Chapter 4
Common Network Pharmacology Software

Xuezhong Zhou, Ning Wang, and Xing Zhai

Abstract As a growing trend in current pharmacology research and an important medical application in network science research, network pharmacology has become an indispensable complement to traditional pharmacology research with the immense accumulation and integration of large-scale pharmacology and disease molecular network data. In addition to emerging new methods and technologies, a large number of analytical techniques and methods, such as network analysis and molecular functional analysis have matured into related software or programs and are available for researchers to use for free or available as open source, which is an important factor and condition that has helped network pharmacology's robust and rapid development. From the perspective of information processing and computer systems, this chapter classifies and introduces commonly used network pharmacology software by refining the overall functional flow of network pharmacology-related software or programs and illustrates demonstrative application examples by combining actual data processing, analysis, and visualization operations. The practical steps and contents in this chapter, combined with the theoretical analysis, analysis methods, and research cases of network pharmacology in other chapters, can provide researchers or students with relevant software tools and practical operation methods that can be used for reference, as well as provide rapid and convenient software tool selection and practical guidance for actual research on network pharmacology.

X. Zhou (✉) · N. Wang
School of Computer and Information Technology, Beijing Jiaotong University, Beijing, China
e-mail: xzzhou@bjtu.edu.cn; wang_ning@bjtu.edu.cn

X. Zhai
Beijing University of Chinese Medicine, Beijing, China
e-mail: zhaix@bucm.edu.cn

S. Li (ed.), *Network Pharmacology*, https://doi.org/10.1007/978-981-16-0753-0_4

4.1 Software Functional Framework and Classification of Network Pharmacology

4.1.1 Overall Software Functional Requirements

Network pharmacology [1] uses network relationship data such as drug-target relationships, interaction group networks, and phenotype–genotype associations, with the goal of analyzing the regulatory role of drug intervention in molecular networks and using corresponding data analysis models and methods such as complex network, machine learning, and molecular functional analysis, to determine the research direction for the interpretation and discovery of drug molecular mechanisms. The focus of pharmacological research is to find the target of drugs and how to regulate the corresponding targets to achieve the effect of disease treatment. Network pharmacology research focuses on discovering and confirming the multi-target effect of drugs and their network pharmacodynamic mechanisms and analyzing and discovering the systemic therapeutic effect of drugs and their combination on diseases, by analyzing the overall effect of network regulation.

In general, classic network pharmacology research cases involve main links such as network data collection and integration, network structure analysis and prediction, molecular and network function analysis, drug-target (target) relationship analysis, drug interaction or combination analysis, and drug indication analysis. The first three links are the common steps and methods in classic network pharmacology research, while the last three links are typical tasks for specific applications. The core data involved in relevant links are described in detail in other chapters of this book, including the clinical efficacy information of drugs, drug composition structure and its interaction, drug-target relationship, interaction group network, phenotypic genotype association, drug side effects, and drug indications.

Network data collection and integration: Targeting research problems of specific drugs' action mechanism, the lack or limitation of main target relationship data, such as drug-target relationship, is relatively common. In addition to generating the corresponding target data through wet tests, network pharmacology research often uses automatic data extraction or generation methods to collect the target relationship data. At the same time, the integration of network data from different sources (medical literature, structured database, etc.) and different types (drug-target relationship, drug side effect relationship, disease gene relationship, etc.) is often used to integrate network data resources for specific research objectives. For example, as the core research task of network pharmacology, the prediction of drug-target relationship usually adopts relatively single data types in the initial stage, such as drug chemical structure information, drug-target relationship information, etc. However, due to the incompleteness and complementary characteristics of various types of information, in recent years, use of multi-source network data integration to construct the basic network pharmacology data resources has become a guiding research direction. In the following foreseeable future, network data integration will become

the basic research method and foundation. Therefore, network data integration is one of the basic functional requirements of network pharmacology methods or software.

Network structure analysis and prediction: From the perspective of network science or complex network [2, 3], network pharmacology is a classic application of complex network in the pharmacology field, and its wider medical applications can be considered as network medicine [4]. Therefore, analysis methods and models based on complex networks such as node or edge centrality measure, shortest path, link prediction, and community analysis [5, 6] (community detection), and various statistical graph generation models (such as random graph [7], small world network [8], and scale-free network [9]) are the main supporting analysis methods in network pharmacology. For example, the problem of edge prediction using the adjacent structure of the network, the path connection mode of the network, or the attribute information of the node is called link prediction. Its most direct application is drug-target prediction (to determine whether the relationship between a specific drug node and a target node exists). The direct application of community analysis with relatively dense internal connections and relatively sparse external sub-network structures obtained from the overall network is the discovery and confirmation of the disease module [10] or drug-target module. Therefore, the above two methods naturally become the core complex network analysis methods in network pharmacology. In addition, due to the above two kinds of network analysis problems, we can also directly model typical machine learning problems [11], such as the drug-target relationship problem can be regarded as an information recommendation problem [12], a binary classification problem of edge judgment, or a corresponding sorting learning problem [13], and the analysis of disease module can be regarded as a clustering problem of network data. Therefore, supervised learning methods [11] such as regression analysis, support vector machine, Bayesian network, and deep neural network in machine learning can be used in drug-target prediction. However, unsupervised learning methods, such as k-means, spectral clustering, and hierarchical clustering, can be applied to the discovery of disease modules. Moreover, all kinds of community analysis methods, such as graph partition-based method and modularity evaluation-based optimization method, can be considered as clustering methods based on network data. Common complex network analysis and even some machine learning software based on network data are tools and methods that can be used in network pharmacology research.

Molecular and network function analysis: Determining the specific biological functions of drug-targets and their molecular networks is an important task for analyzing drug molecular mechanisms and their pharmacodynamic effects. Therefore, systematic molecular function analysis methods have become an important technical means in network pharmacology to further explain the drug effects and pharmacokinetic mechanisms from the multi-tiered levels of molecules, cells, tissues, organs, and systems, as well as adverse drug reactions and side effects. Among them, gene ontology (GO) analysis is the main functional analysis at the gene or protein level, while molecular pathway analysis can be combined with corresponding pathway databases such as KEGG and Reactome for molecular

function analysis of metabolic pathways, signal transduction pathways, and protein complexes.

Analysis and prediction of drug-target relationship: Drug-target analysis is network pharmacology's core analysis task and goal. Judging from the types of drugs involved and the scope of research, it can be divided into two main research approaches. The first category focuses on drug-target discovery of specific drugs (or TCM compound prescriptions) or diseases. The purpose of this task is to identify novel binding relationship between drugs (corresponding small-molecule chemical components) and targets by means of virtual screening, manual compilation, and review of literature and information extraction or wet tests, using clinical efficacy or phenotype information of specific drugs, to form relatively reliable research results through the interaction information between the target in the molecular network and disease-related genes or biomarkers. The second category focuses on the R&D of large-scale drug-target relationship prediction methods using integrated network pharmacology data or drug association attributes. The first type of research is actually a case-based drug-target relationship study based on network pharmacology, which aims to analyze the mechanism of clinically effective drugs and prescriptions, provide an understanding and interpretation of pharmacological mechanisms, and provide a new record for drug-target relationship data resources. This kind of research is extensively practiced and has varied applications in the field of TCM network pharmacology. Especially in the research of the target of Chinese medicine compound prescriptions and their molecular network, this research type has generated practical results and research value. The second type of research aims at the development of new analytical methods and models, which is one of the core research tasks of network pharmacology. This method mainly includes two kinds of models: complex network analysis and machine learning. So far, researchers have implemented a variety of related drug-target prediction algorithms and models that promote the progress of network pharmacology research. Various algorithms and models for related drug-target prediction implemented by researchers have also promoted network pharmacology research development. In view of the significant performance advantages of deep representation learning and deep neural network models where there is sufficient data volume, the current algorithm and its software research and development has formed a tide and trend that is focused on the deep learning model.

Analysis of drug interactions and combinations: Drug interactions (drug–drug interaction) refer to the mutual influence and action between ingredients caused by simultaneous use of food, beverages, food supplements, and other drugs in the process of drug use. These interactions often lead to side effects and adverse reactions, but may also lead to beneficial medicinal effects [14]. Drug combination analysis refers to the analysis and study of the clinical situation of complex diseases such as co-diseases and concomitant diseases, as well as complex chronic diseases such as cancer and complex infectious diseases such as HIV. It is necessary to analyze and study the simultaneous listing and administration of multiple drugs for the same patient (especially the elderly), so as to find the best combination of drugs and identify the combination of drugs that cause serious side effects. Therefore,

given the widespread use of combination drugs (or even compound drugs), drug interaction analysis has become an important research direction, and drug interaction analysis is also an important cross-sectional research area in the R&D of combination drugs within the network pharmacology framework. The above two research tasks complement each other. As network pharmacology focuses on the multi-target and molecular network effects of drugs, drug interaction and combinatorial analysis have become important applications of network pharmacology methods, as they can help discover and confirm more systematic drug interactions and effective combination drugs. At the same time, network pharmacology research of TCM itself is a compound-oriented pharmacological research. The diversity of compound medicine ingredients makes the systematic research of drug interactions and combined drug mechanisms a research task and scientific problem that is both important and promising for breakthroughs. The discovery and conformation network effect index of optimal combination drugs [15] and even a compound network drug efficacy index that reflects the compatibility of TCM formulation is an important basic research task in TCM network pharmacology.

Analysis and prediction of drug indications: Drug indication analysis is the final goal of network pharmacology research, that is, determining the disease or clinical phenotype that can eventually be effectively treated by drugs. From the perspective of analysis methods, the analysis and prediction of drug indications and the very important drug repositioning (or drug repurposing) [16] issue in the context of new drug research and development are the same issue. For a given drug, to predict its total pharmacophore spectrum (disease or phenotype treated or acted upon), the novel pharmacodynamic phenotype is the target of drug redirection analysis [16]. In addition, in this sense, the side effects and adverse reactions of a drug can also be considered as an effector phenotype in the broad sense of the drug, but the side effects and adverse reactions of the drug are only an unexpected effector phenotype. In view of the concept of drug action based on extensive systematic data integration and network regulation, network pharmacology has natural advantages and characteristics in the overall analysis of drug indications. Therefore, drug redirection research based on network method and network pharmacology has become a widely recognized new idea and method in the research and development of new drugs.

4.1.2 Software Functional Framework and Classification of Network Pharmacology

In the previous section, the data processing and analysis requirements of network pharmacology research were briefly summarized. Network pharmacology research involves data resource collection and integration, network construction and analysis, drug-target relationship prediction, and other diverse methods and software functional requirements. From its functional framework and classification, it basically includes the data processing and analysis function module, as shown in Fig. 4.1. In

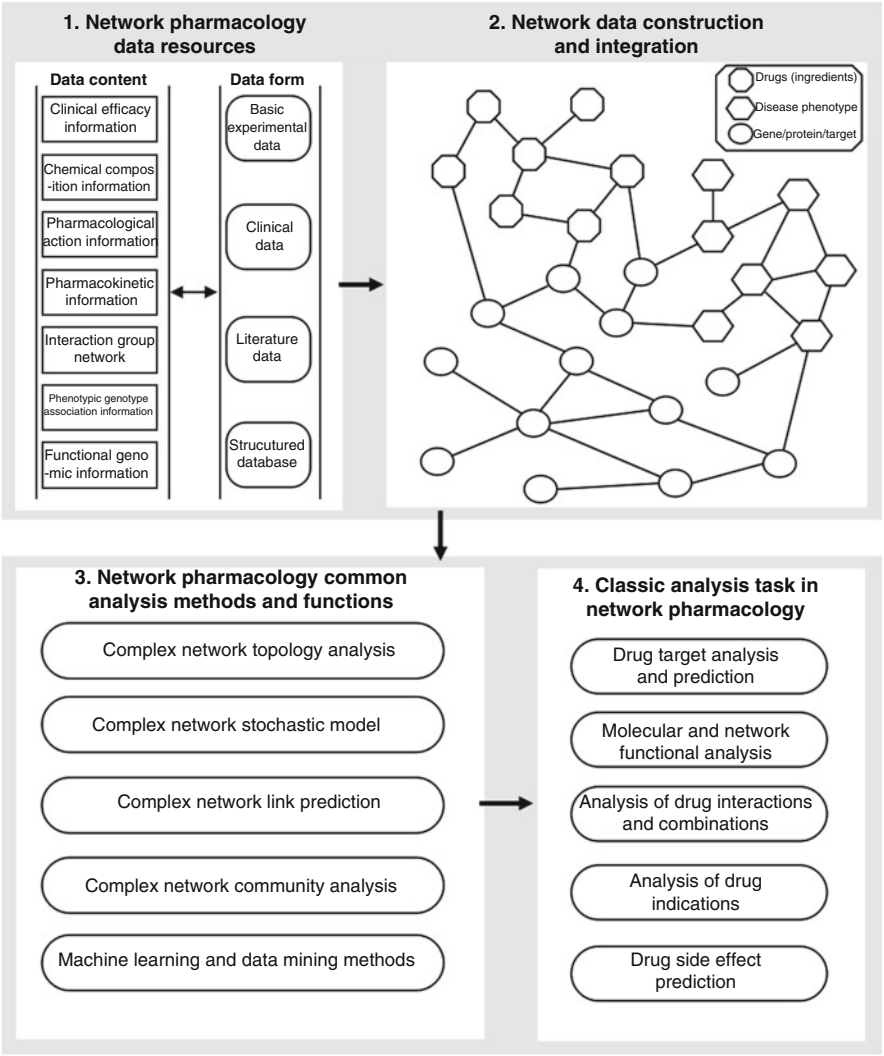


Fig. 4.1 Software functional module framework of network pharmacology

fact, the current research and development of network pharmacology methods and software research and development mainly focus on the following four aspects of functional requirements. For example, in terms of the formation and integration of network pharmacology data resources, a large number of network pharmacology databases have been constructed, such as DrugBank [17], STITCH [18], SIDER [19], PubChem, etc., as well as a large number of high-quality databases of functional genomics and interaction groups. Specifically, in the field of TCM network pharmacology, a database resource platform, including the relationship of TCM-chemical ingredients-targets has also been constructed, which is of immense

help in the research and development of TCM network pharmacology. In addition, drug-target prediction methods and online software for specific diseases such as rare diseases, psychiatric diseases, and cancer, drug interaction prediction, drug combination analysis, drug redirection, and drug side effect analysis software are seeing rapid growth and development. At the same time, as complex networks, machine learning software, and programming language for big data analysis (such as Python) mature further, the applications in the biological field can provide strong technical support for network pharmacology research. The subsequent chapters illustrate and introduce typical methods, software, and practical programming operations related to the above aspects.

4.2 Online Software Commonly Used in Network Pharmacology

Based on the requirements of network pharmacology analysis introduced in the previous section, it can be seen that drug-target and drug indications are important applications. The research and development of convenient and fast online software is an important means to promote drug-target analysis, indication analysis, and other pharmacology research, especially for researchers who are new to network pharmacology technologies and methods. At present, several excellent online analysis tools have been developed for researchers. This section introduces the analysis tools from the perspective of established online software.

4.2.1 Online Software for Drug-Target Prediction

The design and development of new drugs has always been a complex, expensive, and time-consuming process. Moreover, the success rate of new drug research and development is quite low. Usually, only a few drugs can finally pass the FDA evaluation every year and be commercially available for treatment. Therefore, drug research faces problems of low drug development efficiency, rising demand for treatment, and serious shortage of existing therapeutic drugs. The determination of drug-target relationship is an important link in the development of new drugs; however, the screening method based on wet tests is still extremely challenging and difficult, therefore, drug-target prediction analysis is a hot research topic. Teams from research institutions and scientific research institutes around the world have made extensive research and contributions in this respect and have developed various computational models to predict potential drug-target relationships on a large scale. The prediction analysis methods introduced in previous chapters are mainly based on algorithms. In addition, there are also convenient and practical

web-based service tools that can provide online drug-target prediction services, such as DINIES [20], SuperPred [21], and SwissTargetPrediction [22].

DINIES (Drug–Target Interaction Network Inference Engine based on Supervised Analysis) is an online platform that is used to infer potential drug–target interaction networks. DINIES can accept a variety of input data, such as chemical structures, side effects, amino acids, or protein domains. In addition, each dataset is converted into a nuclear similarity, and multiple state-of-the-art machine learning methods are used to predict the drug–target interactions.

SuperPred is an online platform used to predict the structure of small molecular targets. In SuperPred, drug-target prediction is based on similarity distribution through four input options (including the name of the compound searched in the PubChem database, the structure of the compound created through Simplified Molecular Input Line Entry Specification (SMILES), the structure diagram drawn with ChemDoodle, and the uploaded molecular file) to estimate individual threshold value and probability of a specific target.

