running mode by clicking on one of following radio buttons: Sequential on CPU, Parallel on CPU Parallel on GPU.

2. Convert KEGG XML to text: If you want to analyze a XML file downloaded from KEGG, first you need to convert it to a text file by clicking on the radio button “**convert KEGG XML to text**”.

1. **Install KEGGscape for reading network files from the KEGG database**

To visualize KEGG XML file, we recommend that the user should install automatically from the Cytoscape AppStore: Choose the menu Apps → App Manager in Cytoscape. Then type KEGGscape in the search box click on KEGGscape in area two and then click on the Install button. The installation is successful.

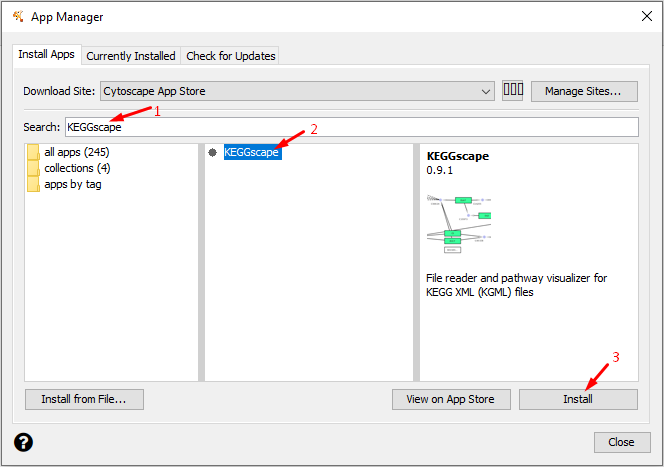


Figure 14: Location of KEGGscape in Cytoscape

# II. Overview of C-Biomarker.net

After installing, C-Biomarker.net will be automatically loaded in the App Cytoscape. The process to identify biomarker genes from C-Biomarker.net include 4 steps: step 1 and step 2 are done in software, and step 3 and step 4 are done by other tools available. Specifically:

* Step 1: Load network by clicking on menu File\Import\Network from file.
* Step 2: Rank candidate biomarker genes by selecting on Biomarker nodes tab, and execute the process in “Briefly introduce the used way of the software”.
* Step 3: Search evidence from PubMed by accessing website address <https://pubmed.ncbi.nlm.nih.gov/> for search the evidence of biomarker genes.
* Step 4: Analyze biological function by the software tool at the website address: https://david.ncifcrf.gov/tools.jsp.

C-Biomarker.net is integrated with Cytoscape and uses the Cytoscape core library to identify cancer marker genes from human networks.

Software features: the ability to quickly compute on large-scale networks, on computers equipped with multi-core CPUs and GPUs.

1. **Case study: Identification of biomarker genes of 17 cancer types from KEGG pathways**

**Step 1: Load networks**

Step 1.1: Download network file (XML) from KEGG database

Users can prepare 17 network files (KEGG xml).

To download network files from KEGG, users need accessing website address:

* <https://www.genome.jp/kegg/pathway.html#disease>

Select the Human Diseases you want to download to your computer.

In this case, select Basal cell carcinoma in Cancer: specific types with the result in Figure 15.

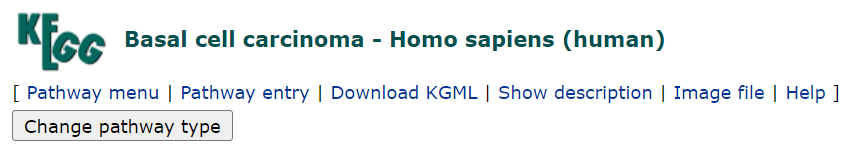


Figure 15: Download KGML

Click on the menu Download KGML to download the data file from KEGG.

For example, the standard data for using the C-Biomarker.net application has the following format: Basal cell carcinoma.xml

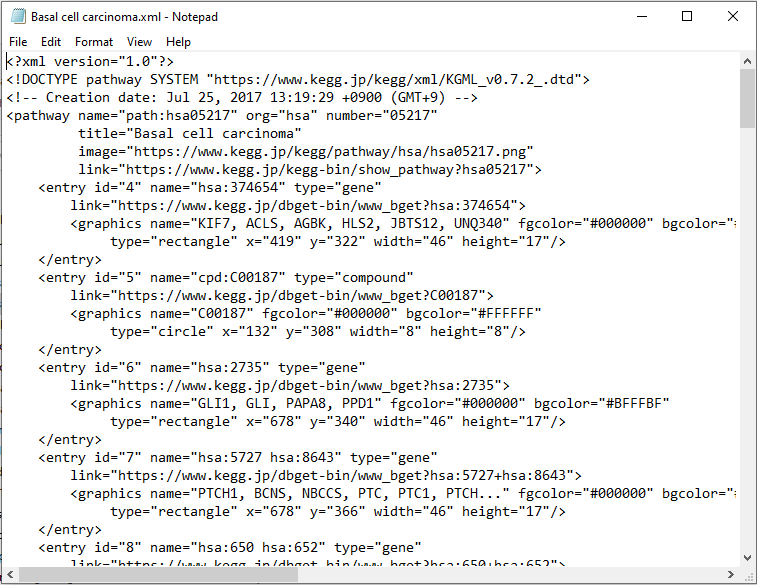


Figure 16: File XML downloaded from KEGG

XML file consists of 2 components: entry (node ​​information), and relationship (network edges).

Step 1.2: Convert the downloaded XML to text file.

Because the XML file cannot be run directly on the Cytoscape app, it is necessary to change the XML file to a TXT file by selecting the Convert KEGG XML to text radio button in Extension tab. The screen to choose file or folder to convert as follows:

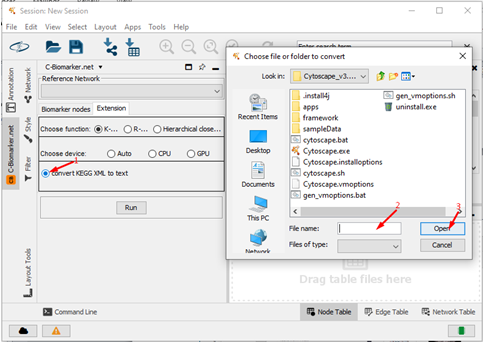


Figure 17: Interface to choose file or folder for converting KEGG to text

At interface of selecting file or folder to convert:

(1) Click on button convert KEGG XML to text

(2) Enter the file name in the field: File name

(3) Click on Open button. The folder to save converted file is displayed on the screen.

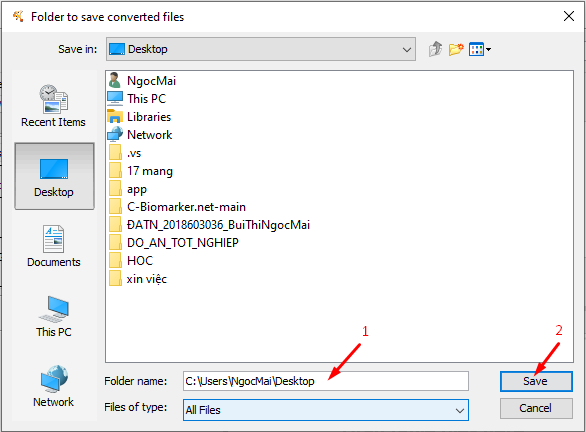


Figure 18: Interface of folder to save converted file

(4) Enter the folder name in the field: Folder name

(5) Click on Save button, and then the resulting folder is displayed on the screen.

After clicking the Save button, a successful conversion message popup will be displayed.

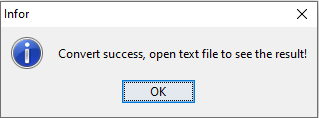


Figure 19: Results screen when running the algorithm

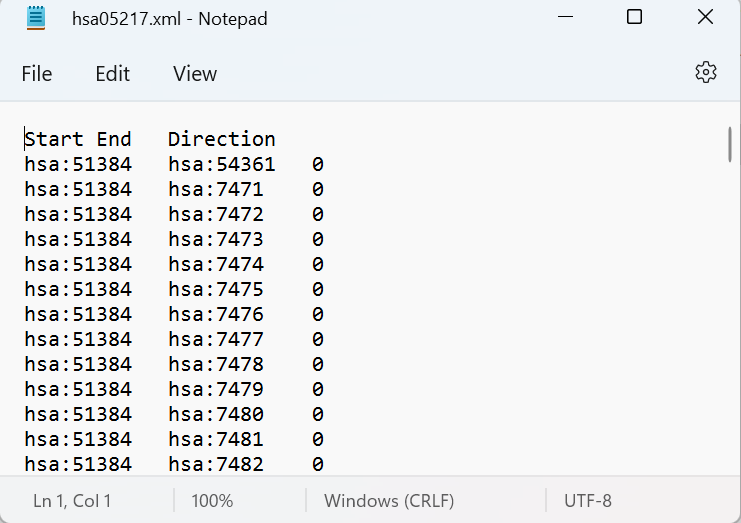


Figure 20: Result after successful conversion

Step 1.3: Load the converted file into Cytoscape by selecting File\Import\Network from the file as follows:

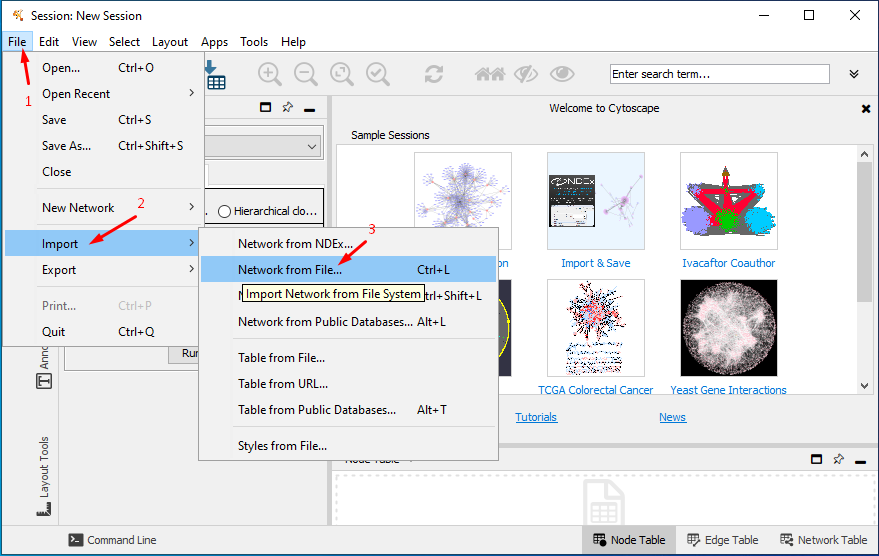


Figure 21: Data file loading interface

Choose the converted network file as follows in Fig.22:

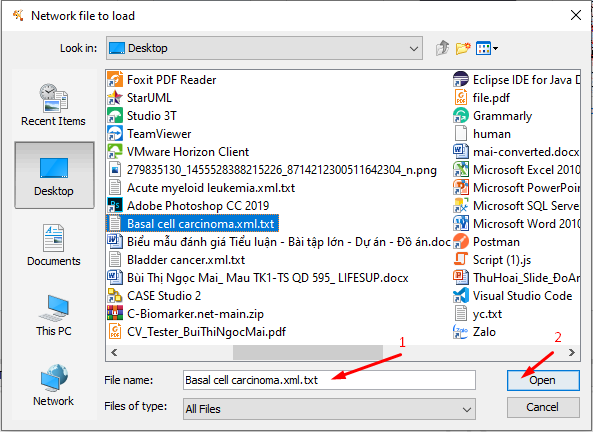


Figure 22: Data file loading interface

If the data file has not specified the source and target columns, you can click on the column headers at the preview interface and choose the type as follows:

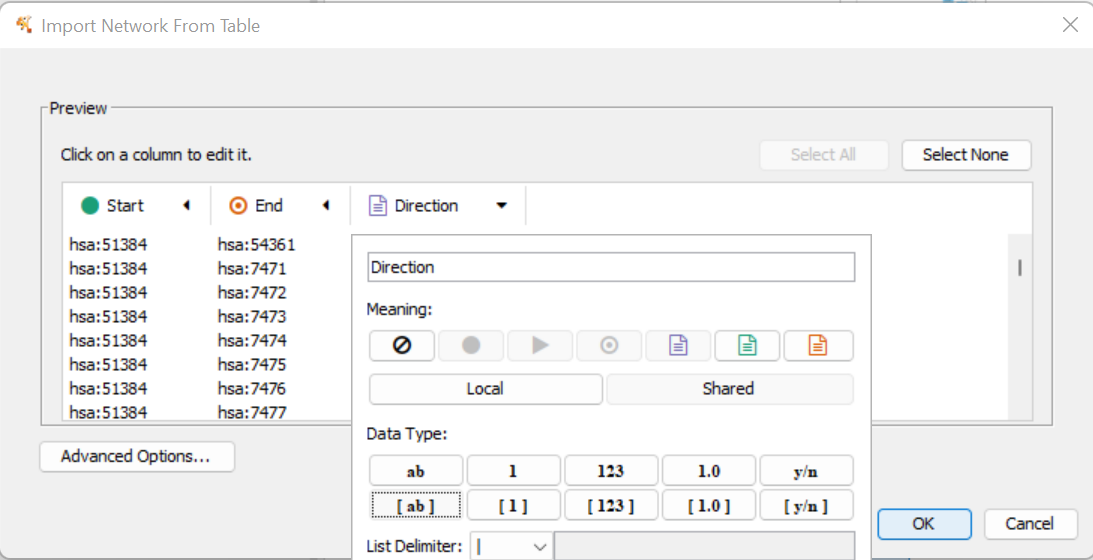


Figure 23: Column format interface

Users are required to specify 3 columns: Start, End, and Direction to be able to import network data files into Cytoscape. Note that data type of Direction must be list of strings [a,b]. After the preview is done, click OK to let Cytoscape load the data.

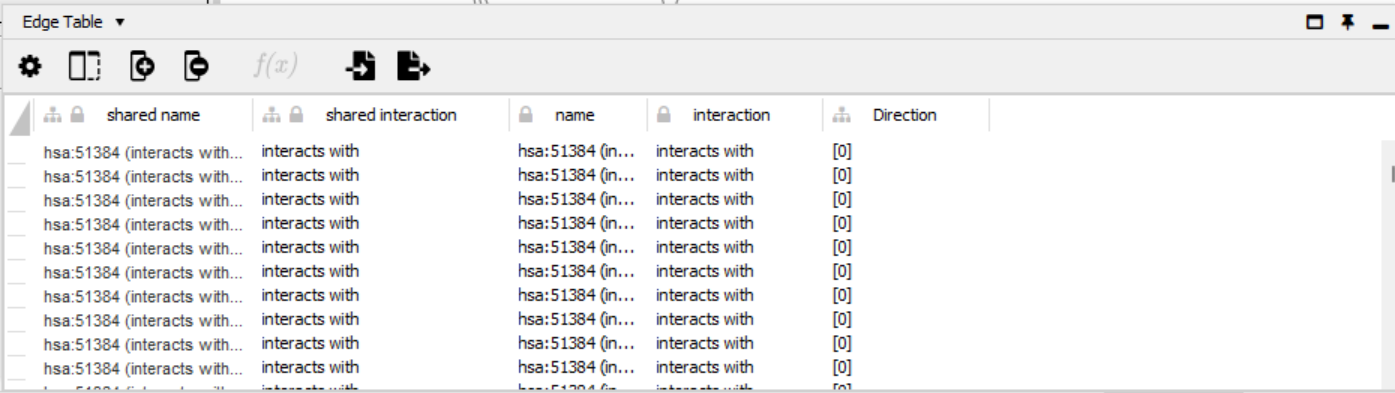


Figure 24: Screen after successful import

**Step 2: Find biomarker nodes from the loading network.**

On the menu C-Biomarker.net:

(1) At the Biomarker nodes tab, click on the Choose device field value (if the node size of the network is smaller than 100, choose Sequential on CPU button; otherwise, if the node size of the network is between 100 and 500 choose Parallel on CPU button, in else case, choose Parallel on GPU).

(2) Click on the Run button.

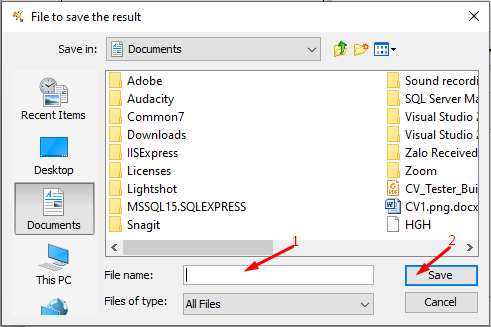


Figure 25: Interface of saving the result

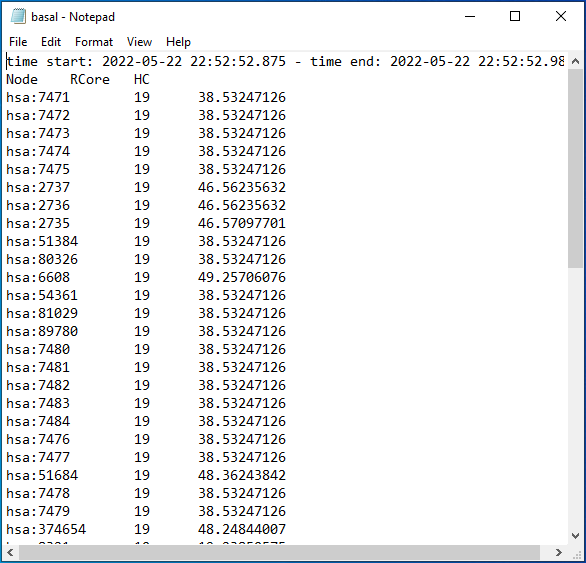


Figure 26: Result interface when running the algorithm

**Step 3: Rank candidate biomarker genes**

Step 3.1. Read the result. In Figure 26, the result is a list of nodes that have been ranked from high to low according to R-Core and HC respectively. In the case that the result is not ranked from high to low, let copy data in the file and paste into excel for sorting by manually. Biomarker nodes are the 3-10 highest ranked nodes located on the innermost core of the network. In other words, biomarkers are sensitive (mutable) nodes located in the innermost core region of the network.