SwissTargetPrediction is an online platform that is used to infer bioactive small molecular targets based on the two-dimensional and three-dimensional similarity values of known ligands. In addition, it can provide prediction results for five different biological tissues (human, house mouse, rat, cattle, and horse).

We use SwissTargetPrediction as an example to demonstrate the specific operation. First, the user can customize the species to be analyzed (in the case of humans, select homo sapiens) (as shown in Fig. 4.2); then, the user can enter the molecular structure of the compound and search for the specific SMILES string of the compound as the input on the ChEMBL website. Here, we take GINSENOSIDE RG1 as an example (as shown in Fig. 4.3); finally, the user can click the "Predict targets" button; the platform runs the calculations and finally the corresponding target prediction analysis results are displayed on the prediction interface (as shown in Fig. 4.4).

4.2.2 Online Software for Drug Indication Analysis

Indication is the phenotypic spectrum of the disease treated by drugs. The main goal of drug indication prediction is to establish the relationship between drugs and the indication spectrum, that is, to determine what is the complete spectrum of disease phenotypes that a specific drug can treat. In view of the different granularity of disease classification, there are two problems in the analysis of drug indications: Optimization of the classification of diseases that have been treated and prediction of new diseases. The view that predicted disease phenotype spectrum contains new major diseases is a widely studied drug repurposing or drug repositioning problem. The drug repositioning method has been successfully applied to the R&D of a variety of disease treatment drugs [23], which can shorten the time of drug R&D and reduce the cost and risk of drug R&D. Drug repositioning can not only expand the application scope of drugs and extend their service life, but also enable the reuse

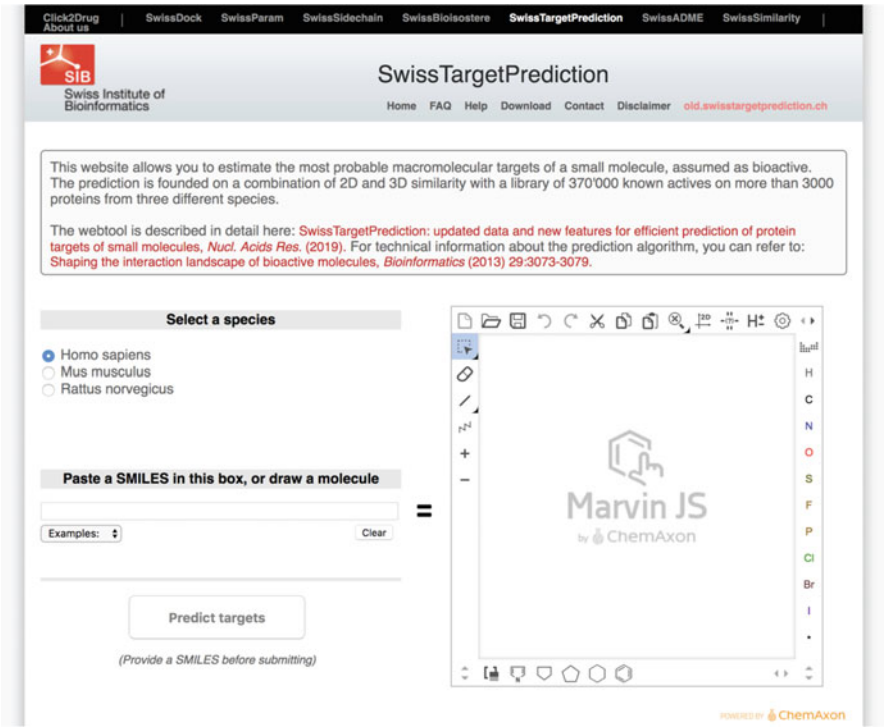


Fig. 4.2 SwissTargetPrediction front-page interface

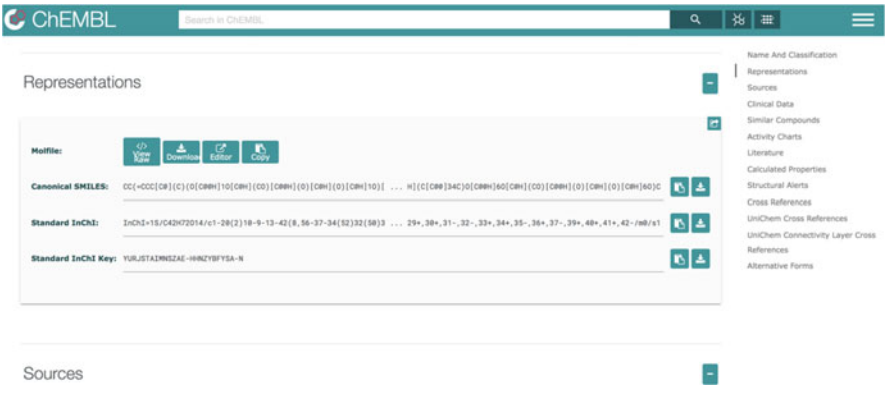


Fig. 4.3 ChEMBL search page

of withdrawn drugs. For example, the original intention of developing sildenafil was to treat cardiovascular diseases such as angina pectoris and hypertension, but it was unexpectedly found in clinical tests that it can be used to treat male erectile dysfunction [24]. Subsequent studies have shown that low doses of sildenafil can

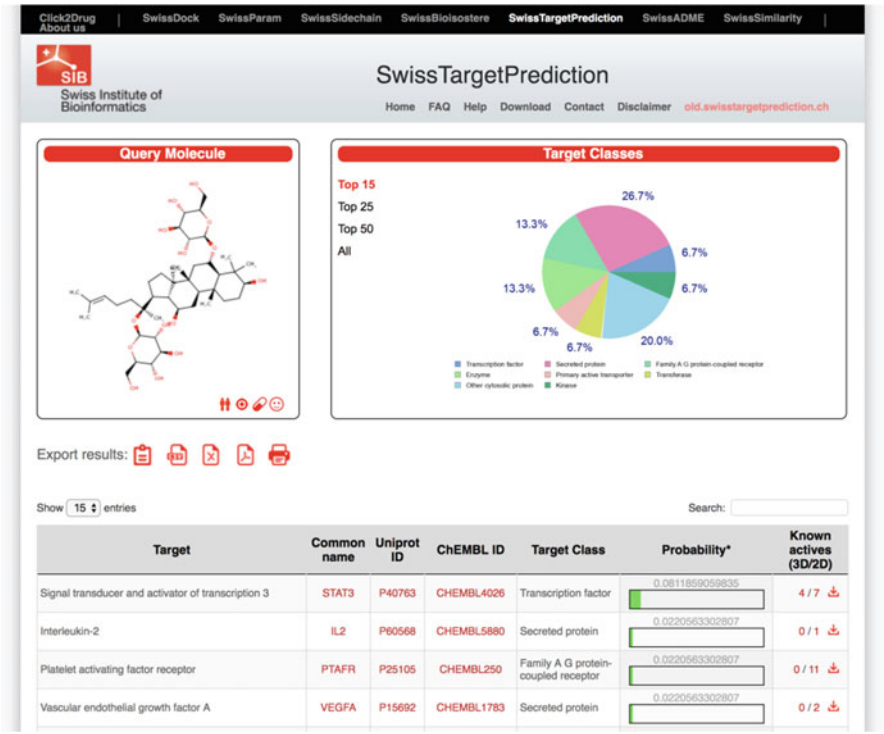


Fig. 4.4 Prediction result page

also be used for the treatment of pulmonary hypertension in rare cases [24]. The discovery of new uses for the above mentioned known drugs is mostly accidental, and not the result of rational design. Due to the large number of types of diseases and the number of known drugs, the cost of screening new uses of known drugs through experiments is still quite high. With the accumulation of omics data and the rapid development of various drug-related databases, such as DrugBank [17] and SIDER [19], drug repositioning prediction by computational methods has become a hot topic in computational biology and systems biology research in recent years [23]. The rational design of the clinical research scheme of drug repositioning assisted by computational methods can provide clues for large-scale experimental screening, further reduce the cost, and make drug repositioning enter the stage of combining rational design and experimental screening.

In recent years, there has been a growing trend in software R&D related to drug indication analysis, such as MeSHDD [25] and RE: fine Drugs [26]. Using the above software, researchers can analyze the properties of existing drugs through online methods to determine whether related drugs can be safely and effectively applied to specific diseases. The next section introduces a typical online tool: MeSHDD [25].

MeSHDD clusters drugs based on the drug–drug similarity of the Medical Subject Heading (MeSH) and then predicts new indications of the drug. Specifically,

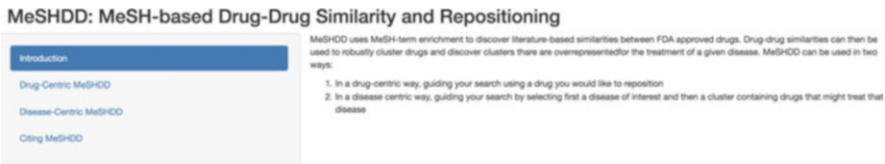


Fig. 4.5 Homepage of MeSHDD's official website

MeSHDD uses hypergeometric distribution to calculate the degree of co-occurrence of drug terms in MeSH and performs Bonferroni correction. Then, the drug–drug similarity is calculated by converting the above calculation result (represented by the P value) to the bit-by-bit distance obtained by the binary representation. Finally, pairing distance and clustering method are used to cluster the drugs, and the enrichment of disease indications is evaluated across multiple categories by comparing with data from TTD. In the verification experiment conducted by the author, it can be seen that MeSHDD can infer the indications for cystic fibrosis of antidiabetic drugs. The specific operation is as follows. First, navigate to the homepage of the official website (as shown in Fig. 4.5) and select the drug to be redirected from the drop-down list on the drug-centered page. Take quinine as an example, the indications corresponding to the drug can be obtained (as shown in Fig. 4.6), and related similar drugs can also be obtained (as shown in Fig. 4.7).

4.2.3 Online Software for Gene Function Enrichment Analysis

Several related gene expressions and interaction group data generated by high-throughput sequencing can provide abundant functional data resources for phenotypic genotype association research, however, they also put forward new requirements for efficient molecular function analysis. Enrichment analysis [27] is the main method to determine the common biological mechanism and medical phenotype association of batch differences or related genes by leveraging existing databases of gene function attributes, phenotypic genotype association data, and interaction group databases (such as molecular pathway database). According to the different related molecular function data used, enrichment analysis is mainly divided into GO enrichment analysis, pathway analysis, and differential gene enrichment analysis. Through gene function enrichment analysis, it is possible to discover the key biological pathways in the biological process in which gene sets are involved, which is an important analysis link in exploring the common rules from the complex omics data.

In short, gene enrichment analysis involves finding gene sets with certain gene functional characteristics and biological processes in a group of genes, which are often used in the follow-up analysis of differentially expressed genes and screened

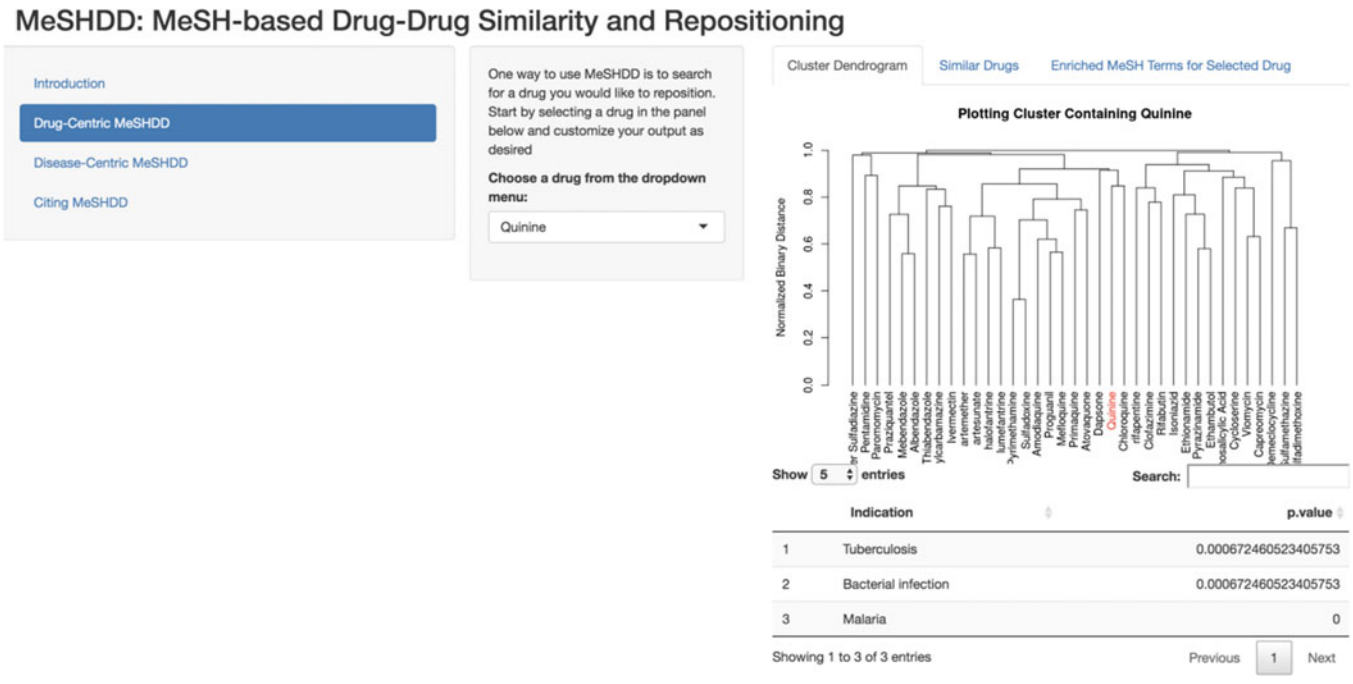


Fig. 4.6 MeSHDD prediction drug indications page

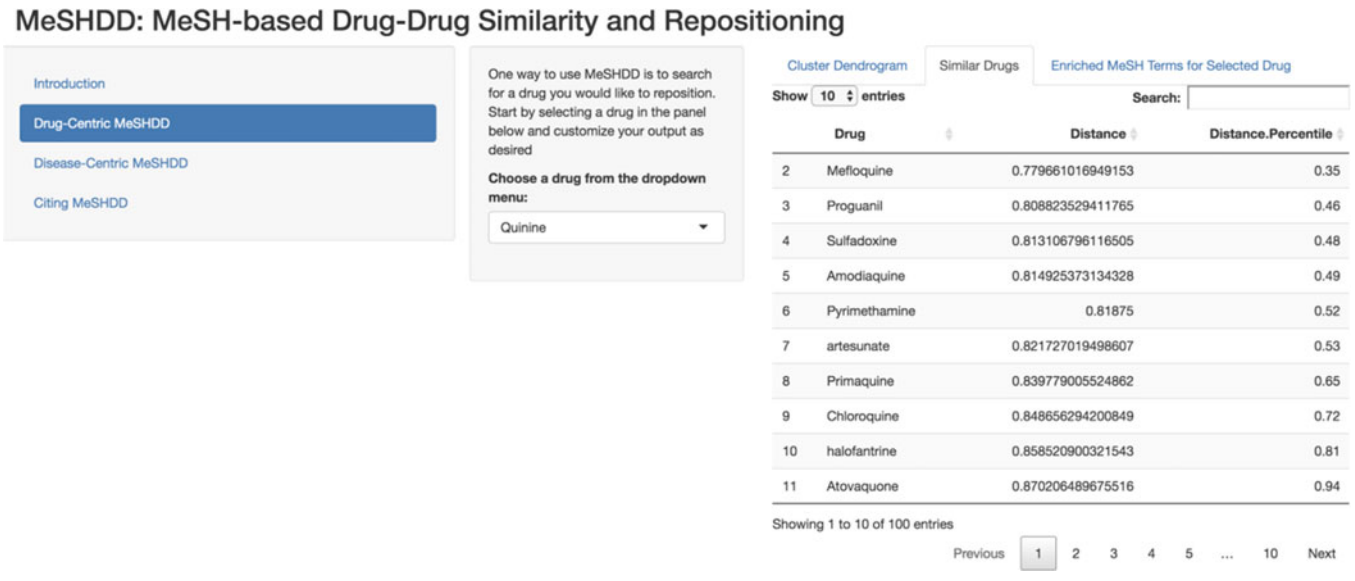


Fig. 4.7 MeSHDD prediction-related drug page

genes. At present, there are nearly 100 kinds of enrichment analysis tools developed by different research institutes. At present, many open source websites have integrated GO enrichment and KEGG pathway analysis functions, such as DAVID [28], KOBAS [29], and STRING [30]. In this section, we introduce DAVID, a well-known and commonly used enrichment analysis tool.