Step 3.2. Convert from gene ID (number) to gene symbol (word) by accessing the website address: <https://www.genome.jp/kegg/tool/conv_id.html>

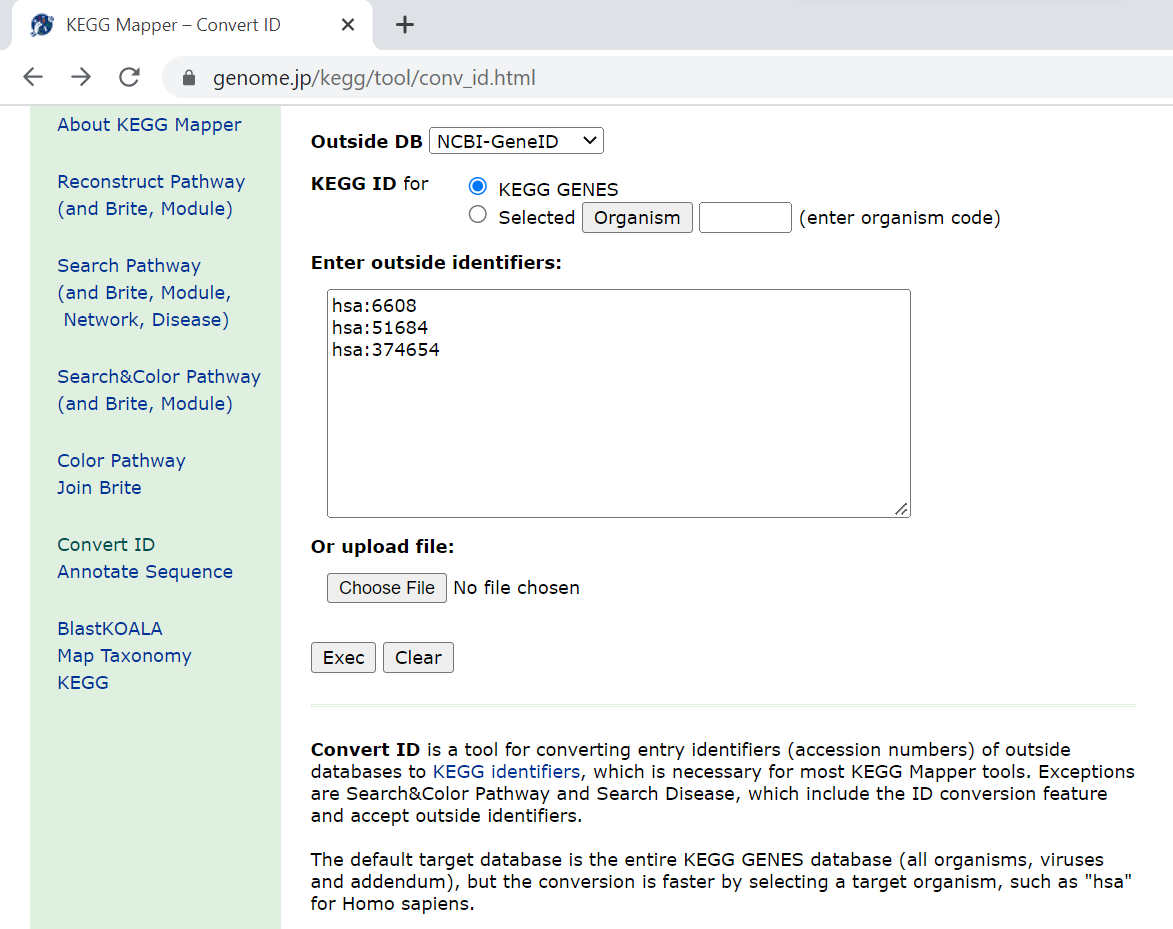


Figure 27: Convert from gene ID to gene symbol

The steps to convert from ID to symbol as follows:

(1) Enter list gene ID.

(2) Click button Exec.

(3) Show the result after converting from gene ID to gene symbol.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| No. | Network | Number of nodes | Number of edges | G1 | G2 | G3 |
|
| 1 | Acute myeloid leukemia | 66 | 53 | FLT3 | KIT | GRB2 |
| 2 | Basal cell carcinoma | 63 | 20 | SMO | SUFU | KIF7 |
| 3 | Bladder cancer | 29 | 16 | CDKN2A | CDKN1A | CCND1 |
| 4 | Breast cancer | 148 | 104 | FGF22 | FGF20 | FGF17 |
| 5 | Chronic myeloid leukemia | 72 | 42 | BCR | ABL1 | GRB2 |
| 6 | Colorectal cancer | 79 | 56 | EGF | AREG | EREG |
| 7 | Endometrial cancer | 55 | 34 | EGF | EGFR | PDPK1 |
| 8 | Gastric cancer | 141 | 73 | FGF22 | FGF20 | FGF17 |
| 9 | Glioma | 78 | 77 | CALML3 | CALML6 | CALM1 |
| 10 | Hepatocellular carcinoma | 169 | 76 | KEAP1 | NFE2L2 | GSTT2B |
| 11 | Melanoma | 71 | 27 | PDGFD | FGF22 | FGF10 |
| 12 | Non-small cell lung cancer | 73 | 59 | EGF | TGFA | ALK |
| 13 | Pancreatic cancer | 78 | 52 | KRAS | EGF | TGFA |
| 14 | Prostate cancer | 94 | 43 | PDGFD | PDGFC | EGF |
| 15 | Renal cell carcinoma | 60 | 35 | RAC1 | CDC42 | PAK6 |
| 16 | Small cell lung cancer | 89 | 36 | LAMA2 | LAMA3 | COL4A6 |
| 17 | Thyroid cancer | 37 | 20 | PAX8 | PPARG | RXRG |

Table 1: Top 3 peaks with the biggest Biomarker results

## Step 4: Search evidence from PubMed

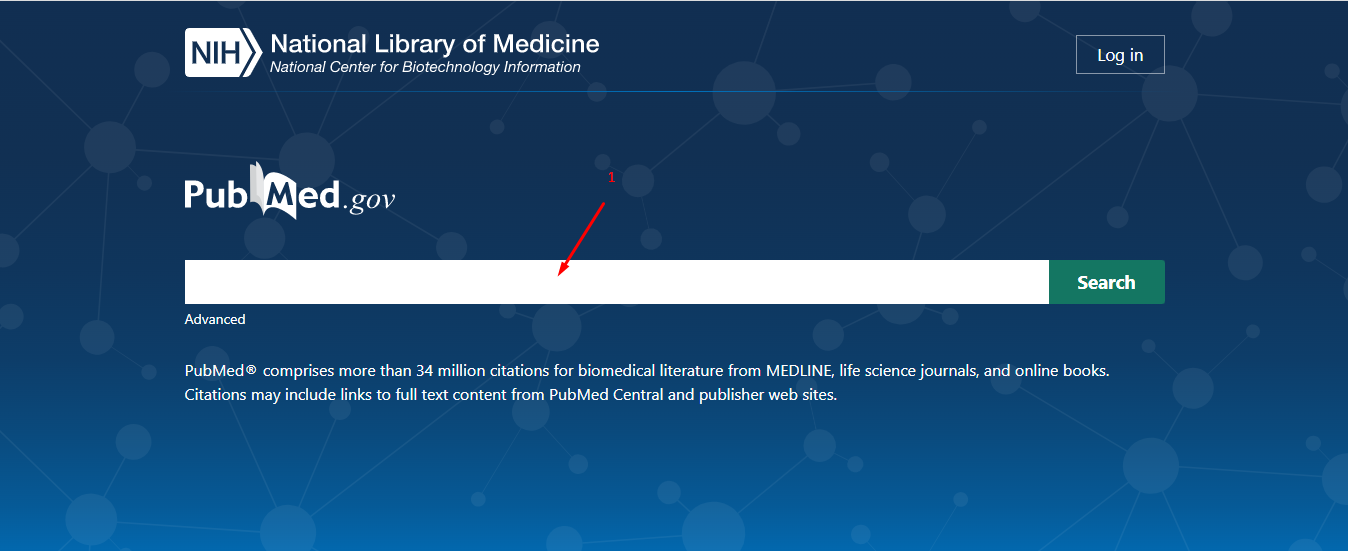


Figure 28: Data file loading interface

The user need to access the website address:

* <https://pubmed.ncbi.nlm.nih.gov/>

In the search box the user searches for the syntax: <gene symbol>< disease name><biomarker>. Then click the Search button.

For example: FLT3 Acute myeloid leukemia biomarker.

**Acute myeloid leukemia**

FLT3: supported by a study with PubMed ID**: 31217189.** PRMT1-mediated FLT3 arginine methylation promotes maintenance of FLT3-ITD + acute myeloid leukemia.