DAVID is taken as an example to conduct GO enrichment and pathway analysis for a given gene set. The homepage of the website is shown in Fig. 4.8. Step 1: Navigate to the Start Analysis page and input the gene set to be analyzed under Enter Gene List. Select Affy_ID under Select Identifier, then select the Gene List in List Type, and click the Submit List button, as shown in Fig. 4.9. Step 2: Select homo sapiens corresponding to the gene set under Select Species, then press the Use button under Select List and then click Functional Annotation Chart to initiate the analysis, as shown in Fig. 4.10; Step 3: Select the content to be analyzed, as shown in Fig. 4.11; select Gene_Ontology and Pathways, and click Functional Annotation Chart to display the analysis result, as shown in Fig. 4.12. The corresponding GO enrichment analysis and KEGG path analysis results can be obtained at the bottom of the analysis page, and the above analysis results can be downloaded by clicking the appropriate button.

4.2.4 Online Software for Constructing Protein Interaction Network

Proteins and their interactions are the pillars of cellular mechanism. Proteins are important macromolecules that constitute organisms and regulate a large number of basic life activities and biological behavior of organisms [31]. The protein interaction network is composed of individual proteins and their interactions, which can participate in all aspects of life processes such as biological signal transmission, gene expression regulation, energy and substance metabolism, and cell cycle regulation [32]. In network pharmacology correlation analysis, the protein interaction network is often used in drug-target and gene enrichment analysis and other studies and is of great significance for understanding the working principle of proteins in biological systems, the reaction mechanism of biological signals, the energy substance metabolism, as well as the functional connection between proteins. At present, there are many databases that provide protein interaction relationships, such as STRING, MINT [33], and BioGRID [34]. In this section, we introduce the STRING database that is well known and commonly used in research.

Currently, the STRING database has been updated to version 11, which includes known and predicted protein interaction relationships. The database contains 5090 species, 24.58 million proteins, and 3123.05 million protein interactions. The interaction relationships are derived from high-throughput experiments, text mining, other database data, and bioinformatics prediction data.

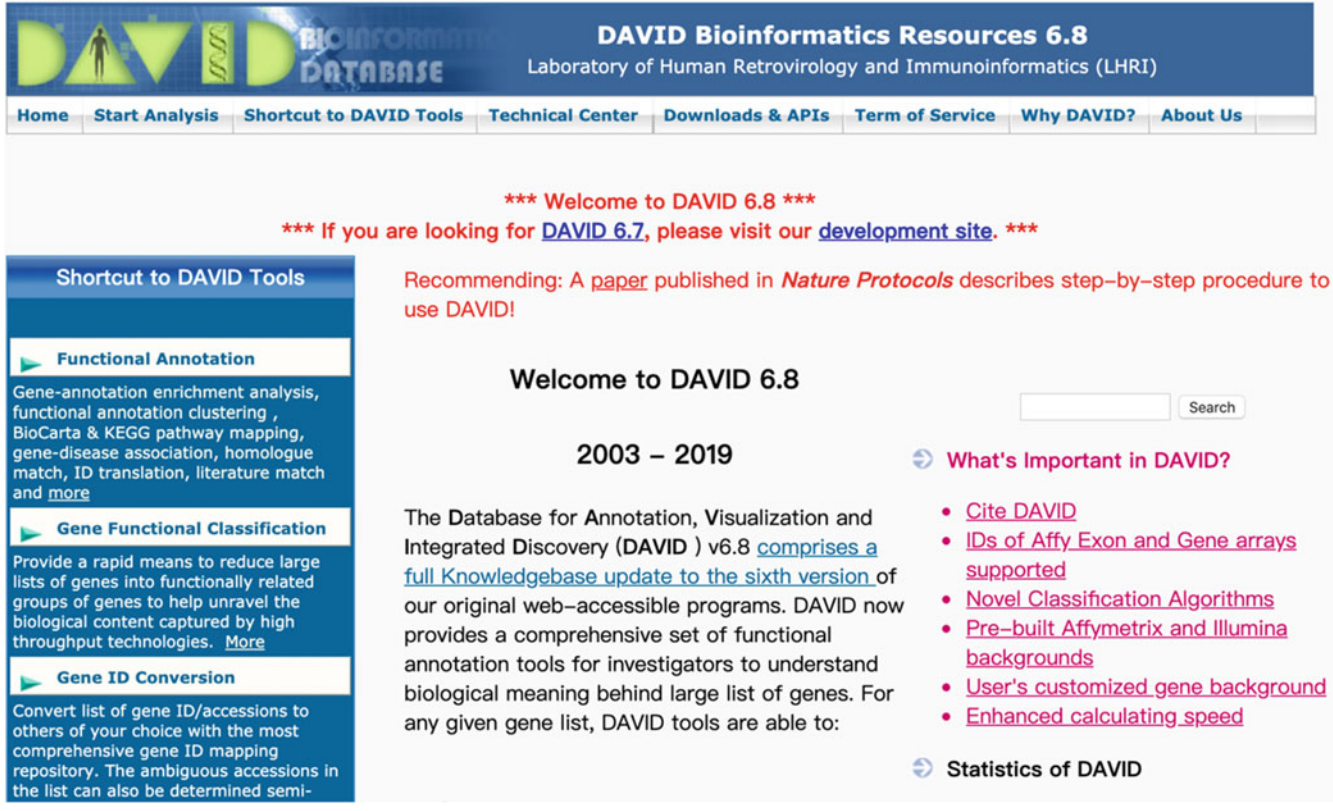


Fig. 4.8 Homepage of DAVID

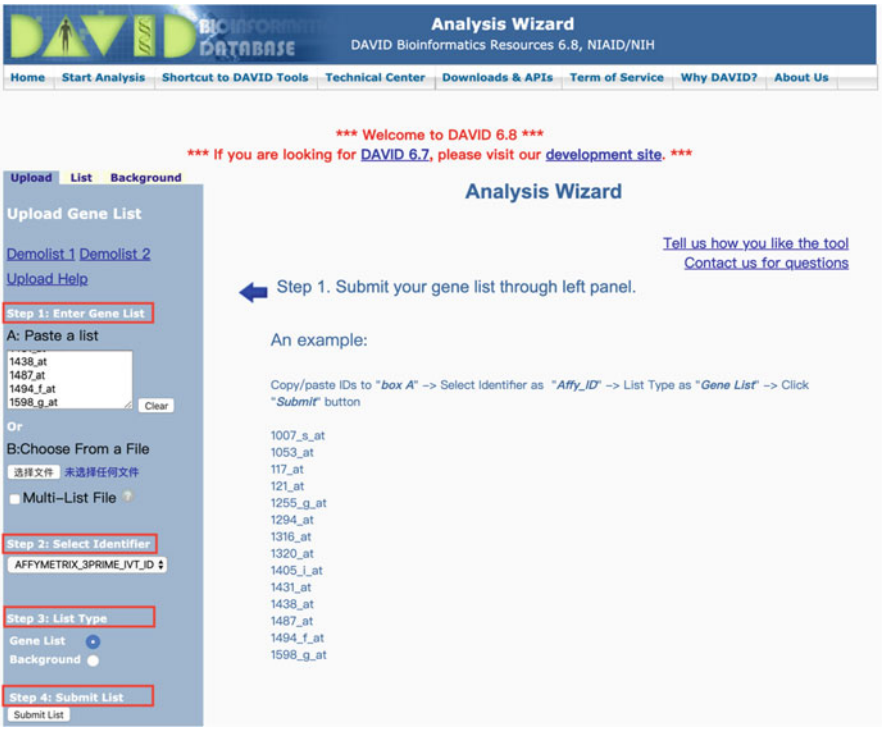


Fig. 4.9 Input data interface display

Users can query a single protein or a collection of multiple proteins. A sample operation for querying a single protein is as follows: Navigate to the website's homepage; the default page has the query tool for a single protein. Enter the gene to be analyzed (CASP3 as an example) into the Protein Name text box, select Homo sapiens under Organism, and then click the Search button, as shown in Fig. 4.13. Click the Continue button in the next page (as shown in Fig. 4.14) to display the analysis results; the protein interaction relationship related to the input gene can be obtained, as shown in Fig. 4.15. You can also click the Exports button to download the corresponding analysis results, as shown in Fig. 4.16.

A sample operation for querying a collection of multiple proteins is as follows: Navigate to the website's homepage and click the Multiple Proteins button. Then, enter the gene set or gene list to be analyzed into the List Of Names text box, select Homo sapiens under Organism, and click the Search button, as shown in Fig. 4.17. Click the Continue button in the next page (as shown in Fig. 4.18) to display the analysis results; the protein interaction relationship related to the input gene set is displayed, as shown in Fig. 4.19. Similarly, you can also click the Exports button to download the corresponding analysis results, as shown in Fig. 4.20.

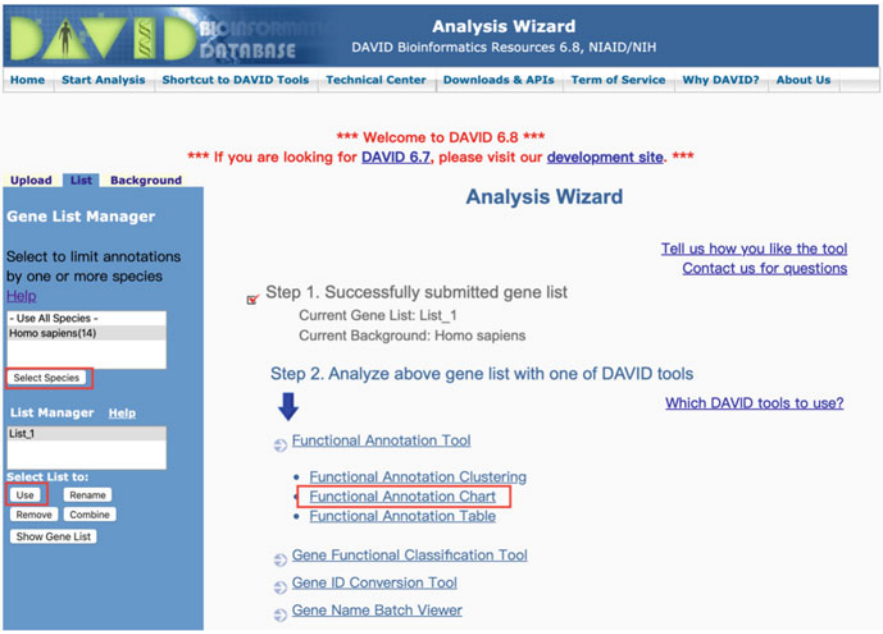


Fig. 4.10 Interface display for selecting and analyzing species

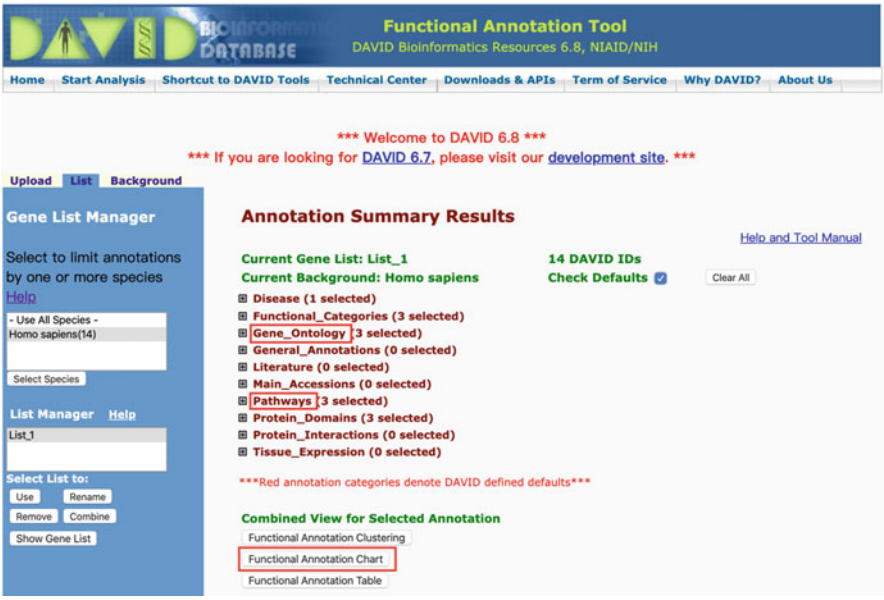


Fig. 4.11 Interface display for selecting and analyzing contents

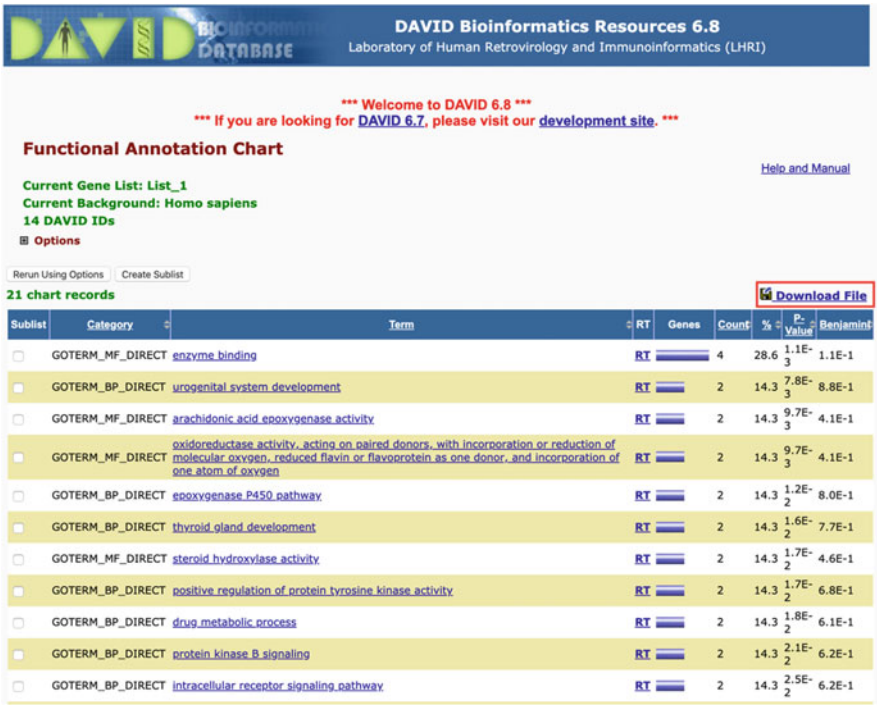


Fig. 4.12 Analysis results interface display

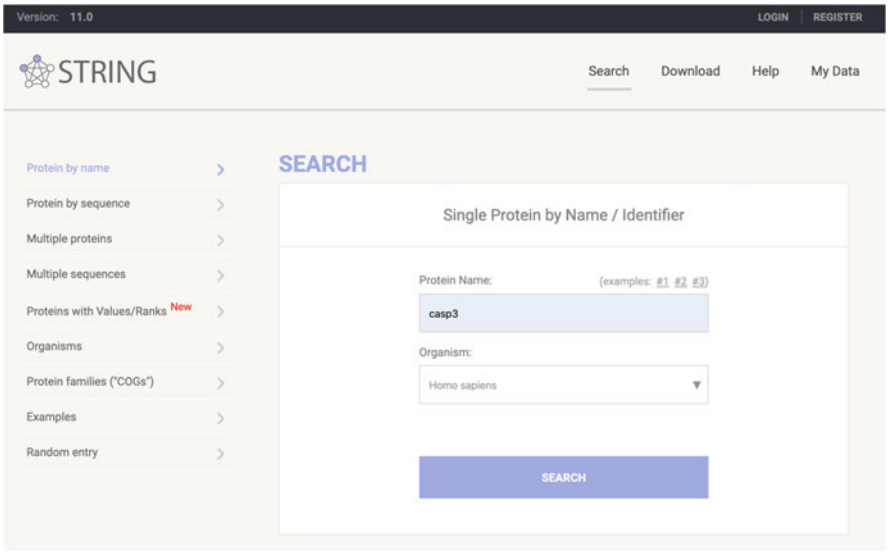


Fig. 4.13 Single gene query homepage

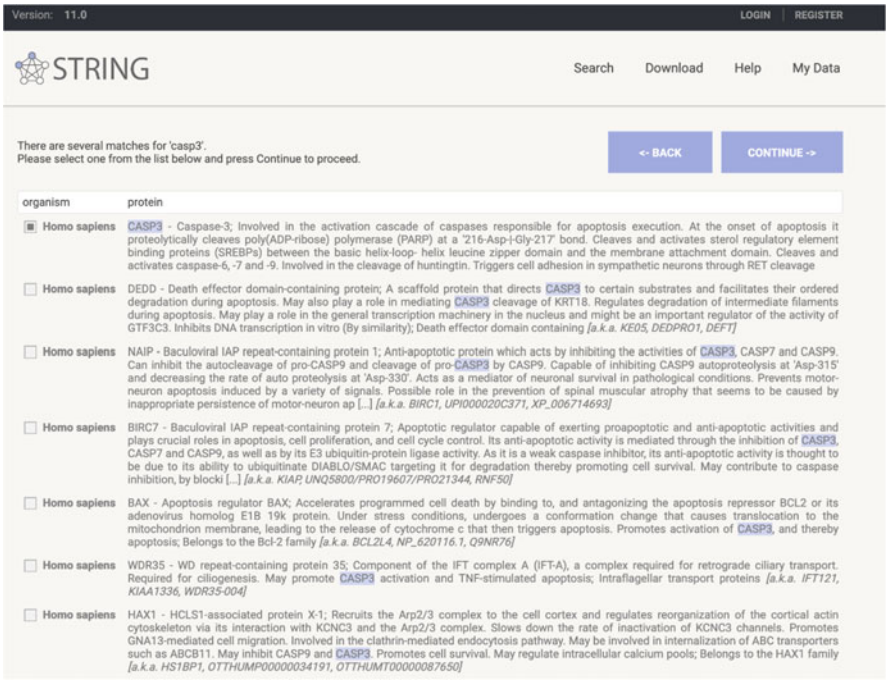


Fig. 4.14 Single gene query information confirmation page

4.3 Software Based on Graphical Interface Operation

Based on the network pharmacology analysis requirements introduced in the first section, we know that a complex network is one of the important methods. A complex network is not only a formal tool but also a scientific research method. Due to its universality for solving problems in various fields, it has been widely used in the fields of medicine, sociology, physics, information science, and ecology. At present, the accumulation of network data in various fields, such as protein interaction network [35], disease relation network [36], social network [37], power network, aviation network, and transportation network, has further promoted research on complex network methods. For example, in social network research, we study the law of group behavior [38] and the law of information dissemination [39] by constructing a social network, whereas in the biomedical field, we study drug interaction [40] and drug-target relationship [41] by using complex network methods. At present, a large amount of network data has a large scale and many network nodes and edges. Therefore, it is necessary to rely on visual network analysis methods to obtain effective results. In view of this, researchers have developed several excellent visual network analysis tools, including visual software based on graphical interfaces and software that can be programmed to call the package (For example: Python package, R language package, Java package, etc.),