KIT: supported by a study with PubMed ID**:** 32678289**.** KIT pathway upregulation predicts dasatinib efficacy in acute myeloid leukemia.

GRB2: supporteded by a study with PubMed ID**:** 26895103**.** The expression of GADS enhances FLT3-induced mitogenic signaling.

**Basal cell carcinoma**

SMO: supported by a study with PubMed ID**:** 32796174. **Hedgehog Pathway Alterations Downstream of Patched-1 Are Common in Infundibulocystic Basal Cell Carcinoma.**

SUFU: supported by a study with PubMed ID**:** 29186568.Germline SUFU mutation carriers and medulloblastoma: clinical characteristics, cancer risk, and prognosis.

**Bladder cancer**

CDKN2A: supported by a study with PubMed ID**:** 30258198**. CDKN2A as a transcriptomic marker for muscle-invasive bladder cancer risk stratification and therapy decision-making.**

CDKN1A: supported by a study with PubMed ID: 29602637. Importin-11 overexpression promotes the migration, invasion, and progression of bladder cancer associated with the deregulation of CDKN1A and THBS1.

CCND1: supported by a study with PubMed ID:23887292. CCND1 status in metastasizing bladder cancer: a prognosticator and predictor of chemotherapeutic response.

**Chronic myeloid leukemia**

BCR: supported by a study with PubMed ID:31311809. Targeting BCR-ABL1 in Chronic Myeloid Leukemia by PROTAC-Mediated Targeted Protein Degradation.

ABL1: supported by a study with PubMed ID:34185393. BCR-ABL1 p210 screening for chronic myeloid leukemia in patients with peripheral blood cytoses.

GRB2: supported by a study with PubMed ID:23399893. Gads (Grb2-related adaptor downstream of Shc) are required for BCR-ABL-mediated lymphoid leukemia.

**Colorectal cancer**

EGF: supported by a study with PubMed ID:33833529. miR-944 Suppresses EGF-Induced EMT in Colorectal Cancer Cells by Directly Targeting GATA6.

AREG: supported by a study with PubMed ID: 32943459 . Amphiregulin Expression Is a Predictive Biomarker for EGFR Inhibition in Metastatic Colorectal Cancer: Combined Analysis of Three Randomized Trials.

EREG: supported by a study with PubMed ID:26869404. AREG and EREG as Predictive Biomarkers for RAS Wild-Type Colorectal Cancer Treated With Panitumumab: A Fresh Approach to an Old Puzzle.

**Endometrial cancer**

EGF: supported by a study with PubMed ID:20579378. EGFR isoforms and gene regulation in human endometrial cancer cells.

EGFR: supported by a study with PubMed ID:27092881 .MUC1 stimulates EGFR expression and function in endometrial cancer.

**Hepatocellular carcinoma**

KEAP1: supported by a study with PubMed ID:27650414. Clinical implication of Keap1 and phosphorylated Nrf2 expression in hepatocellular carcinoma.

NFE2L2: supported by a study with PubMed ID:34299031 **.**Somatic Mutations in Circulating Cell-Free DNA and Risk for Hepatocellular Carcinoma in Hispanics.

**Melanoma**

PDGFD:supported by a study with PubMed ID:23462921. Gene variants in angiogenesis and lymphangiogenesis and cutaneous melanoma progression.

**Non-small cell lung cancer**

EGF: supported by a study with PubMed ID: 28348561. CIMAvax-EGF: A New Therapeutic Vaccine for Advanced Non-Small Cell Lung Cancer Patients.

TGFA: supported by a study with PubMed ID: 21528670. Expression analysis of angiogenesis-related genes in Bulgarian patients with early-stage non-small cell lung cancer.

ALK: supported by a study with PubMed ID:29455675. The function and therapeutic targeting of anaplastic lymphoma kinase (ALK) in non-small cell lung cancer (NSCLC).

**Pancreatic cancer**

KRAS: supported by a study with PubMed ID: 32005945. Role of oncogenic KRAS in the diagnosis, prognosis, and treatment of pancreatic cancer.

EGF: supported by a study with PubMed ID: 29125273.Expression and clinical significance of EGF and TGF-α in chronic pancreatitis and pancreatic cancer.

**Prostate cancer**

PDGFD: supported by a study with PubMed ID:33918816 . Neuropilin-2 and Its Transcript Variants Correlate with Clinical Outcome in Bladder Cancer.

EGF: supported by a study with PubMed ID:24435707. A ROS/STAT3/HIF-1α signaling cascade mediates EGF-induced TWIST1 expression and prostate cancer cell invasion.

**Renal cell carcinoma**

RAC1: supported by a study with PubMed ID:32371578. Rac Signaling Drives Clear Cell Renal Carcinoma Tumor Growth by Priming the Tumor Microenvironment for an Angiogenic Switch.

CDC42: supported by a study with PubMed ID: 16343437. Splicing variant of Cdc42 interacting protein-4 disrupts beta-catenin-mediated cell-cell adhesion: expression and function in renal cell carcinoma.

**Thyroid cancer**

PAX8: supported by a study with PubMed ID:21878896. Molecular genetics and diagnosis of thyroid cancer.

PPARG: supported by a study with PubMed ID:27250077. Key genes and pathways predicted in papillary thyroid carcinoma based on bioinformatics analysis.

## Step 5: Biological function analysis

For biological function analysis, the user needs to access the website address: <https://david.ncifcrf.gov/tools.jsp>