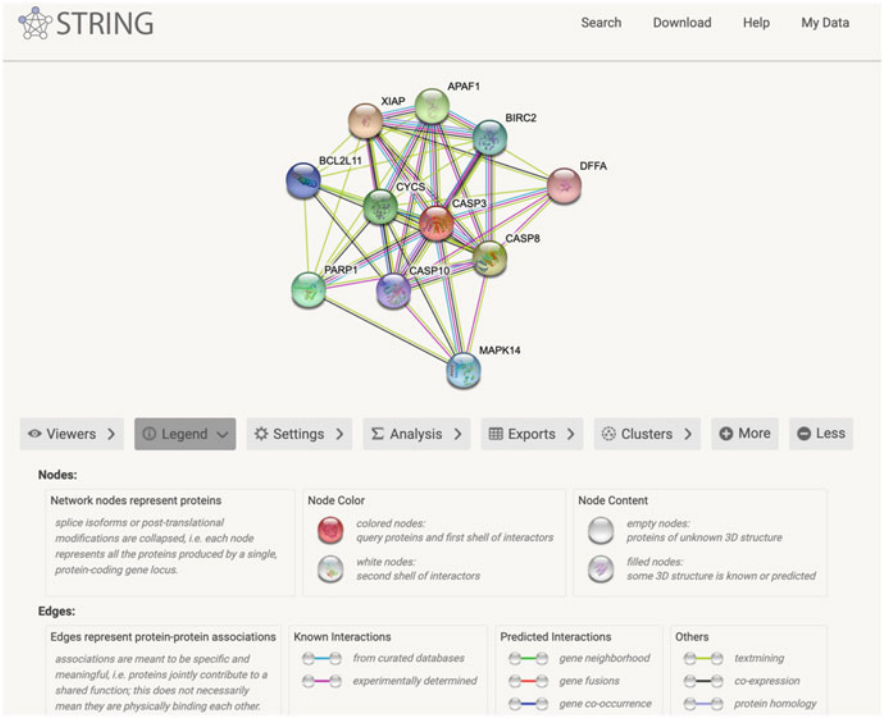


Fig. 4.15 Single gene query result page

among them, a visualization tool based on a graphical interface is easy to install and operate and is more intuitive to operate than a programming language package. Next, this section briefly introduces and demonstrates an independent system software from two aspects: differential gene enrichment analysis and network analysis.

4.3.1 Differential Gene Enrichment Analysis Software

The GO function and KEGG pathway enrichment analysis introduced in the previous section aims to discover the characteristic molecular function and pathway information of the identified gene set. In addition, another type of enrichment analysis is mainly used to identify differential genes for specific conditions such as phenotypes. For example: Gene Set Enrichment Analysis (GSEA) [42] is a widely used method that can be used to assess the distribution trend of genes in a gene set in phenotypic correlation ranking and determine their association with a specific phenotype. Different from KEGG pathway analysis, GSEA considers the influence of genes with little expression difference but important functions, on the pathway, and compared with KEGG pathway analysis, it can retain more relevant information.

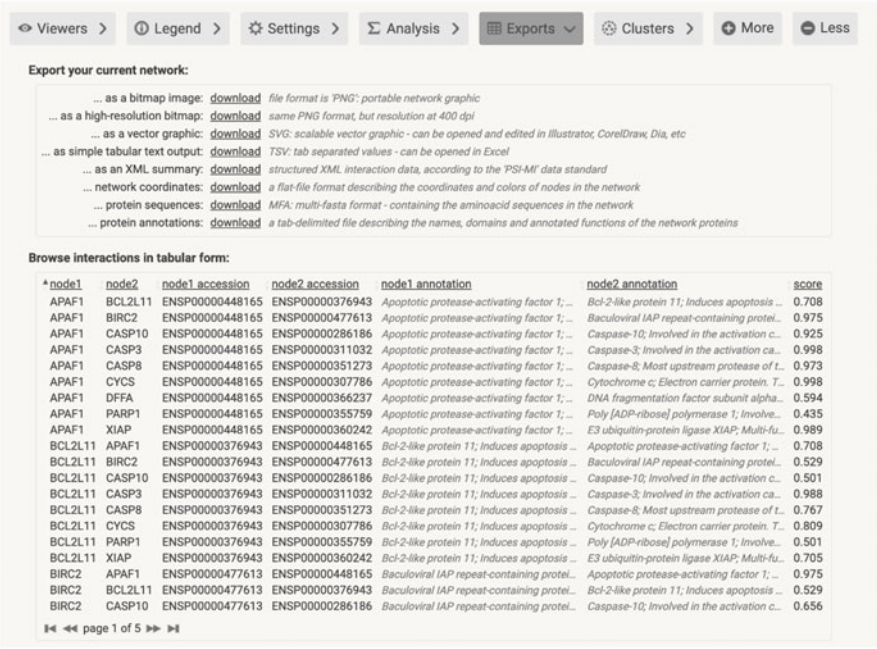


Fig. 4.16 Single gene query download page

The GSEA algorithm and software were developed by the Broad Institute in the USA. The installation and analysis process of GSEA software is introduced below.

4.3.1.1 Software Installation

Two methods are officially recommended. The first is the Java-based GSEA desktop application. Navigate to the GSEA official download page and click the Launch icon on the right to download (as shown in Fig. 4.21), however, the installation requires an internet connection. The second type is the Java-based GSEA application package. Click download on the right to download. The installation does not require an internet connection and starts quickly. The software startup interface is shown in Fig. 4.22. We use the second method as an example in the following analysis and introduction.

4.3.1.2 Data Preparation and Import

GSEA provides the sample dataset on its official website, as shown in Fig. 4.23. Users can download a selected gene expression matrix file and sample grouping information file for analysis.

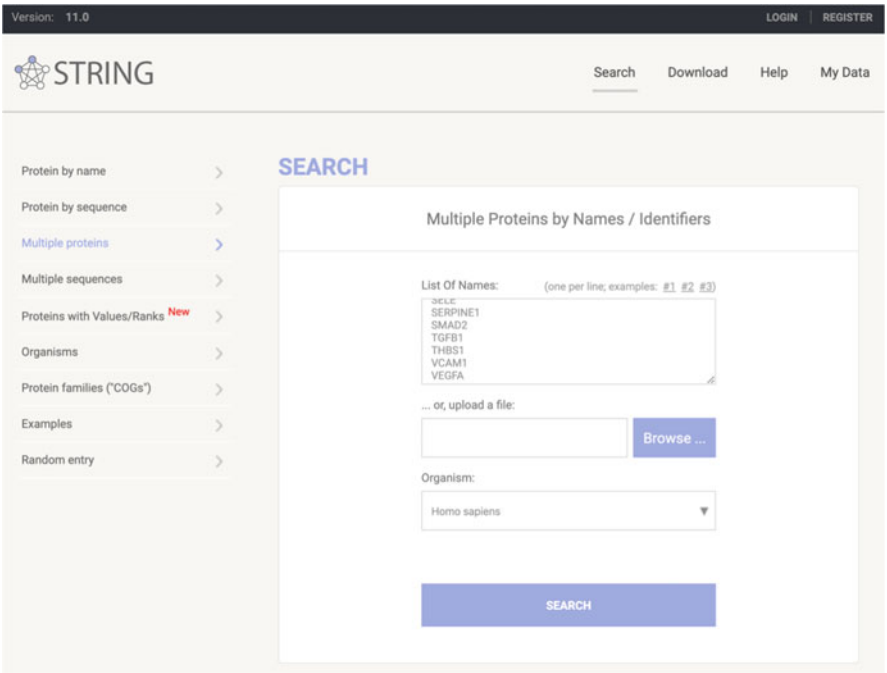


Fig. 4.17 Multiple gene query homepage

The gene expression matrix Diabetes_collapsed_symbols.gct, sample grouping information Diabetes.cls, and gene function classification data c5.all.v6.2.symbols.gmt provided by the GSEA website are selected here as an example. According to the steps shown in Fig. 4.24, click Load data— > Browse for File— > find the file to be imported in the pop-up box, select and click open to import data.

4.3.1.3 Setting Parameters and Running the Software

Click Run GSEA on the left panel of the interface; the parameter selection bar pops up. Parameter settings are divided into three parts: Mandatory parameter settings, basic parameter settings, and advanced parameter setting. Generally, the latter two parameters do not need to be modified, and the default value can be used. The following is a brief description of the fields in the mandatory parameter settings (as shown in Fig. 4.25).

Select the expression dataset file Diabetes_collapsed_symbols.gct in the Expression dataset field. Select the gene function set database c5.all.v6.2.symbols.gmt in the Gene sets database field. Number of permutations indicates the number of permutation tests, and the default value is 1000. Select the comparison method in the Phenotype labels field. GSEA automatically extracts the corresponding data from the expression dataset file for comparison based on the group information during the

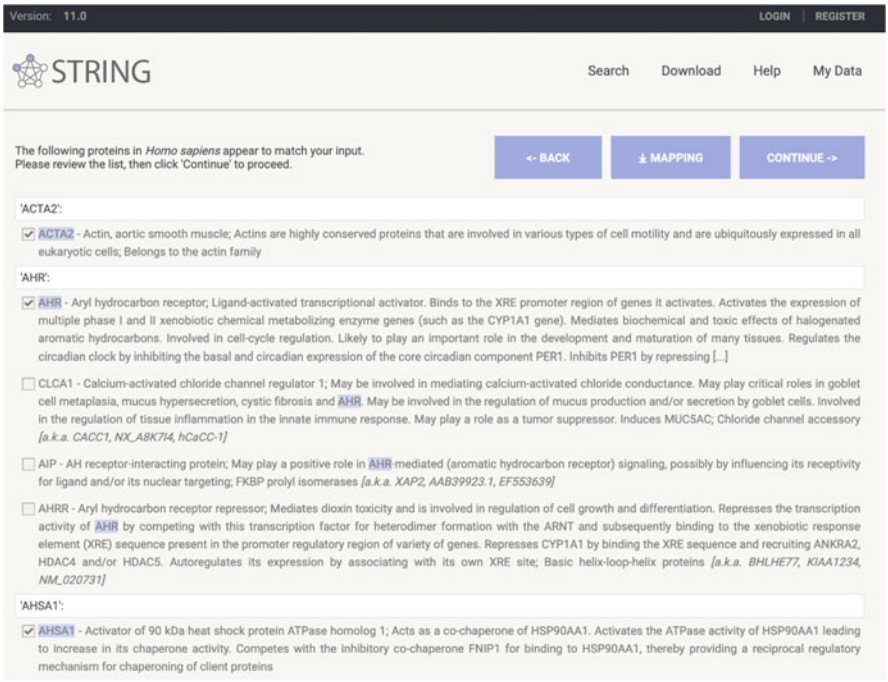


Fig. 4.18 Multiple gene query information confirmation page

analysis process. Select true in the Collapse dataset to gene symbols field. As the number of samples in each group is greater than 7, select phenotype in the Permutation type field. The Chip platform option is for the annotation conversion of the ID, which is not required in this example.

After the above parameters are set, click the Run button under the parameter settings column; the running status is displayed in the GSEA reports at the bottom left of the interface. If it displays Running, it means the operation is successful, and if it displays Error, it means the operation failed, as shown in Fig. 4.26. In case of an error, click Error to view the Error report.

4.3.1.4 View Results

The results of data analysis are saved to the set path. Click index.html to view the web version of the analysis report, as shown in Fig. 4.27.

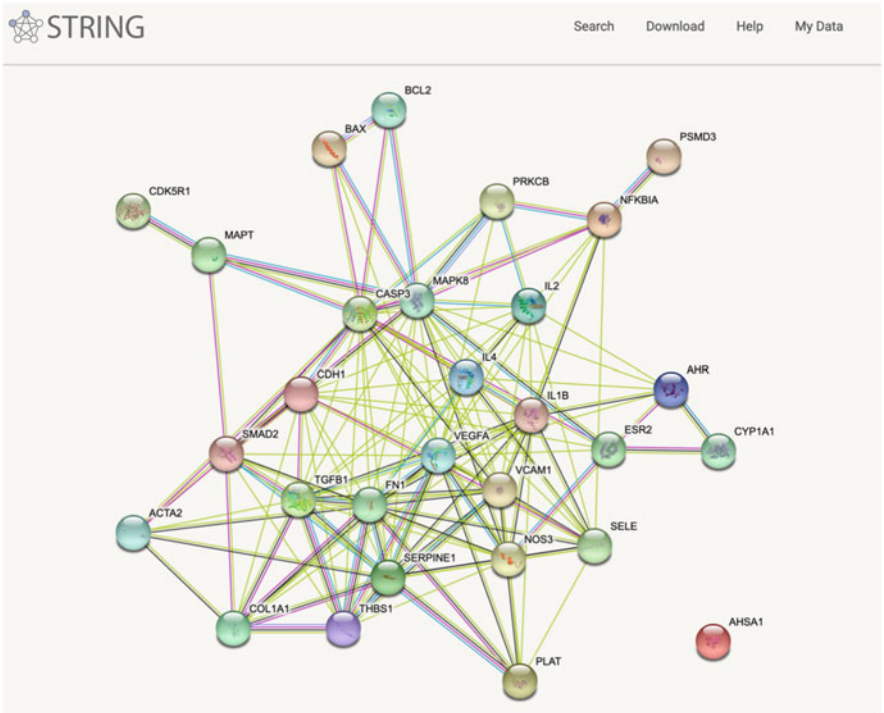


Fig. 4.19 Multiple gene query result page

4.3.2 Network Analysis Software

At present, there are many open source and commercial-use complex network construction and network analysis software. For example, Gephi [43] and Cytoscape [44] are open source and free, as shown in Table 4.1. These powerful software not only provide network graph creation, visualization, and abundant network graph layout methods, but also provide large-scale network analysis algorithms, such as community division algorithm, centrality measurement algorithm, and shortest path calculation method. Cytoscape currently has 14,650 citations and Gephi has 4704. Ruth et al. used Cytoscape to analyze the evolutionary network of mammals and their gut microbes [45], Zhong et al. used Cytoscape to analyze the overall distribution of Saccharomyces cerevisiae protein complex [46], Barberán et al. used Gephi to use network analysis to explore the symbiosis mode of soil microbial communities [47]. In this section, we imitate and introduce two common complex network visualization analysis software: Cytoscape and Gephi. To more intuitively and pragmatically explain the basic functions of software data processing and analysis, we use a small amount of Protein–Protein Interaction (PPI) network dataset (as shown in Table 4.2) to carry out a practical operation and demonstration in

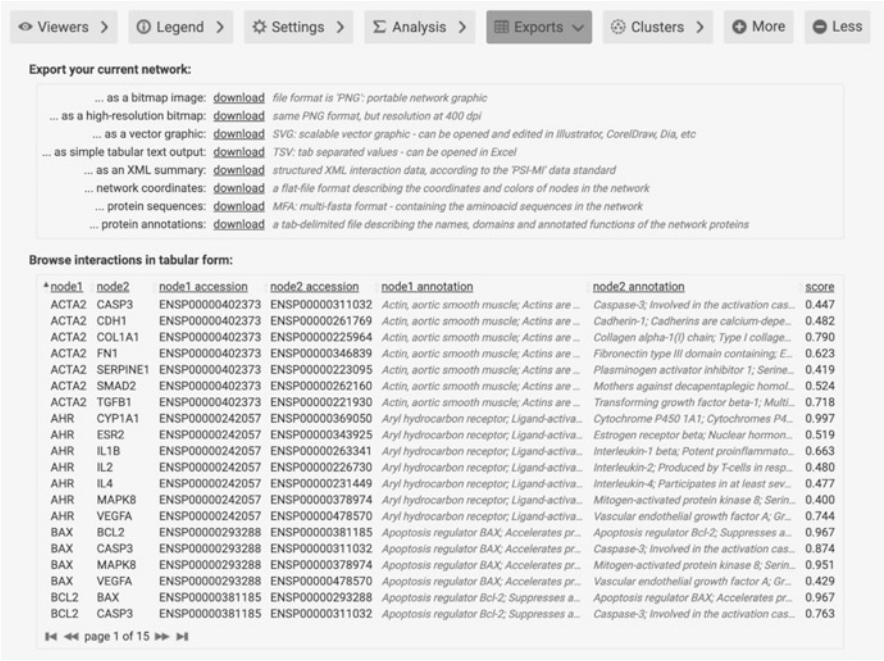


Fig. 4.20 Multiple gene query download page

combination with the functions of the corresponding software and generate relatively intuitive analysis results.

4.3.2.1 Cytoscape

Cytoscape is an open source software platform (latest version is 3.7.1) for visualizing molecular interaction networks and biological pathways, and integrating these networks with annotations, gene expression profiles, and other status data. Although Cytoscape was originally designed for biological research, it has become a universal platform for complex network analysis and visualization. Its dominant function is to analyze the relationships among large-scale protein interactions, protein–DNA, and genetic interactions. Cytoscape's core functions provide the basic components for data integration, analysis, and visualization. Additional extended functions are provided in the form of small programs (apps, formerly called plug-ins). Various apps can be used for molecular network analysis, new layouts, additional file format support, script writing, and connection with databases. The system also supports the development of open APIs based on Java, which can be published to the Cytoscape application store for free download or installation by users. As the software is developed and run on Java, the corresponding Java runtime library needs to be

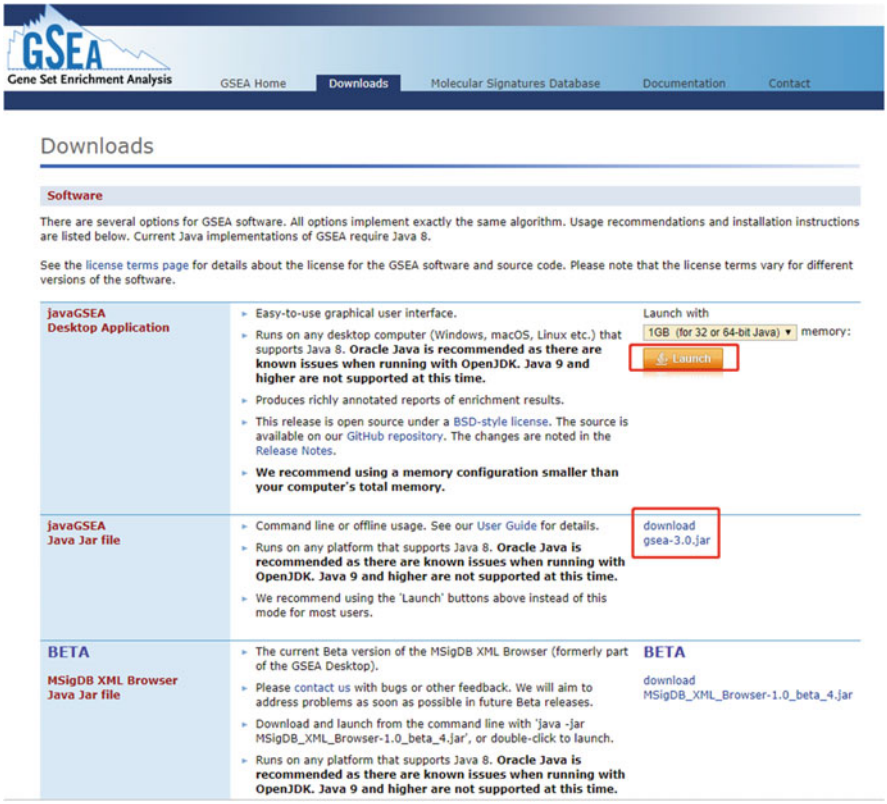


Fig. 4.21 GSEA official download page

installed beforehand. Cytoscape's main interface after an operation is shown in Fig. 4.28.

Basic Use

Launch the software, you can see the upper menu bar, select import under File to import the network; the imported data format for Cytoscape is as shown in Table 4.3. Interaction represents the relationship between nodes; this option can be defined according to actual data. Network layout can be selected as the Layout and includes grid layout, hierarchical layout, and circular layout. In addition to some basic software functions you can also search and install corresponding app plug-ins based on your needs. Cytoscape's core functions are also provided in the form of plug-ins. After importing the network diagram, you can select the layout mode and set the color, size, and shape of the nodes. The operation process is shown in Fig. 4.29.

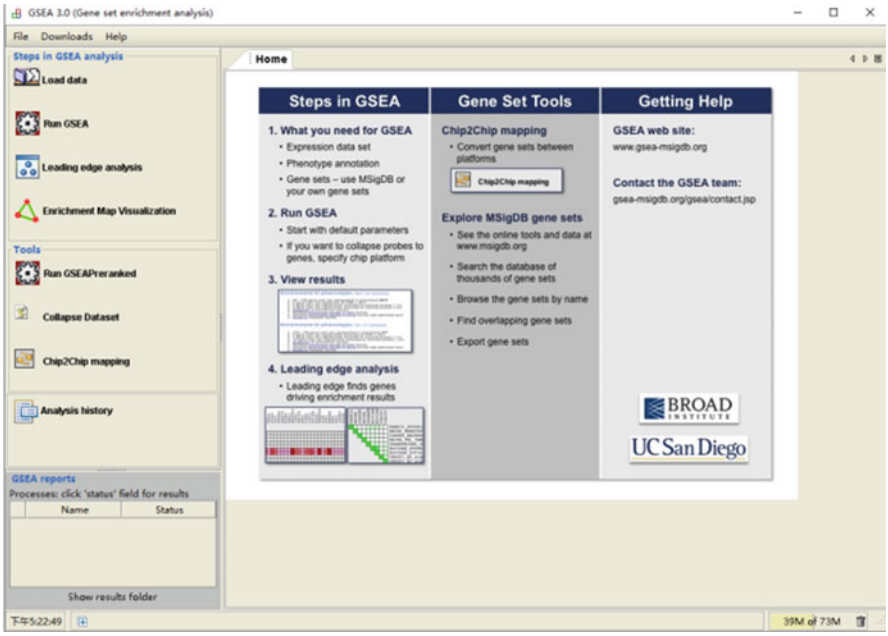


Fig. 4.22 GSEA software startup interface

Exemplary Functional Components

Common requirements in biological network analysis are analysis of network topology characteristics, community analysis, etc. To introduce the functions of Cytoscape more clearly, this chapter demonstrates the practical operation of apps such as CentiScaPe [49] and MCODE [50]. The data used in this section is the PPI network data, as shown in Table 4.3. The following is the actual operation of the above two components combined with PPI data:

Analysis of centrality measurement using CentiScaPe:

CentiScaPe is an app for network centrality measurement calculation, which can be used for analyzing undirected and directed networks. The supported centrality metrics can be divided into three aspects: network, node, and edge and include Network Diameter, Degree, Strength, Betweenness, Closeness, Eccentricity, etc. A simple operation demonstration of CentiScaPe is shown in Fig. 4.30.

Launch the Cytoscape software, select CentiScaPe from the Apps menu bar, the CentiScaPe menu is displayed. In this menu, select the network characteristics that need to be calculated, such as network diameter, node degree, etc. Then select undirected graph or directed graph and press the start button to start the calculation. Each indicator has a corresponding meaning and function. Click the button to the right of the indicator to view the details of that indicator.

Using MCODE for community analysis:

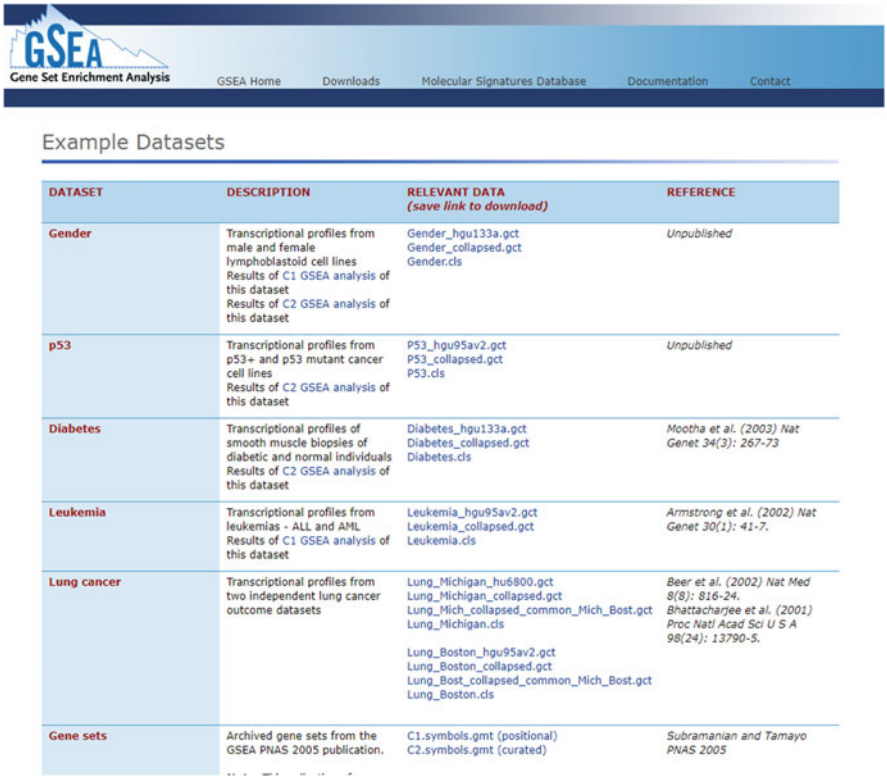


Fig. 4.23 GSEA sample data download page

The MCODE plug-in adopts a Molecular Complex Detection algorithm, which is used to detect the closely connected subnet structure (highly inter-connected local network structures) in the network. This closely connected subnet is also known as community. Communities usually have different practical meanings in different networks. The communities in protein interaction networks are usually part of protein complexes and molecular pathways, whereas the communities in similar networks of protein structures usually represent the protein family. With respect to community extraction, MCODE also supports visualization analysis of the community structure. A simple operation demonstration of MCODE is shown in Fig. 4.31.

Launch the Cytoscape software and import the network. Select the MCODE plug-in from the Apps menu bar. Then select and set the relevant parameters, such as degree coefficient, etc. Click clustering; the corresponding clustering results are displayed in the right panel. Click the appropriate category to display the specific community analysis results in the graph.

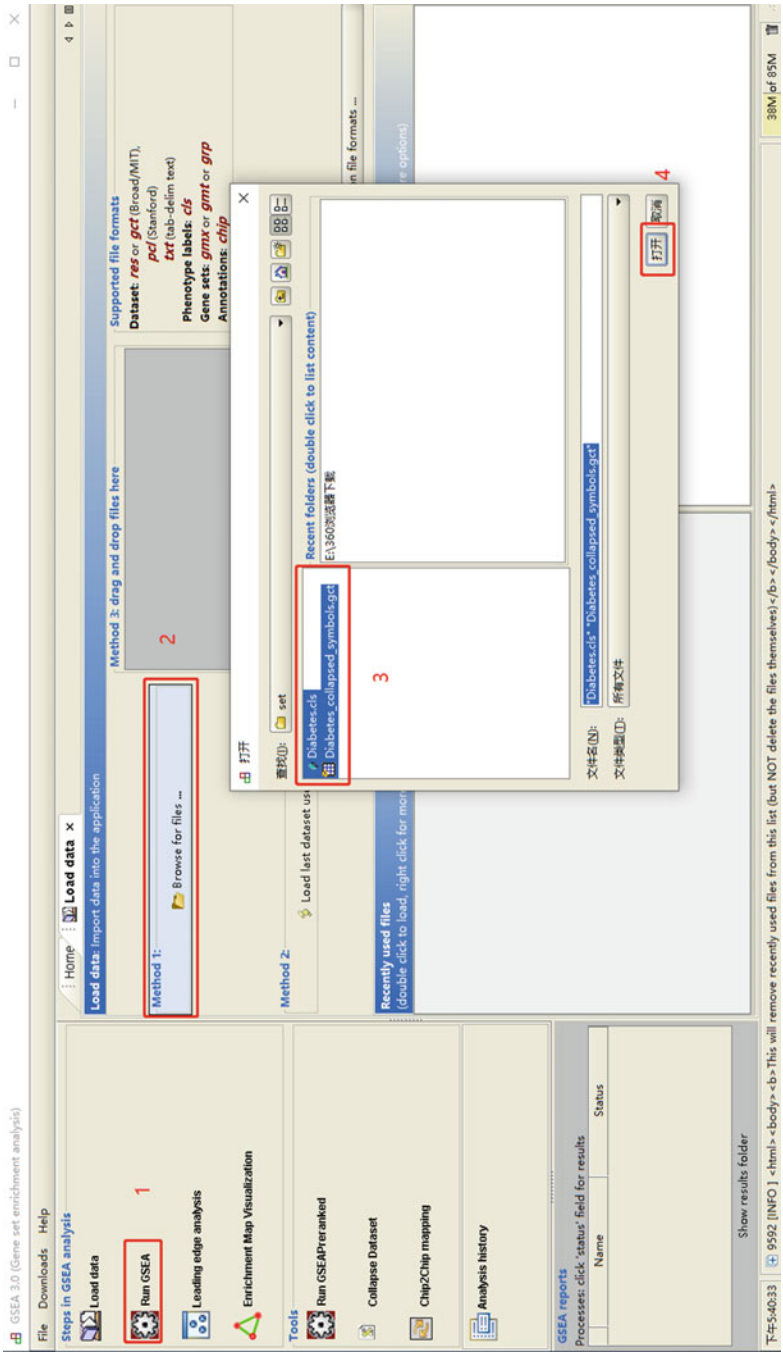


Fig. 4.24 Import data interface

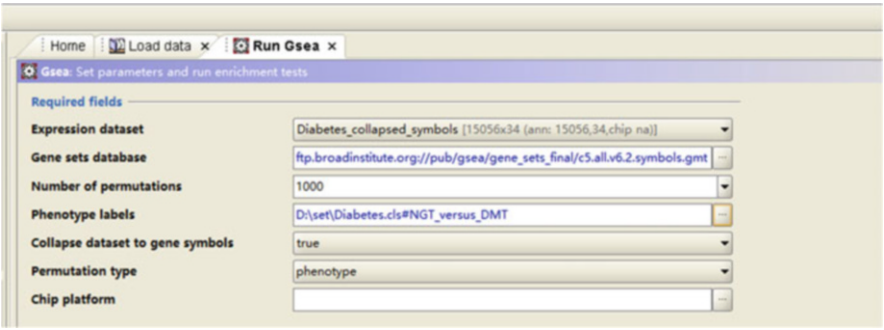


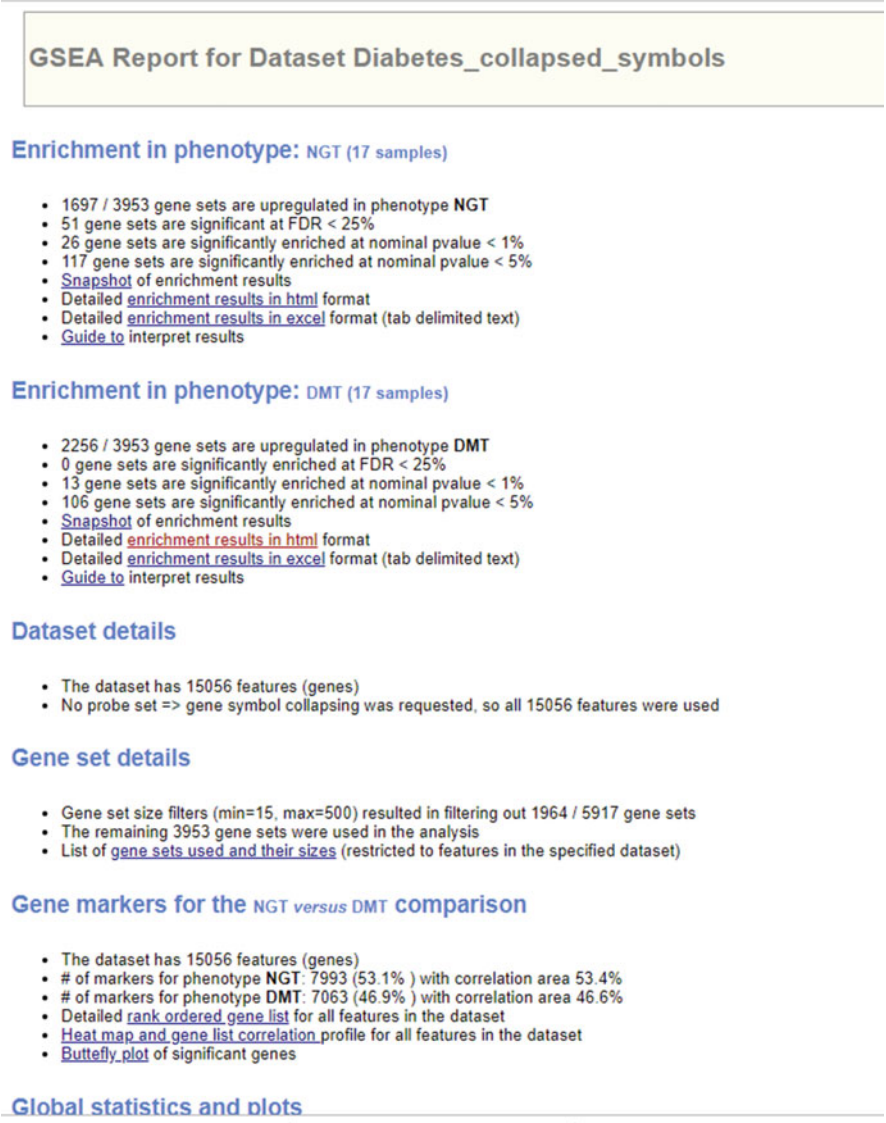
Fig. 4.25 Mandatory parameter setting interface