Figure 29:Search Biological function analysis

(1) Paste a list: id of the top 3 genes

(2) Click on Combobox select value: Entrez\_gene\_id

(3) Click on the radio button and select value: Gene list

(4) Click on button Submit List

|  |
| --- |
| **Acute myeloid leukemia:** Three candidate cancer biomarker genes, including: FLT3, KIT, GRB2 share the same biological functions as follows:  Acute myeloid leukemia: is aggressive and rapidly lethal blood cancer, characterized by the accumulation of immature leukemic blasts in the bone marrow and blood, leading to organ infiltration and failure of normal hematopoietic function.  Ras signaling pathway: The Ras/Raf/MAPK pathway is probably the best-characterized signal transduction pathway in cell biology. The function of this pathway is to transduce signals from the extracellular milieu to the cell nucleus where specific genes are activated for cell growth, division, and differentiation.  MAPK signaling pathway: MAPK pathways relay, amplify and integrate signals from a diverse range of stimuli and elicit an appropriate physiological response including cellular proliferation, differentiation, development, inflammatory responses, and apoptosis in mammalian cells.  PI3K-Akt signaling pathway: The PI3K/AKT signaling pathway is frequently in a dysregulated state in tumors, and has now become an important anticancer target (32). The PI3K/AKT signaling pathway itself serves a major role in regulating cell physiology and pathology, including cell proliferation, survival, and invasion.  Pathways in cancer: Twelve signaling pathways, which are responsible for 3 cancer core functions involving cell survival, fate, and genome maintenance.  Receptor: Receptors are proteins, usually cell surface receptors, which bind to ligands and cause responses in the immune system. Receptors can be found in various immune cells like B cells, T cells, NK cells, monocytes, and stem cells.  Plasma membrane: The plasma membrane, or the cell membrane, protects a cell. It also provides a fixed environment inside the cell. And that membrane has several different functions. One is to transport nutrients into the cell and also to transport toxic substances out of the cell. |
| **Basal cell carcinoma:** Three candidate cancer biomarker genes, including: SMO, SUFU, KIF7 share the same biological functions as follows:  Ciliary tip: Ciliary Tip Signaling Compartment Is Formed and Maintained by Intraflagellar Transport.  Hedgehog signaling pathway: The hedgehog signaling pathway is a mechanism that directs the development of embryonic cells in animals, from invertebrates to vertebrates. The hedgehog signaling pathway is a system of genes and gene products, mostly proteins, that convert one kind of signal into another, called transduction.  Basal cell carcinoma: Basal cell carcinoma is a type of skin cancer. Basal cell carcinoma begins in the basal cells — a type of cell within the skin that produces new skin cells as old ones die off. Basal cell carcinoma often appears as a slightly transparent bump on the skin, though it can take other forms.  Pathways in cancer: Twelve signaling pathways, which are responsible for 3 cancer core functions involving cell survival, fate, and genome maintenance. |
| **Bladder cancer:** Three candidate cancer biomarker genes, including: CDKN2A, CDKN1A, CCND1 share the same biological functions as follows:  Regulation of G1/S transition of the mitotic cell cycle: The G1/S transition is highly regulated by transcription factor p53 to halt the cell cycle when DNA is damaged. It is a "point of no return" beyond which the cell is committed to dividing; in yeast, this is called START and in multicellular eukaryotes, it is termed the restriction point (R-Point).  Regulation of cyclin-dependent protein serine/threonine kinase activity: Modulates the activity of a cyclin-dependent protein serine/threonine kinase, enzymes of the protein kinase family that are regulated through association with cyclins and other proteins.  G1/S transition of mitotic cell cycle: The G1/S transition is highly regulated by transcription factor p53 in order to halt the cell cycle when DNA is damaged. It is a "point of no return" beyond which the cell is committed to dividing; in yeast this is called START and in multicellular eukaryotes it is termed the restriction point (R-Point).  Bladder cancer: Bladder cancer occurs when there are abnormal, cancerous cells growing uncontrollably in the lining of the bladder, which is the hollow organ in the lower abdomen that stores urine. These cancerous cells begin to affect the normal function of the bladder and can spread to surrounding organs.  Cytosol: The cytosol serves several functions within a cell. It is involved in signal transduction between the cell membrane and the nucleus and organelles. It transports metabolites from their production site to other parts of the cell. It is important for cytokinesis, when the cell divides in mitosis. The cytosol plays a role in eukaryote metabolism. In animals, this includes glycolysis, gluconeogenesis, protein biosynthesis, and the pentose phosphate pathway. |
| **Breast cancer:** Three candidate cancer biomarker genes, including: FGF22, FGF20, FGF17 share the same biological functions as follows:  Cytoplasm: The cytoplasm is an integral part of both prokaryotic and eukaryotic cells and functions to house and maintain an optimal environment for the cellular organelles.  Breast cancer: Breast cancer is a type of cancer that starts in the breast. It can start in one or both breasts.  Heparin-binding growth factor/Fibroblast growth factor: Heparin is a potent modulator of receptor binding of growth factors such as fibroblast growth factor (FGF), vascular endothelial growth factor, and heparin-binding epidermal growth factor (HB-EGF), that play a role in wound repair.  Fibroblast growth factor receptor binding: The fibroblast growth factor receptors (FGFRs) regulate important biological processes including cell proliferation and differentiation during development and tissue repair.  Cytokine, IL-1-like: Members of IL-1 family cytokines are involved in the process of innate and adaptive immunity as well as fibrosis in systemic sclerosis (SSc). IL-1 family gene polymorphisms, abnormal expression of IL-1, and its potential role in the fibrosis process have been explored in SSc.  FGF: Fibroblast Growth Factors (FGFs) are potent regulators of cell proliferation and differentiation. |
| **Chronic myeloid leukemia:** Three candidate cancer biomarker genes, including: BCR, ABL1, GRB2 share the same biological functions as follows:  Acetylation: Acetylation is a modification that can dramatically change the function of a protein through alteration of its properties, including hydrophobicity, solubility, and surface properties, all of which may influence protein conformation and interactions with substrates, cofactors, and other macromolecules.  Cytosol: The cytosol serves several functions within a cell. It is involved in signal transduction between the cell membrane and the nucleus and organelles. It transports metabolites from their production site to other parts of the cell. It is important for cytokinesis when the cell divides in mitosis. The cytosol plays a role in eukaryote metabolism. In animals, this includes glycolysis, gluconeogenesis, protein biosynthesis, and the pentose phosphate pathway.  Chronic myeloid leukemia: In chronic myeloid leukemia, the bone marrow produces too many white blood cells. Initially, these cells function relatively normally. However, as the condition progresses, immature white blood cells called myeloblasts (or blasts) accumulate in the blood and bone marrow.  Pathways in cancer: Twelve signaling pathways, which are responsible for 3 cancer core functions involving cell survival, fate, and genome maintenance. |
| **Colorectal cancer:** Three candidate cancer biomarker genes, including: EGF, AREG, EREG share the same biological functions as follows:  Colorectal cancer: Colorectal cancer starts in the colon or the rectum. These cancers can also be called colon cancer or rectal cancer, depending on where they start.  Positive regulation of epidermal growth factor-activated receptor activity.  Epidermal growth factor receptor binding: The epidermal growth factor receptor protein is involved in cell signaling pathways that control cell division and survival.  Epidermal growth factor receptor signaling pathway: The epidermal growth factor receptor (EGFR) signaling pathway is one of the most important pathways that regulate growth, survival, proliferation, and differentiation in mammalian cells.  Clathrin-coated endocytic vesicle membrane: Clathrin-coated vesicles (CCVs) mediate endocytosis of plasma membrane proteins and deliver their content to the endosomes for subsequent recycling to the plasma membrane or transport to the vacuole for degradation. |
| **Endometrial cancer:** Three candidate cancer biomarker genes, including: EGF, EGFR, PDPK1 share the same biological functions as follows:  Endometrial cancer: Endometrial cancer is cancer that arises from the endometrium (the lining of the uterus or womb). It is the result of the abnormal growth of cells that can invade or spread to other parts of the body. The first signis most often vaginal bleeding not associated with a menstrual period.  Epidermal growth factor receptor signaling pathway: The epidermal growth factor receptor (EGFR) signaling pathway is one of the most important pathways that regulate growth, survival, proliferation, and differentiation in mammalian cells.  Non-small cell lung cancer: Non-small-cell lung carcinoma (NSCLC) is any type of epithelial lung cancer other than small-cell lung carcinoma (SCLC).  Prostate cancer: The prostate's primary function is to produce the fluid that nourishes and transports sperm (seminal fluid). |
| **Gastric cancer**: Three candidate cancer biomarker genes, including: FGF22, FGF20, FGF17 share the same biological functions as follows:  Cytoplasm: One of the major functions of cytoplasm is to enable cells to maintain their turgidity, which enables the cells to hold their shape. Other functions of cytoplasm are as follows:   * The jelly-like fluid of the cytoplasm is composed of salt and water and is present within the membrane of the cells and embeds all of the parts of the cells and organelles. * The cytoplasm is home to many activities of the cell as it contains molecules, and enzymes that are crucial in the breakdown of the waste.   Pathways in cancer: Twelve signaling pathways, which are responsible for 3 cancer core functions involving cell survival, fate, and genome maintenance.  Heparin-binding growth factor/Fibroblast growth factor: Heparin is a potent modulator of receptor binding of growth factors such as fibroblast growth factor (FGF), vascular endothelial growth factor, and heparin-binding epidermal growth factor (HB-EGF), that play a role in wound repair.  Fibroblast growth factor receptor binding: The fibroblast growth factor receptors (FGFRs) regulate important biological processes including cell proliferation and differentiation during development and tissue repair.  Cytokine, IL-1-like: Members of IL-1 family cytokines are involved in the process of innate and adaptive immunity as well as fibrosis in systemic sclerosis (SSc). IL-1 family gene polymorphisms, abnormal expression of IL-1 and its potential role in the fibrosis process have been explored in SSc.  FGF: Fibroblast Growth Factors (FGFs) are potent regulators of cell proliferation and differentiation.  Gastric cancer: Gastric cancer is a disease in which malignant (cancer) cells form in the lining of the stomach. |