Fig. 4.26 Running status interface

4.3.2.2 Gephi Visualization Software

Gephi is a free and open source network analysis and visualization software developed on Java. It supports three different operating systems, Mac OS, Windows, and Linux, and supports interfaces in different languages such as English, Simplified Chinese, and French. Gephi was first released in 2006, and the latest version is V0.92. Gephi can visualize any network data represented by nodes and edges, such as social networks, power networks, disease transmission networks, protein interaction networks, etc. At the same time, Gephi supports dozens of algorithms in the form of an extended library, which can be used to calculate the average degree, graph density, and average clustering coefficient of a network and to screen the network according to various criteria, such as edge weight, node degree, etc. Gephi can also be used for community division and visualization of networks. The division algorithms include Fast Unfolding of Communities in Large Networks (BGLL) [51], etc. Table 4.4 shows the data format of the data imported in Gephi. The naming of the node name field should strictly include the Source and Target. Select the appropriate value in the Type field; the available options are undirected graph and directed graph.

Next, we take PPI network data import and analysis in Table 4.4 as an example to demonstrate the corresponding functions of Gephi software. Network data import is the first step in the analysis. Network data can be imported using the main function interfaces by selecting the file menu in the Gephi interface based on different file formats (the data format is shown in Table 4.4 and Fig. 4.32). After the data is



GSEA Report for Dataset Diabetes_collapsed_symbols

Enrichment in phenotype: NGT (17 samples)

- 1697 / 3953 gene sets are upregulated in phenotype **NGT**
- 51 gene sets are significant at FDR < 25%
- 26 gene sets are significantly enriched at nominal pvalue < 1%
- 117 gene sets are significantly enriched at nominal pvalue < 5%
- Snapshot of enrichment results
- Detailed enrichment results in html format
- Detailed enrichment results in excel format (tab delimited text)
- Guide to interpret results

Enrichment in phenotype: DMT (17 samples)

- 2256 / 3953 gene sets are upregulated in phenotype **DMT**
- 0 gene sets are significantly enriched at FDR < 25%
- 13 gene sets are significantly enriched at nominal pvalue < 1%
- 106 gene sets are significantly enriched at nominal pvalue < 5%
- Snapshot of enrichment results
- Detailed enrichment results in html format
- Detailed enrichment results in excel format (tab delimited text)
- Guide to interpret results

Dataset details

- The dataset has 15056 features (genes)
- No probe set => gene symbol collapsing was requested, so all 15056 features were used

Gene set details

- Gene set size filters (min=15, max=500) resulted in filtering out 1964 / 5917 gene sets
- The remaining 3953 gene sets were used in the analysis
- List of gene sets used and their sizes (restricted to features in the specified dataset)

Gene markers for the NGT *versus* DMT comparison

- The dataset has 15056 features (genes)
- # of markers for phenotype **NGT**: 7993 (53.1%) with correlation area 53.4%
- # of markers for phenotype **DMT**: 7063 (46.9%) with correlation area 46.6%
- Detailed rank ordered gene list for all features in the dataset
- Heat map and gene list correlation profile for all features in the dataset
- Buttefly plot of significant genes

Global statistics and plots

Fig. 4.27 Analysis result page

imported, the main interface can be selected through the corresponding properties to flexibly view the network diagram. For example, multiple network layouts and styles can be selected (such as Fruchterman Reingold), and then the corresponding visualization effect can be obtained by running the operation. In addition, the color and size of nodes in the network and the color and size of edges can be adjusted and processed accordingly (as shown in Fig. 4.33). Various topological statistical

Table 4.1 Network analysis software

Three commonly used visualization software			
Directory	Cytoscape	Gephi	Pajek [48]
Developer	UCSD	Mathieu Jacomy, Sebastien Heymann	eytanAdar
Development language	Java	Java	Java
Supported Platforms	Mac OS, Windows, Linux	Mac OS, Windows, Linux	Mac OS, Windows, Linux
Supported languages	English	English, Simplified Chinese, etc.	English
Open source and free	Yes	Yes	Yes

Table 4.2 Examples of PPI network data

Source	Target	Weight
FKBP4	HSP90AA1	1
CFTR	HSPA8	1
CFTR	SLC9A3R1	1
CFTR	UBC	1
CYP51A1	LSS	1
USH1C	CDH23	1
RALA	RALBP1	1
RALA	EXOC2	1
RALA	EXOC8	1
CX3CL1	CX3CR1	1

features of the network, such as average degree, network diameter, and betweenness, can be conveniently calculated and displayed (as shown in Fig. 4.34). Community analysis is an important algorithm for complex network analysis and is also one of the basic functions. Gephi integrates the classic community analysis method into a toolbar called "statistics." After clicking and running, the results of the community structure analysis in the network are displayed. The visualization of specific community structure can be classified and displayed (as shown in Fig. 4.35) through the color rendering mode (the selection is based on modules) of nodes in the menu on the left side of the main interface. It is worth noting that several other Gephi analysis functions are integrated in the form of plug-ins; users can load the corresponding plug-ins through the menu to obtain new analysis functions.

4.3.2.3 Pajek Complex Network Visualization Software

Among the more complex network analysis software, Pajek is a free large-scale complex network analysis tool with a more than two decade-long research and development history (since 1996). Compared with other software, most network

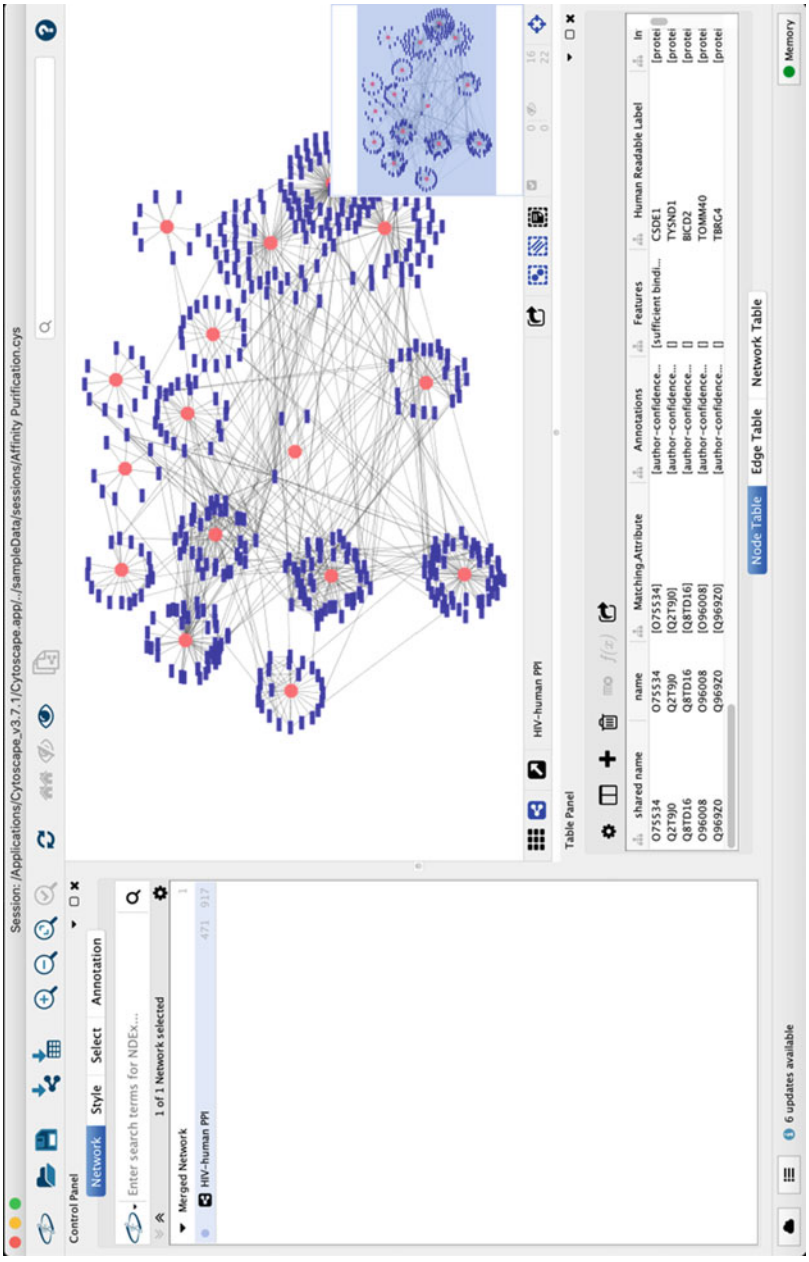


Fig. 4.28 Cytoscape software homepage

Table 4.3 Import data format of cytoscape

source	interaction	target
FKBP4	Pp	HSP90AA1
CFTR	pp	HSPA8
CFTR	pp	SLC9A3R1
CFTR	pp	UBC
CYP51A1	pp	LSS
USH1C	pp	CDH23
RALA	pp	RALBP1
RALA	pp	EXOC2
RALA	pp	EXOC8
CX3CL1	pp	CX3CR1
TRAPPC6A	pp	TRAPPC3
NDUFAB1	pp	NDUFB7

analysis algorithms implemented in Pajek have a low computational time complexity. Therefore, an ultra-large-scale network that can handle hundreds of millions of nodes is a powerful analysis tool for developing various large-scale complex nonlinear networks. The latest version of Pajek is V5.08 (supports 32-bit and 64-bit operating systems), with Windows, Linux, and Mac versions. Pajek is updated on a regular basis. It uses network exploratory analysis methods such as centrality measurement and community analysis; however, it has poor visualization effect. In addition, through the recent development (2019) of the R language interface package, the statistical analysis function of the R language can be used to create powerful network structure statistical analysis capability.

4.4 Toolkit Based on Programming Languages

Current visualization toolkits based on programming language calls basically use the network topology statistical measurement, classic graph algorithms, community division, and link prediction analysis methods. They are more flexible in terms of network operations, can accurately control nodes and edges, and can easily adjust the corresponding calculation functions as required. However, in general, programming language-call-based tools are suitable for backend batch computing and system integration. Based on the programming language, we have selected a representative common network visualization package for C++, Java, Python, and R programming language. Some common visualization toolkits are listed in Table 4.5.

C++ and Java toolkits are briefly introduced below:

1. Boost Graph Library: A C++ Network Visualization Toolkit

The Boost Graph Library (BGL) is a C++ visualization toolkit that provides generic interfaces that can access the internal structure of a graph while hiding implementation details. It has an open interface and the graph library that

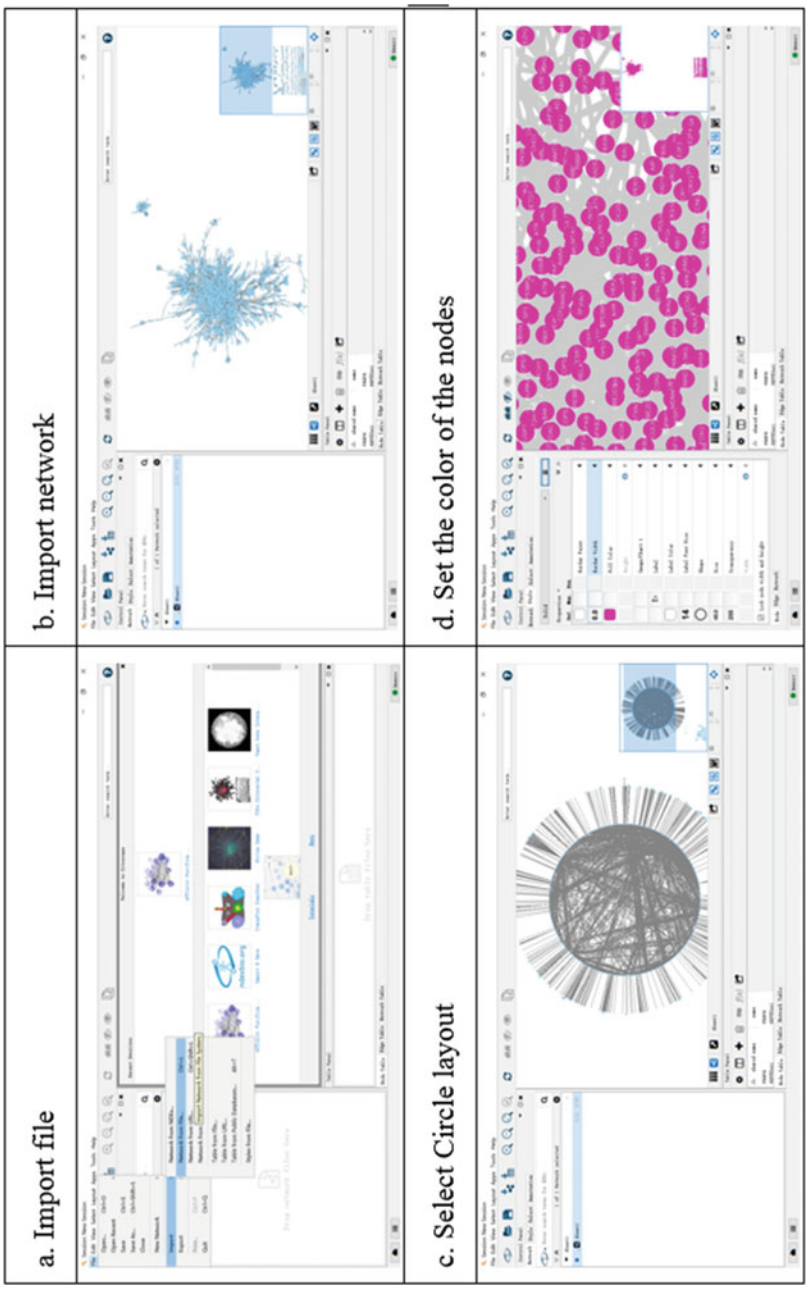


Fig. 4.29 Cytoscape visualization network

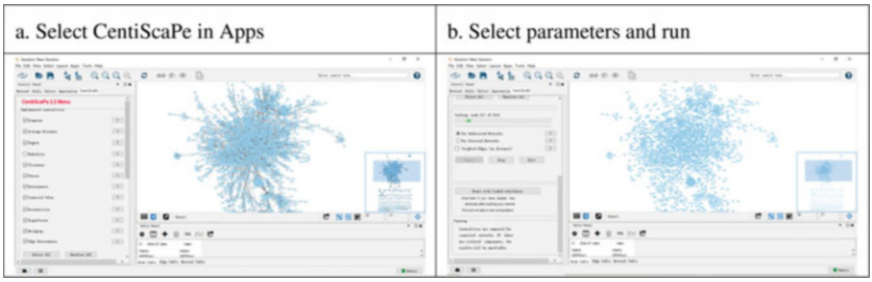


Fig. 4.30 Use of Centiscape plug-in

implements this interface can interoperate with the BGL algorithm and other algorithms that implement this structure. It supports three kinds of data format—adjacency list, adjacency matrix, and edge list. BGL can be used for visualization and provides many graph-related algorithms, such as Dijkstra algorithm for the shortest path, Kruskal algorithm for the minimum spanning tree, topological sorting, etc.