| **Glioma:** Three candidate cancer biomarker genes, including: CALML3, CALML6, CALM1 share the same biological functions as follows:  Pathways in cancer: Twelve signaling pathways, which are responsible for 3 cancer core functions involving cell survival, fate, and genome maintenance.  Glioma: Gliomas begin in the gluey supportedive cells (glial cells) that surround nerve cells and help them function.  Enzyme regulator activity: Molecular function.  DOMAIN:EF-hand 4: Biochemical and structural studies are now uncovering the mechanisms by which EFCaBPs work and are helping to define their biological activities, while simultaneously expanding knowledge of the roles of these proteins in normal cellular physiology and the pathology of the disease.  Phototransduction: This process involves the sequential activation of a series of signaling proteins, leading to the eventual opening or closing of ion channels in the photoreceptor cell membrane.  Long-term potentiation: It is an important process in the context of synaptic plasticity. LTP recording is widely recognized as a cellular model for the study of memory.  Amphetamine addiction: Amphetamines are highly addictive drugs that stimulate the central nervous system. |
| **Hepatocellular carcinoma:** Three candidate cancer biomarker genes, including: KEAP1, NFE2L2, GSTT2B share the same biological functions as follows:  Fluid shear stress and atherosclerosis: Fluid shear stress has emerged as an essential feature of atherogenesis. This fluid drag force acting on the vessel wall is mechanotransduced into a biochemical signal that results in changes in vascular behavior.  Hepatocellular carcinoma: Hepatocellular carcinoma (HCC) is the most common type of primary liver cancer.  Chemical carcinogenesis - reactive oxygen species: The role of oxidative stress in chemical carcinogenesis.  Pathways in cancer: Twelve signaling pathways, which are responsible for 3 cancer core functions involving cell survival, fate, and genome maintenance.  Nucleoplasm: The main function of the nucleoplasm is to provide the proper environment for essential processes that take place in the nucleus and to store the structures that are used in these processes.  Cytoplasm: One of the major functions of cytoplasm is to enable cells to maintain their turgidity, which enables the cells to hold their shape. Other functions of cytoplasm are as follows:   * The jelly-like fluid of the cytoplasm is composed of salt and water and is present within the membrane of the cells and embeds all of the parts of the cells and organelles. * The cytoplasm is home to many activities of the cell as it contains molecules, enzymes that are crucial in the break down of the waste.   Cytosol: The cytosol serves several functions within a cell. It is involved in signal transduction between the cell membrane and the nucleus and organelles. It transports metabolites from their production site to other parts of the cell. It is important for cytokinesis, when the cell divides in mitosis. The cytosol plays a role in eukaryote metabolism. In animals, this includes glycolysis, gluconeogenesis, protein biosynthesis, and the pentose phosphate pathway. |
| **Melanoma:** Three candidate cancer biomarker genes, including: PDGFD, FGF22, FGF10 share the same biological functions as follows:  Melanoma: Melanoma, also redundantly known as malignant melanoma, is a type of skin cancer that develops from the pigment-producing cells known as melanocytes.  Growth factor activity: The function that stimulates a cell to grow or proliferate. Most growth factors have other actions besides the induction of cell growth or proliferation.  Rap1 signaling pathway: the Rap1 signaling pathway exists in many important cellular processes such as the information and control of cell adhesion and cell junction, cell migration, polarization, and cell proliferation and survival.  Positive regulation of cell proliferation: Any process that activates or increases the rate or extent of cell proliferation.  MAPK signaling pathway: MAPK pathways relay, amplify and integrate signals from a diverse range of stimuli and elicit an appropriate physiological response including cellular proliferation, differentiation, development, inflammatory responses and apoptosis in mammalian cells. |
| **Non-small cell lung cancer:** Three candidate cancer biomarker genes, including: EGF, TGFA, ALK share the same biological functions as follows:  Pathways in cancer: Twelve signaling pathways, which are responsible for 3 cancer core functions involving cell survival, fate, and genome maintenance.  Non-small cell lung cancer: Non-small-cell lung carcinoma (NSCLC) is any type of epithelial lung cancer other than small-cell lung carcinoma (SCLC).  TOPO\_DOM:Extracellular: Involved in an extracellular matrix organization (By similarity). Involved in brain organization and function (By similarity).  Plasma membrane: The plasma membrane, or the cell membrane, provides protection for a cell. It also provides a fixed environment inside the cell. And that membrane has several different functions. One is to transport nutrients into the cell and also to transport toxic substances out of the cell.  PI3K-Akt signaling pathway: The PI3K/AKT signaling pathway itself serves a major role in regulating cell physiology and pathology, including cell proliferation, survival and invasion. |
| **Pancreatic cancer:** Three candidate cancer biomarker genes, including: KRAS, EGF, TGFA share the same biological functions as follows:  Plasma membrane: The plasma membrane, or the cell membrane, provides protection for a cell. It also provides a fixed environment inside the cell. And that membrane has several different functions. One is to transport nutrients into the cell and also to transport toxic substances out of the cell.  Pathways in cancer: Twelve signaling pathways, which are responsible for 3 cancer core functions involving cell survival, fate, and genome maintenance.  PI3K-Akt signaling pathway: The PI3K/AKT signaling pathway itself serves a major role in regulating cell physiology and pathology, including cell proliferation, survival and invasion.  Positive regulation of MAP kinase activity: Any process that activates or increases the frequency, rate or extent of MAP kinase activity.  Non-small cell lung cancer: Non-small-cell lung carcinoma (NSCLC) is any type of epithelial lung cancer other than small-cell lung carcinoma (SCLC).  Glioma: Gliomas begin in the gluey supportedive cells (glial cells) that surround nerve cells and help them function.  Pancreatic cancer: can affect the function of the pancreas, grow into nearby blood vessels and organs, and eventually spread to other parts of the body through a process called metastasis. |
| **Prostate cancer:** Three candidate cancer biomarker genes, including: PDGFD, PDGFC, EGF share the same biological functions as follows:  Positive regulation of MAP kinase activity: Any process that activates or increases the frequency, rate, or extent of MAP kinase activity.  Growth factor activity: The function that stimulates a cell to grow or proliferate. Most growth factors have other actions besides the induction of cell growth or proliferation.  Melanoma: Melanoma, also redundantly known as malignant melanoma,is a type of skin cancer that develops from the pigment-producing cells known as melanocytes.  Positive regulation of cell proliferation: Any process that activates or increases the rate or extent of cell proliferation.  MAPK signaling pathway: MAPK pathways relay, amplify and integrate signals from a diverse range of stimuli and elicit an appropriate physiological response including cellular proliferation, differentiation, development, inflammatory responses and apoptosis in mammalian cells. |
| **Renal cell carcinoma:** Three candidate cancer biomarker genes, including: RAC1, CDC42, PAK6 share the same biological functions as follows:  Cytoplasm: One of the major functions of cytoplasm is to enable cells to maintain their turgidity, which enables the cells to hold their shape. Other functions of cytoplasm are as follows:   * The jelly-like fluid of the cytoplasm is composed of salt and water and is present within the membrane of the cells and embeds all of the parts of the cells and organelles. * The cytoplasm is home to many activities of the cell as it contains molecules, enzymes that are crucial in the break down of the waste.   Cytosol: The cytosol serves several functions within a cell. It is involved in signal transduction between the cell membrane and the nucleus and organelles. It transports metabolites from their production site to other parts of the cell. It is important for cytokinesis, when the cell divides in mitosis. The cytosol plays a role in eukaryote metabolism. In animals, this includes glycolysis, gluconeogenesis, protein biosynthesis, and the pentose phosphate pathway.  Renal cell carcinoma: Renal cell carcinoma (RCC) is a kidney cancer that originates in the lining of the proximal convoluted tubule, a part of the very small tubes in the kidney that transport primary urine.  Axon guidance: is a subfield of neural development concerning the process by which neurons send out axons to reach their correct targets.  Ras signaling pathway: The function of this pathway is to transduce signals from the extracellular milieu to the cell nucleus where specific genes are activated for cell growth, division and differentiation. |
| **Small cell lung cancer:** Three candidate cancer biomarker genes, including: LAMA2, LAMA3, COL4A6 share the same biological functions as follows:  Basement membrane: The basement membrane (BM) is a special type of extracellular matrix that lines the basal side of epithelial and endothelial tissues. Functionally, the BM is important for providing physical and biochemical cues to the overlying cells, sculpting the tissue into its correct size and shape.  Extracellular matrix structural constituent: Extracellular matrices (ECMs) are multiplicate well-organized 3-dimensional architectural networks with critical structural and functional roles in tissue organization and remodeling and the regulation of cellular processes.  ECM-receptor interaction: ECM-receptor interaction: ECM-receptor interaction pathways were the most upregulated gene-enriched signaling pathways. They play an important role in the process of tumor shedding, adhesion, degradation, movement, and hyperplasia.  Small cell lung cancer: Small cell lung cancer is a disease in which malignant (cancer) cells form in the tissues of the lung. There are two main types of small cell lung cancer. Smoking is the major risk factor for small cell lung cancer. Signs and symptoms of small cell lung cancer include coughing and shortness of breath.  Amoebiasis: Amoebiasis, or amoebic dysentery, is an infection caused by Entamoeba histolytica. Amoebiasis can be present with no, mild, or severe symptoms.  Extracellular matrix: In biology, the extracellular matrix (ECM) is a three-dimensional network consisting of extracellular macromolecules and minerals, such as collagen, enzymes, glycoproteins, and hydroxyapatite that provide structural and biochemical supported to surrounding cells. |
| **Thyroid cancer:** Three candidate cancer biomarker genes, including: PAX8, PPARG, RXRG share the same biological functions as follows:  Thyroid cancer: is a rare but well-described phenomenon and must be considered when evaluating thyroid carcinoma with concurrent hyperthyroidism.  Pathways in cancer: Twelve signaling pathways, which are responsible for 3 cancer core functions involving cell survival, fate, and genome maintenance.  Transcription factor activity, sequence-specific DNA binding: A transcription regulator activity that modulates transcription of gene sets via selective and non-covalent binding to a specific double-stranded genomic DNA sequence (sometimes referred to as a motif) within a cis-regulatory region.  Cell differentiation: is the process by which dividing cells change their functional or phenotypical type. All cells presumably derive from stem cells and obtain their functions as they mature.  Chromatin: The primary function is to package long DNA molecules into more compact, denser structures.  RNA polymerase II transcription factor activity, sequence-specific DNA binding: DNA-binding transcription factor activity that activates or increases transcription of specific gene sets transcribed by RNA polymerase II. |

1. **Case study: Identification of biomarker genes from a large human signaling network**

**Step 1: Load networks**

Users can download data files from [1]. To use the C-Biomarker.net application, the downloaded network data files must meet the following criteria:

There are at least 2 columns in the data file specified in the Start and End formats, for the users specify 2 columns of the Start and End formats. The purpose of this is to define the source and destination nodes so that the application's K-Core and R-Core calculation functions can be used. The name of each node in these two columns is written next to each other without spaces.

If you want to use 2 functions to calculate HC (Hierarchical closeness) and biomarker, the input data is required to add 2 columns Direction and Weight, in which:

1. Direction: Indicates whether the current link between the Start and End node pairs is undirected or directed. 0 is an undirected link and 1 is a directed link.
2. Weight: Represents the weight of the link between 2 nodes. For calculating R-Core or HC, this value can be set to 0.

Example: The standard data for using C-Biomarker.net application has the following column format:

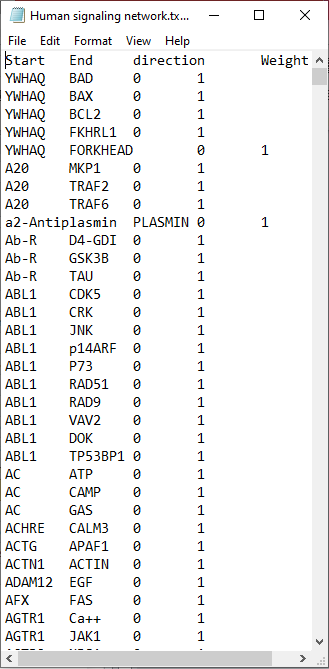


Figure 30: Data column format used in C-Biomarker.net

**Step1: Load network**

From the Cytoscape menu, select File -> Import -> Network from file as follows:

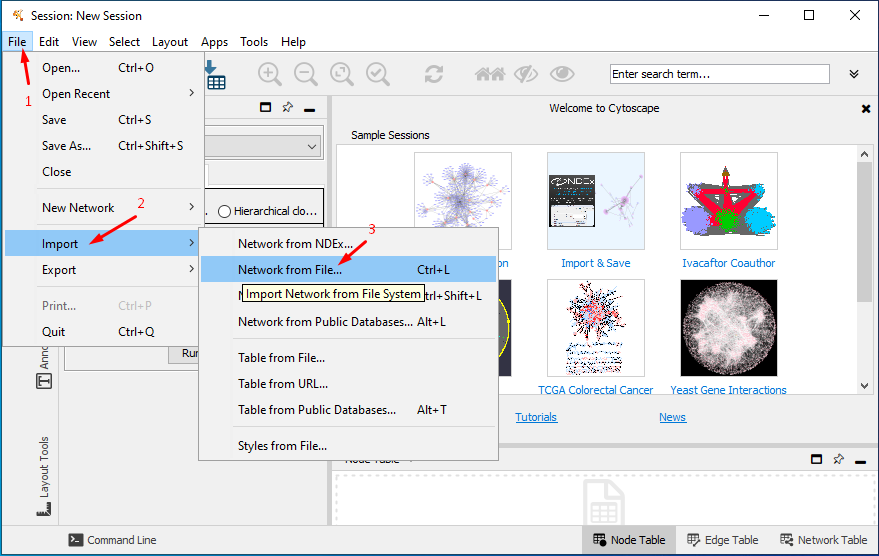


Figure 31: Data file loading interface

After selecting the prepared network data file, the preview interface of the data file will appear as Fig.32.

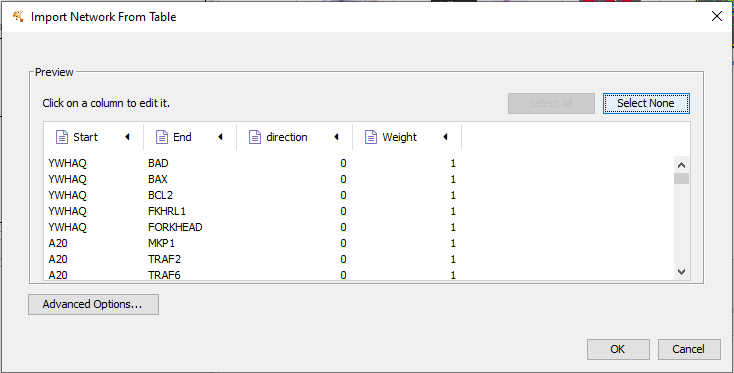


Figure 32: Network data preview interface

If the data file has not specified the source and target columns, you can click on the column headers at the preview interface and choose the type as follows:

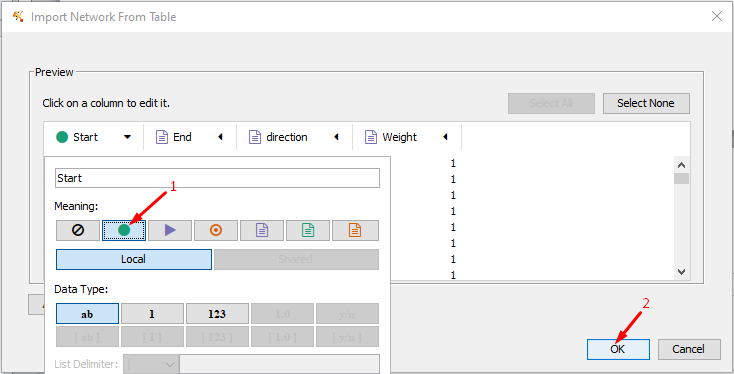


Figure 33: Column format interface

Users are required to specify 2 columns Start and End to be able to import network data files into Cytoscape. Cytoscape will rely on these two columns to visualize the file's data into a network of nodes. After the preview is done, click OK to let Cytoscape load the data.

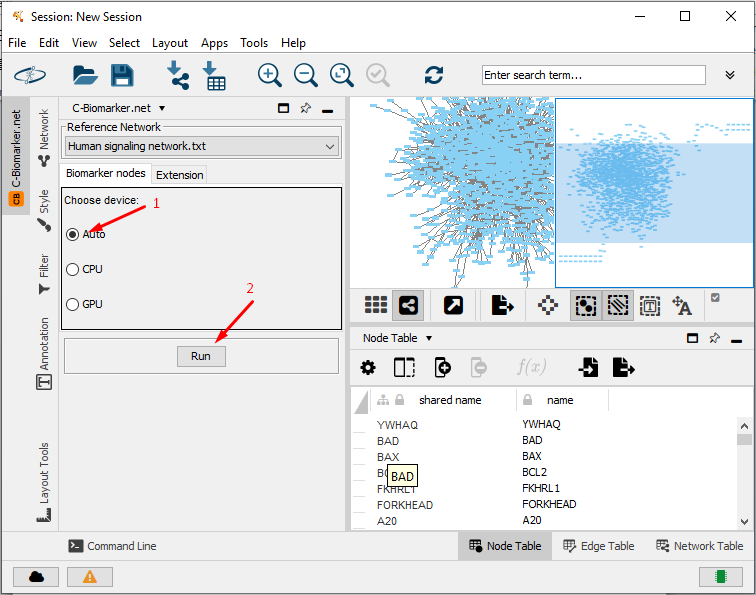


Figure 34: Screen after successful import

**Step 2: Find biomarker nodes from the loading network.**

At the menu C-Biomarker.net:

(1) At the Biomarker nodes tab, click on the Choose device field value.

(2) Click on button Run.

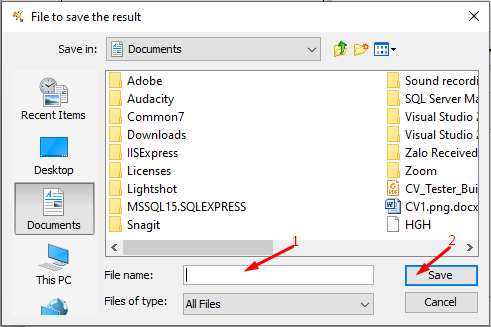


Figure 35: Interface to save the result

At interface to save the result:

(1) Enter the file name in the field of File name.

(2) After clicking on button Save, the result file is displayed on the screen.

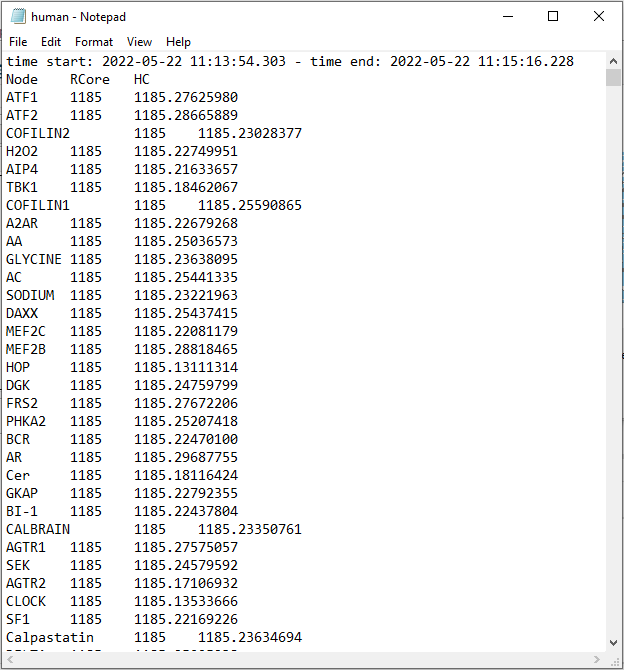


Figure 36: Result interface when running the algorithm

**Step 3: Rank candidate biomarker genes**

The result is a list of nodes that have been ranked from high to low according to R-Core and HC respectively. Biomarker nodes are the top 3-10 highest ranking nodes located on the innermost core of the network. In other words, biomarkers are sensitive (mutable) nodes located in the innermost core region of the network.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| No. | Network | Number of nodes | Number of edges | G1 | G2 | G3 |
|
| 1 | Human signaling network | 1561 | 5089 | CHRD | NOG | BMPR1A |