2. GraphStream: A Java Network Visualization Toolkit

GraphStream is a graph library for processing Java which focuses on the dynamic representation of graphs. The main research object of this library is the modeling of dynamic interactive networks of various scales. The goal of this library is to provide a method to represent graphs and process them. To this end, GraphStream provides several graph classes that allow directed and undirected, 1-graph, or P-graph (that is, multiple graphs, graphs that can have multiple edges between two nodes) modeling. GraphStream allows any type of data attributes to be stored on graph elements: numbers, strings, or any objects. In addition, the graphic flow also provides a method to process graph evolution over time, which can be used to display the way in which nodes and edges are added and removed, and the possible way data attributes can appear, disappear, and evolve.

The following is a brief introduction and operation demonstration of Python and R toolkit combined with some cases.

4.4.1 NetworkX

The first version of NetworkX was released in May 2002, and the current number of citations has reached 2149. It is a graph theory and complex network modeling tool developed in Python language. It has built-in commonly used graphs and complex network analysis algorithms that can easily carry out complex network data analysis, simulation modeling, and other work. NetworkX makes it easy to generate both classical and random graphs, such as scale-free networks (where a few nodes have several), which is convenient for carrying out some network analysis without data. NetworkX supports the creation of simple undirected graphs, directed graphs, and

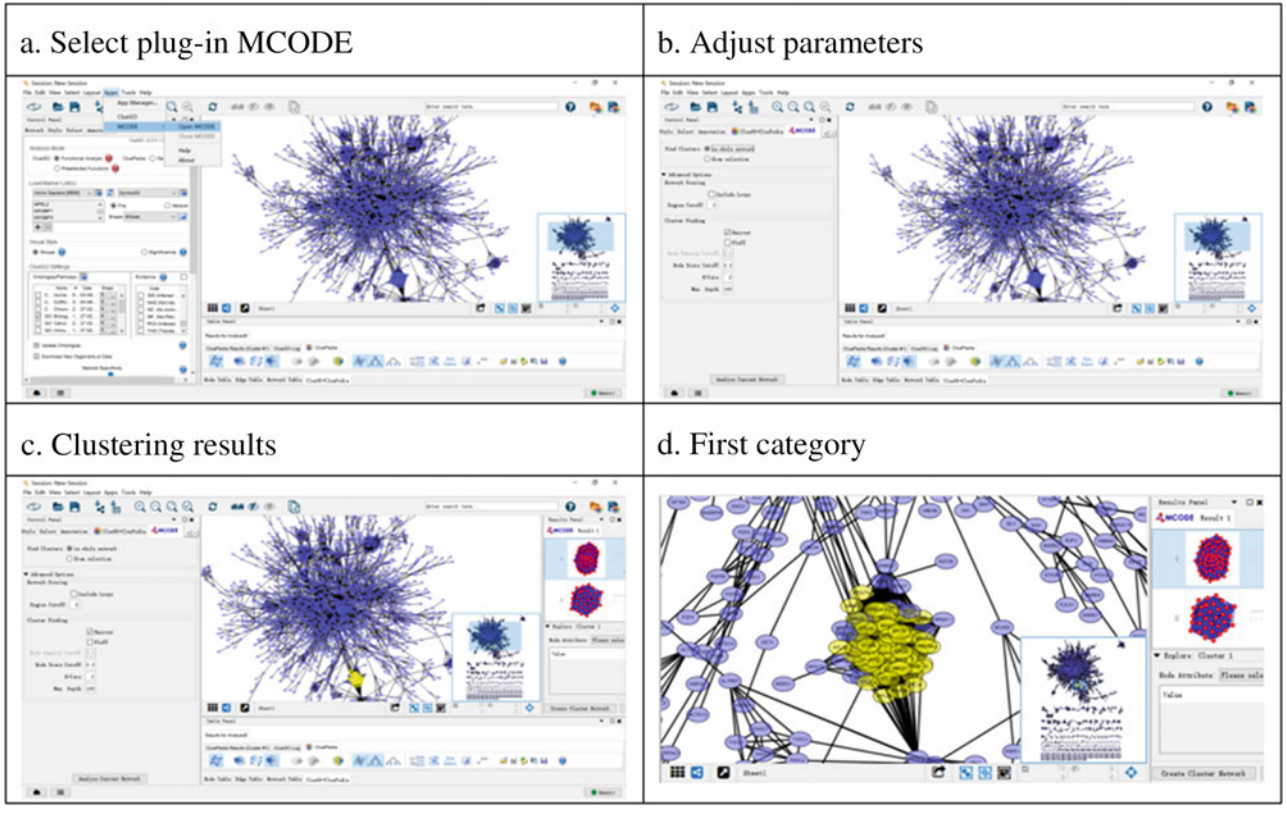


Fig. 4.31 MCODE plug-in use

Table 4.4 Gehpi data format

Source	Target	Weight
FKBP4	HSP90AA1	1
CFTR	HSPA8	1
CFTR	SLC9A3R1	1
CFTR	UBC	1
CYP51A1	LSS	1
USH1C	CDH23	1
RALA	RALBP1	1
RALA	EXOC2	1
RALA	EXOC8	1
CX3CL1	CX3CR1	1

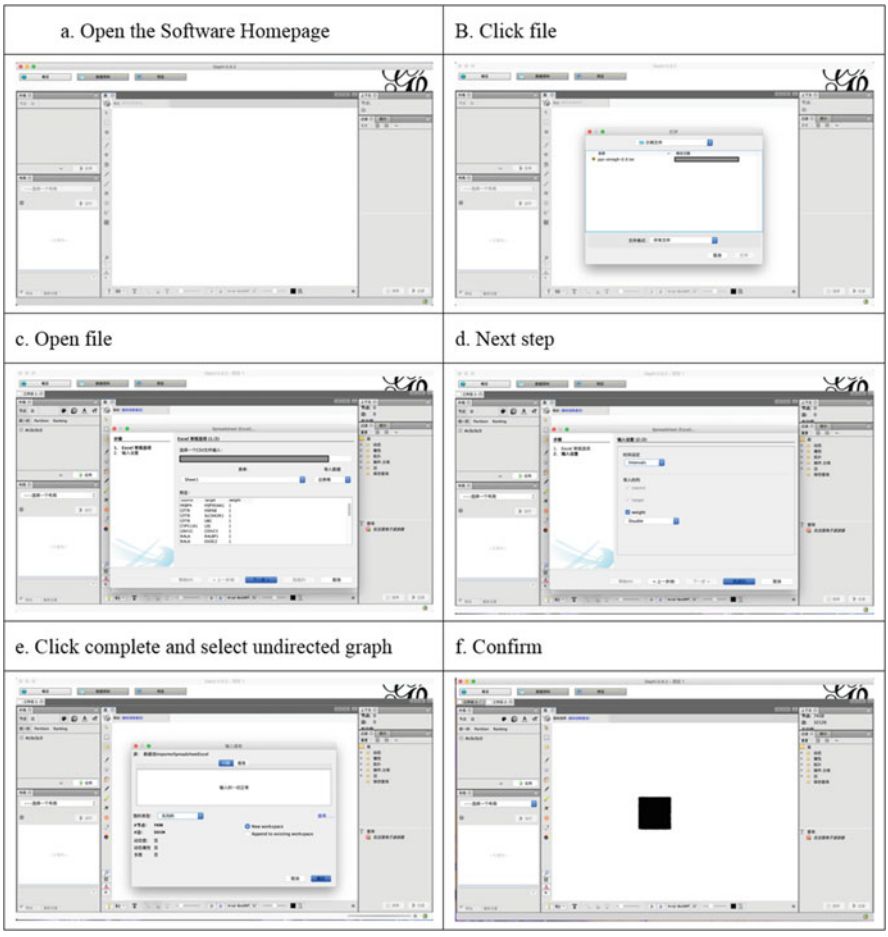


Fig. 4.32 Gephi import network

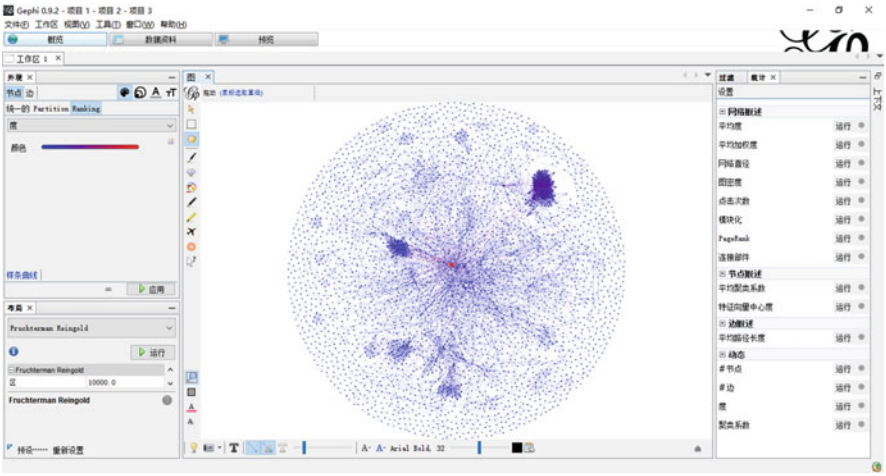


Fig. 4.33 Network layout and node edge settings

multiple graphs. It has many standard graph theory algorithms built-in, and the nodes can be any data. It has rich functions and is easy to use. For programmers who are familiar with Python, the NetworkX visualization Python package is a very convenient tool with simple and efficient operation. Table 4.6 shows some basic functions of NetworkX in the Python environment. For detailed functions, please check the documents on the official website.

To more vividly demonstrate the visualization effect of NetworkX, this chapter uses an exemplary small amount of clinical disease merger relational data and uses Python code to generate a visualized network diagram. The disease merger relational network data contains 51 disease nodes and 150 disease merger relational edges (as shown in Table 4.7). The core codes are shown in Table 4.8. The visualization result is shown in Fig. 4.36. The case diagram uses a circular layout. The larger the node degree in the network, the larger the node. The color is also set according to the node degree. The larger the node degree, the closer the node color is to blue. The greater the weight of the edge between the two nodes, the wider the line of the edge; the width of the edge between hypertension and renal insufficiency in the merger disease network is the widest, as hypertension and renal insufficiency are more likely to occur as co-morbidity.

4.4.2 igraph

igraph is a simple and easy-to-use network analysis tool. Several of its functions are developed using C language. It has high computing efficiency and is highly suitable for solving large and complex network problems. R, Python, and C/C++ can be used to call the corresponding packages for visualization. The latest version is V1.0.0. In

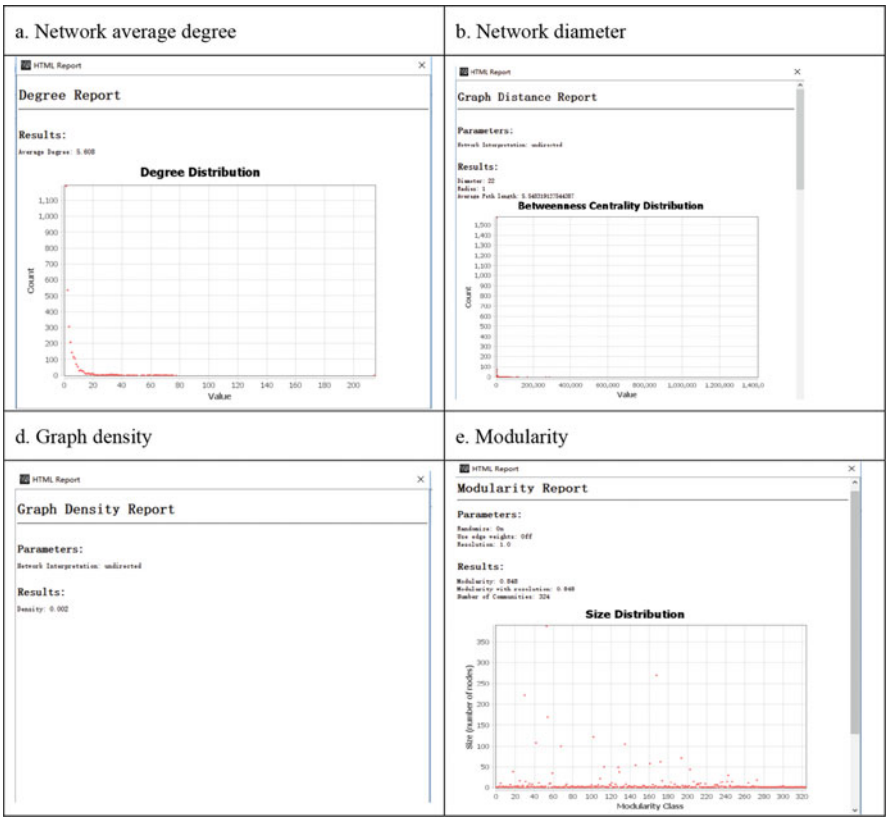


Fig. 4.34 Statistical index of the Gephi diagram

the igraph network graph we can set the node color and calculate node degree, edge density, clustering coefficient, and other statistics and their distribution and can also cluster the network and visualize each category. In this section, we first call the igraph R language package with the R language to create a visualization example. Table 4.9 lists the basic functions of the igraph R package.

To show the visualization effect of igraph, we perform a simple operation demonstration using the disease network dataset. First, install RStudio, and then install the igraph R package. Table 4.10 shows the core codes of the case (based on R language), and Fig. 4.37 shows the visualization effect of the case. Different colors in the figure represent different communities, and each community represents a set of closely related diseases.

Taking into account the wide range of network pharmacology applications, this chapter only provides a high-level introduction to the main network pharmacology methods and software. It focuses on introducing common software and methods such as common complex network analysis and visualization, molecular and network function analysis, drug-target prediction, and drug indication prediction that

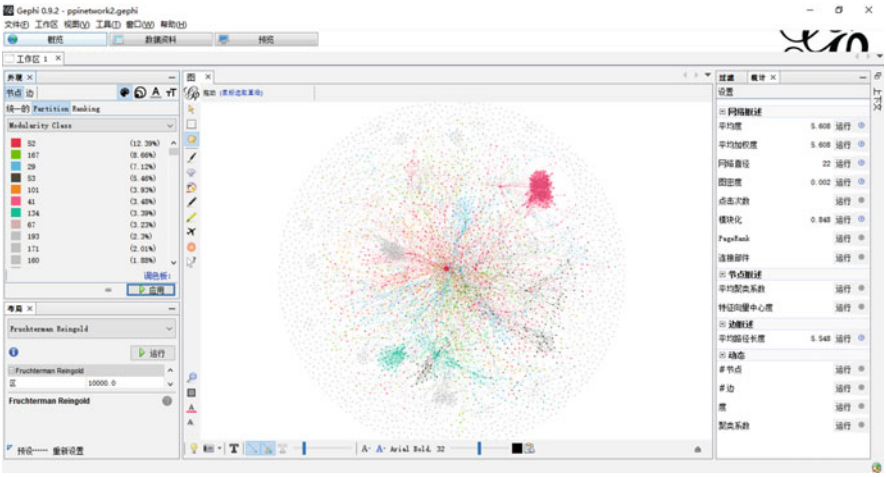


Fig. 4.35 Gephi module rendering

Table 4.5 Common web visualization toolkits

Name	NetworkX [52]	igraph [53]	Boost Graph Library [54]	GraphStream [55]
Developers	Aric Hagberg, Pieter J. Swart	Szabolcs Horvát	Douglas Gregor Andrew Lumsdaine	Julien Baudry Antoine Dutot
Programming languages	Python	R, Python, C/C++	C++	Java
Open source and free	Yes	Yes	Yes	Yes
Latest versions	2.3	1.0.0	1.70.0	1.3

Table 4.6 NetworkX functions

Functions	Description
add_node(node)	Add node to graph
get_node_attributes(G, name)	Get node attributes from graph
all_neighbors(G, node)	Return to neighboring node of the node
common_neighbors(G, u, v)	Return to common neighbor of the two nodes
is_directed(G)	Judge whether graph is a directed graph
nodes(G)	Return to nodes of the graph
number_of_nodes(G)	Find the number of graph nodes of the graph
add_edge(node1, node2)	Add edge
get_edge_attributes(G, name)	Get edge attributes
number_of_edges(G)	Return to number of edges
clear()	Delete nodes and edges in the graph