Table 2: Top 3 nodes with the biggest Biomarker results

**Step 4: Search evidence from PubMed**

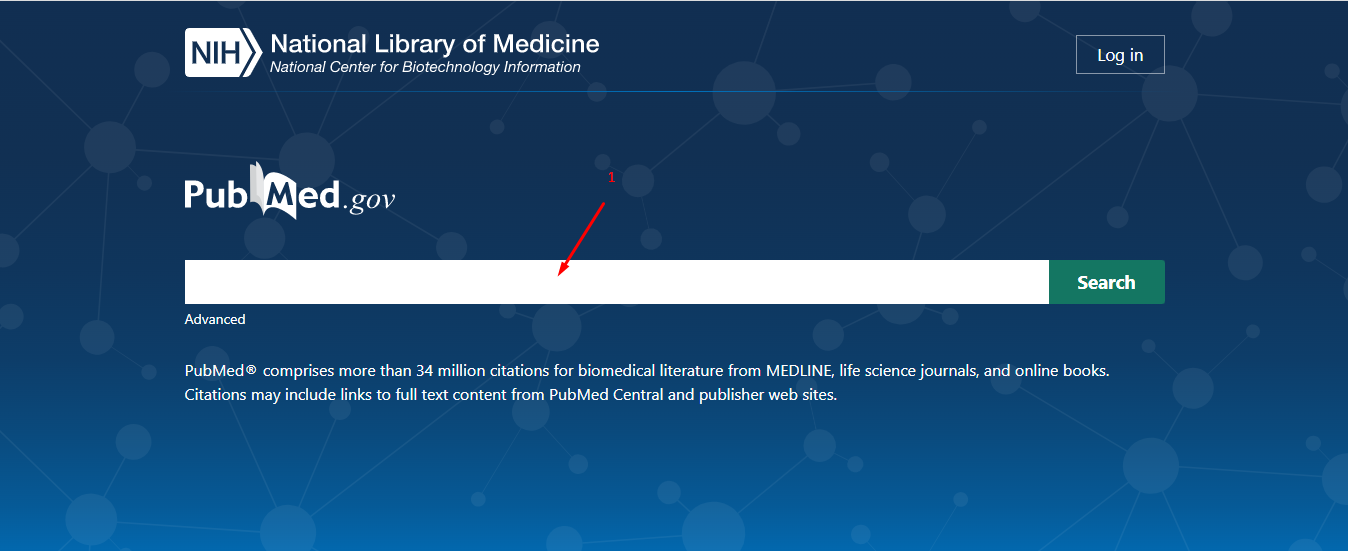


Figure 37: Search evidence from PubMed

The user need to access the website address: <https://pubmed.ncbi.nlm.nih.gov/>

In the search box the user searches for the syntax: <gene symbol<biomarker>. Then click the Search button.

CHRD: reported as biomarkers for heart failure by a study with PubMed ID35346191.

NOG: reported as potential biomarkers of atrial fibrillation-related stroke by a study with PubMed ID30760287.

BMPR1A: reported as biomarker of dedifferentiated liposarcomas by a study with PubMed ID27114889.

## Step 5: Biological function analysis

Because genes are in symbol form, users need to access the website address to convert the gene symbol to Entrez ID by accessing the address: https://www.syngoportal.org/convert.html

After having Entrez ID, the user needs to access the below website address for biological functional analysis:

* <https://david.ncifcrf.gov/tools.jsp>

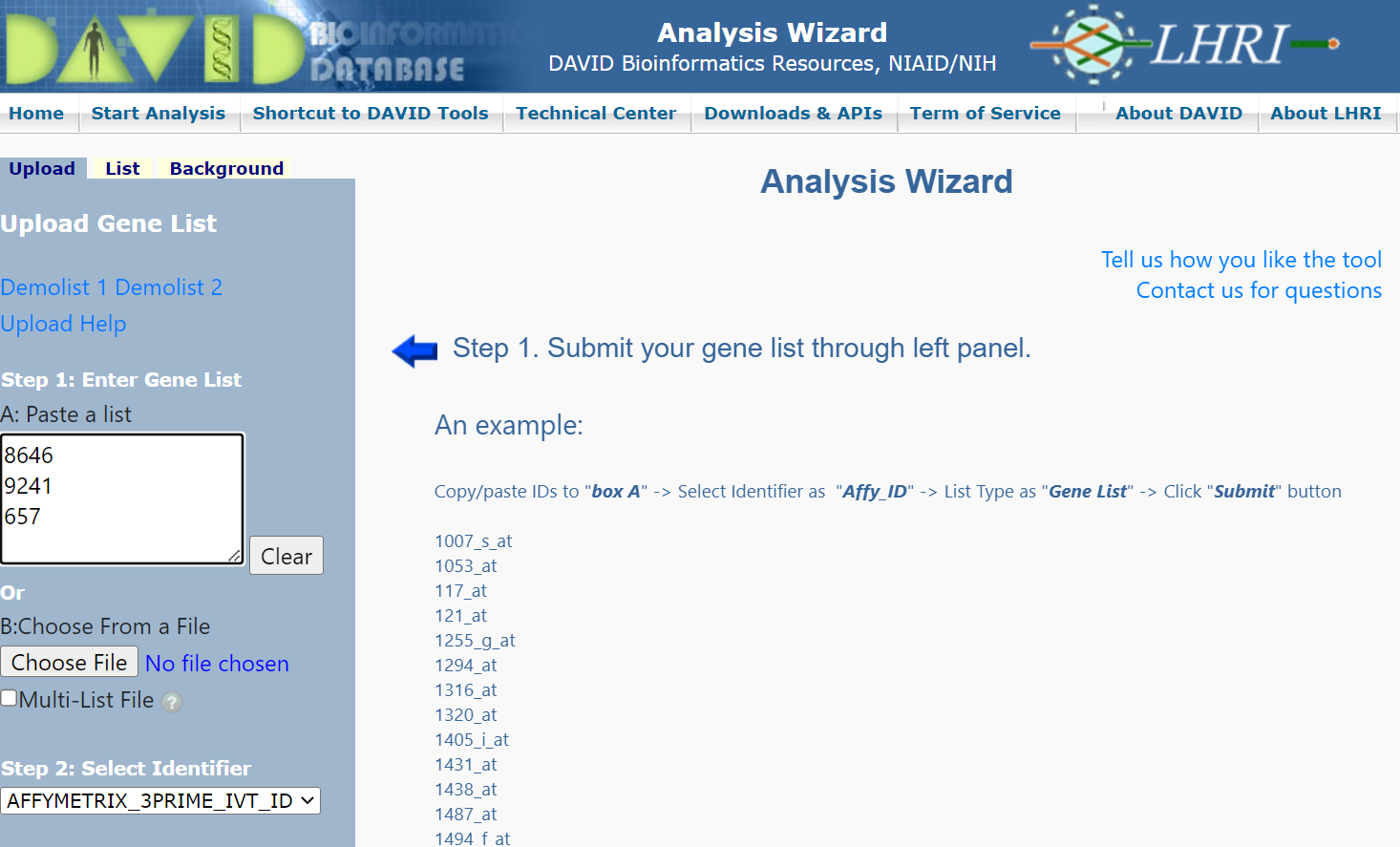


Figure 38: Search Biological function analysis

(1) Paste a list: ID of top 3 genes

(2) Click on the combobox and select value: Entrez\_gene\_id

(3) Click on the radio button and select value: Gene list

(4) Click on the button Submit List

**Human signaling network**: Three candidate biomarker genes, including: CHRD, NOG, BMPR1A share the same biological functions as follows:

TGF-beta signaling pathway: The transforming growth factor beta (TGF-β) proteins are cytokines that are critical during the development and homeostasis of somatic tissue and whose downstream signaling regulates tumor progression. View our interactive TGF-β signaling pathway and find the products that will help your research.

Dorsal/ventral pattern formation: The regionalization process in which the areas along the dorsal/ventral axis are established that will lead to differences in cell differentiation. The dorsal/ventral axis is defined by a line that runs orthogonal to both the anterior/posterior and left/right axes. The dorsal end is defined by the upper or back side of an organism. The ventral end is defined by the lower or front side of an organism.

Glycoprotein: Glycoproteins are proteins which contain oligosaccharide chains (glycans) covalently attached to amino acid side-chains. The carbohydrate is attached to the protein in a cotranslational or posttranslational modification. This process is known as glycosylation. Secreted extracellular proteins are often glycosylated. Glycoproteins are also often important integral membrane proteins, where they play a role in cell–cell interactions. It is important to distinguish endoplasmic reticulum-based glycosylation of the secretory system from reversible cytosolic-nuclear glycosylation.

1. **Reference**

[1] X. Liu and L. Pan, "Identifying Driver Nodes in the Human Signaling Network Using Structural Controllability Analysis," (in eng), *IEEE/ACM Trans Comput Biol Bioinform,* vol. 12, no. 2, pp. 467-72, Mar-Apr 2015.