Table 4.7 Dataset of diabetes with combination of diseases

Disease 1 (node 1)	Disease 2(node 2)	Weight (Weight)
Hypertension	Renal insufficiency	1
Lipid metabolism disorder	Hypertension	0.995283019
Hypertension	Heart disease	0.872641509
Cerebral infarction	Hypertension	0.613207547
Hypertension	Hepatic disease	0.514150943
Lipid metabolism disorder	Hepatic disease	0.504716981
Lipid metabolism disorder	Renal insufficiency	0.367924528
Gastrointestinal lesions	Hypertension	0.29245283
Lipid metabolism disorder	Cerebral infarction	0.245283019
Retinopathy	Renal insufficiency	0.240566038
Hypertension	Cervical spondylosis	0.240566038
Heart disease	Renal insufficiency	0.231132075
Neuropathy	Hypertension	0.202830189
Cerebral infarction	Heart disease	0.202830189

Table 4.8 NetworkX drawing core codes

Core Codes of NetworkX Drawing Network Diagram

import networkx as nx # Import networkx package
import matplotlib.pyplot as plt # Import drawing kit
G = nx.Graph() # Create a graph

G.add_nodes_from(nodes) #Add nodes to graph, nodes are edge list of the graph
G.add_weighted_edges_from(edges)# Add edges to graph, edges are edge list of the graph
de = G.degree() # Calculate degree of graph
node_color = [convert_to_hex(de[i]) for i in nodes]#Set node color according to degree
node_size = [de[i]*30 for i in nodes]#Set node size according to degree
widths = [int(line[2])/50 for line in data]#Set width of edge according to weight of edge
nx.draw(G,pos=nx.spring_layout(nodes), with_labels = True,font_size=10, node_size=node_size,width=widths,node_color=node_color)
plt.rcParams['font.sans-serif'] = ['SimHei'] #Display Chinese tag
plt.savefig("ba.png") #Save figure
plt.show() #Display

are unique to network pharmacology. Traditional computational pharmacology software such as virtual screening (Docking) software is not involved [56]. At the same time, this chapter provides some further details from the perspective of network pharmacology technology and application scope including the construction methods of network pharmacology-related resources, such as the information extraction method of drug-target and drug side effect relationship [57], the transformation network pharmacology method combining clinical and basic medicine [58], the network pharmacology prediction method based on deep learning, etc., which have become new and key research topics [59]. This chapter does not elaborate much on the above aspects. Particularly, there are several important studies in the

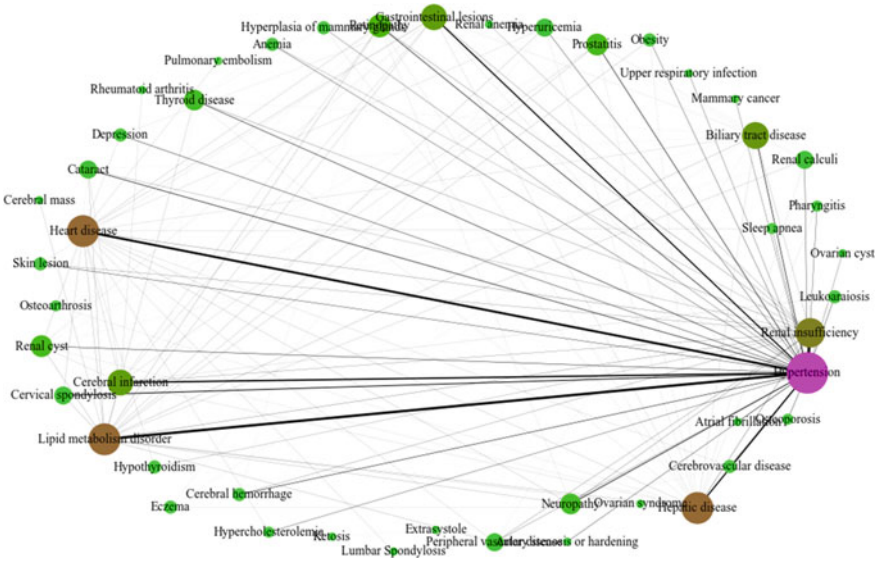


Fig. 4.36 NetworkX graph visualization display

Table 4.9 igraph basic functions

Functions	Description
nodes<-read.table("nodes.txt")	Read information of the node
links<-read.table("links.txt")	Read information of the edge
net<-graph_from_data_frame(d = links, vertices = nodes, directed = F)	Create network
E(net)$color<- "red"	Designate node color
plot(net,vertex.color="gray50",edge.color = E(net)$color)	Graph visualization
plot(net, layout = layout_in_circle)	Layout
degree(g, mode="all")	Compute all node degrees in the graph
edge_density(net)	Calculate edge density
transitivity(g, type="global")	Calculate clustering coefficient
cfg <- cluster_fast_greedy(net)	Clustering, mining community in the network

prediction and analysis of adverse drug reactions and drug side effects, as well as the prediction method of drug interaction relationship. The related research plays a vital role in network pharmacology research. However, this chapter does not cover the methods and software information from this aspect. Readers who are interested can refer to other research works [60, 61].

Conversely, current network pharmacology software and analysis processing algorithms are focused on the functions of independent technical links, such as

Table 4.10 igraph visualization core codes

Core codes: igraph visualization
library(igraph)#load igraph package
net <- graph_from_data_frame(relation, directed=FALSE)#Create graph
cfg <- cluster_fast_greedy(net)#Community division algorithm
plot(cfg, net)#visualization
Degree(net)#Add up node degree
edge_density(net)#Calculate edge density
transitivity(net, type="global")#Calculate clustering coefficient

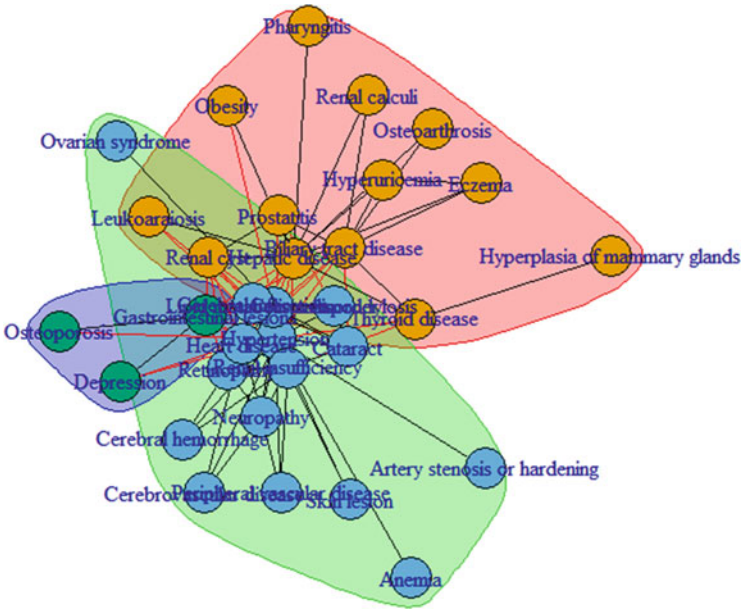


Fig. 4.37 igraph disease network visualization

network analysis and visualization, drug-target relationship prediction, etc. However, as network pharmacology research involves many upstream and downstream technologies and functional links, researchers need to combine and apply different software and algorithms to generate corresponding research results. To this end, in order to improve the effectiveness of network pharmacology research, there is an urgent need to develop an integrated, high-performance, and service-oriented network pharmacology software platform. The platform needs to include network data integration, network analysis and prediction, visualization processing, functional enrichment analysis, and related literature validation, to support the integrated network pharmacology research process.

References

1. Hopkins AL. Network pharmacology: the next paradigm in drug discovery. Nat Chem Biol. 2008;4(11):682–90.
2. Newman MEJ. Networks: An Introduction. Astronomische Nachrichten. 2010;327(8):741–3.
3. Barabasi AL. Network science: Luck or reason. Nature. 2012;489(7417):507–8.
4. Barabasi AL, Gullbahce N, Loscalzo J. Network medicine: a network-based approach to human disease. Nat Rev Genet. 2011;12(1):56–68.
5. Lu LY. Link Prediction on Complex Networks. J Univ Electron Sci Technol China. 2010;39 (5):651–61.
6. Watts DJ. Identity and Search in Social Networks. Science. 2002;296(5571):1302–5.
7. Barabasi AL, Albert R. Emergence of scaling in random networks. Science. 1999;286 (5439):509–12.
8. Watts DJ, Strogatz SH. Collective Dynamics of Small World Networks. Nature. 1998;393 (6684):440–2.
9. Barabasi AL. Scale-Free Networks: A Decade and Beyond. Science. 2009;325(5939):412–3.
10. Menche J, Harma A, Kitsak M. Uncovering disease-disease relationships through the incomplete interactome. Science. 2015;347(6224):1257601.
11. Zhou ZH. Machine learning [M]. Beijing: Tsinghua University Press; 2016.
12. Duan H, Zhai CX. Mining coordinated intent representation for entity search and recommendation[C]. New York: ACM; 2015.
13. Deng J, Yuan Q, Mamitsuka H, et al. DrugE-rank: predicting drug-target interactions by learning to rank. Data Mining Syst Biol. 2018;1807:195–202.
14. Juurlink DN, Mamdani M, Kopp A, et al. Drug-Drug Interactions Among Elderly Patients Hospitalized for Drug Toxicity. JAMA. 2003;289(13):1652.
15. Cheng FX, Kovacs IA, Barabasi AL. Network-based prediction of drug combinations. Nat Commun. 2019;10(1):1197.
16. Govindaraj RG, Naderi M, Singha M, et al. Large-scale computational drug repositioning to find treatments for rare diseases. NPJ Syst Biol Appl. 2018;4(1):13.
17. Law V, Knox C, Djoumbou Y, et al. DrugBank 4.0: shedding new light on drug metabolism. Nucleic Acids Res. 2014;42(D1):D1091–7.
18. Kuhn M, Von MC, Campillos M, et al. STITCH: interaction networks of chemicals and proteins. Nucleic Acids Res. 2007;36(Database issue):D684–8.
19. Kuhn M, Letunic I, Jensen LJ, et al. The SIDER database of drugs and side effects. Nucleic Acids Res. 2016;44(D1):D1075–9.
20. Yamanishi Y, Kotera M, Moriya Y, et al. Dinies: drug–target interaction network inference engine based on supervised analysis. Nucleic Acids Res. 2014;42(W1):W39–45.
21. Nickel J, Gohlke BO, Erehman J, et al. SuperPred: update on drug classification and target prediction. Nucleic Acids Res. 2014;42(W1):W26–31.
22. Gfeller D, Grosdidier A, Wirth M, et al. SwissTargetPrediction: a web server for target prediction of bioactive small molecules. Nucleic Acids Res. 2014;42(W1):W32–8.
23. Vanhaelen Q. Computational methods for drug repurposing[M]. New York: Springer Science Business Media; 2018.
24. Xie DF, Li P, Li F, et al. Computational Approaches to Analyze the Strategies of Drug Repositioning. Prog Biochem Biophys. 2012;39(11):1029–36.
25. Brown AS, Patel CJ. meshdd: literature-based drug-drug similarity for drug repositioning. J Am Med Inform Assoc. 2016;24(3):614–8.
26. Moosavinasab S, Patterson J, Strouse R, et al. RE: fine drugs: an interactive dashboard to access drug repurposing opportunities. Database. 2016; pii: baw083
27. Wang X, Yin TS, Li BY, et al. Progress in Gene Functional Enrichment Analysis. Scientia Sinica(Vitae). 2016;46(4):363.
28. Huang DW, Sherman BT, Lempicki RA. Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. Nucleic Acids Res. 2009;37(1):1–13.

29. Wu J, Mao X, Cai T, et al. KOBAS server: a web-based platform for automated annotation and pathway identification. Nucleic Acids Res. 2006;34:W720–4.
30. Szklarczyk D, Gable AL, Lyon D, et al. STRING v11: protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. Nucleic Acids Res. 2018;47(D1):D607–13.
31. Wang J. Protein-protein Interaction Databases. Chin J Biochem Mol Biol. 2017;2017(33):767.
32. Gao JR. Molecular Biology [M]. Chemical Industry Press; 2011.
33. Licata L, Briganti L, Peluso D, et al. MINT, the molecular interaction database: 2012 update. Nucleic Acids Res. 2012;40(Database issue):D857–61.
34. Stark C, Breitkreutz BJ, Reguly T, et al. BioGRID: a general repository for interaction datasets. Nucleic Acids Res. 2006;34(Database issue):D535–9.
35. Stelzl U, Worm U, Lalowski M, et al. A human protein-protein interaction network: a resource for annotating the proteome. Cell. 2005;122(6):957–68.
36. Goh KI, Cusick ME, Valle D, et al. The human disease network. PNAS. 2007;104 (21):8685–90.
37. Scott J. Trend report social network analysis. Sociology. 1988;22(1):109–27.
38. Dholakia UM, Baggozzi RP, Pearo LK. A social influence model of consumer participation in network- and small-group-based virtual communities. Int J Res Mark. 2004;21(3):241–63.
39. Cha M, Mislove A, Gummadi KP. A measurement-driven analysis of information propagation in the flickr social network[C]//. In: Proceedings of the 18th International Conference on World Wide Web; 2009. p. 721–30.
40. Johnell K, Klarin I. The Relationship between Number of Drugs and Potential Drug-Drug Interactions in the Elderly. Drug Saf. 2007;30(10):911–8.
41. Muhammed AY, Kwang G, Michael EC, et al. Drug-target network. Nat Biotechnol. 2007;25 (10):1119–26.
42. Subramanian A, Kuehn H, Gould J, et al. GSEA-P: A desktop application for Gene Set Enrichment Analysis. Bioinformatics. 2007;23(23):3251–3.
43. Bastian M, Heymann S, Jacomy M. Gephi: an open source software for exploring and manipulating networks[C]. In: International AAAI Conference on Weblogs and Social Media; 2009.
44. Shannon P, Markiel A, Ozier O, et al. Cytoscape: A Software Environment for Integrated Models of Biomolecular Interaction Networks. Genome Res. 2003;13(11):2498–504.
45. Ley RE, Hamady M, Lozupone C, et al. Evolution of mammals and their gut microbes. Science. 2008;320(5883):1647–51.
46. Krogan NJ, Cagney G, Yu H, et al. Global landscape of protein complexes in the yeast Saccharomyces cerevisiae. Nature. 2006;440(7084):637–43.
47. Barberan A, Bates ST, Casamayor EO, et al. Using network analysis to explore co-occurrence patterns in soil microbial communities. ISME J. 2012;6(2):343–51.
48. Batagelj V, Mrvar A. Pajek - Analysis and Visualization of Large Networks. In: Lecture Notes in Computer Science. Berlin: Springer; 2004. p. 77–103.
49. Scardoni G, Tosadori G, Faizan M, et al. Biological network analysis with CentiScaPe: centralities and experimental dataset integration. F1000research. 2014;3:139.
50. Bader GD, Hogue CW. An automated method for finding molecular complexes in large protein interaction networks. BMC Bioinform. 2003;4(1):1–27.
51. Blondel VD, Guillaume JL, Lambiotte R, et al. Fast unfolding of community hierarchies in large networks. J Statistical Mech Theory Exper. 2008;2008(10):P10008.
52. Hagberg A, Schult D, Swart P, et al. Exploring Network Structure, Dynamics, and Function using NetworkX[C]//. In: Proceedings of the 7th Python in Science Conference (SciPy); 2008. p. 11–5.
53. Csardi G, Nepusz T. The igraph software package for complex network research. Inter J Complex Syst. 2006;2006:1695.
54. Gregor D, Lumsdaine A. The Parallel BGL: A Generic Library for Distributed Graph Computations. Parallel Object-Oriented Scientific Computing (POOSC). 2005;2:1–18.

55. Antoine D, Frederic G, Damien O, et al. GraphStream: A Tool for bridging the gap between Complex Systems and Dynamic Graphs. 2008.
56. Pagadala NS, Syed K, Jack T. Software for molecular docking: a review. Biophys Rev. 2017;9 (2):91–102.
57. Xue H, Li J, Xie H, et al. Review of Drug Repositioning Approaches and Resources. Int J Biol Sci. 2018;14(10):1232–44.
58. Messinis DE, Melas IN, Hur J, et al. Translational systems pharmacology-based predictive assessment of drug-induced cardiomyopathy. Cpt Pharmacometrics Syst Pharmacol. 2018;7 (3):166–74.
59. Geoffrey H. Deep learning-a technology with the potential to transform health care. JAMA. 2018;320(11):1101–2.
60. Lounkine E, Keiser MJ, Whitebread S, et al. Large-scale prediction and testing of drug activity on side-effect targets. Nature. 2012;486(7403):361.
61. Zitnik M, Agrawal M, Leskovec J. Modeling polypharmacy side effects with graph convolutional networks. Bioinformatics. 2018;34(13):i457–66.