An Overview of Systems Biology & Metabolomics

Kirubanandan Shanmugam*

Independent Research Professional, Chennai, India

***Corresponding author:** Kirubanandan Shanmugam, ksh1005@yahoo.com

Abstract: Systems biology is an interdisciplinary strategy that seeks to comprehend intricate biological systems by amalgamating diverse forms of data, such as genomics, proteomics, and metabolomics. Metabolomics, which investigates small molecules generated during cellular processes, offers insights into the biochemical activities taking place within an organism. Through the examination of the metabolome, researchers can acquire knowledge about the biochemical processes happening within an organism. Metabolomics proves to be particularly advantageous in systems biology as it furnishes a momentary view of the metabolic condition of a biological system at a specific point in time. By merging metabolomics data with other omics data, researchers can construct comprehensive models to enhance their understanding of how biological systems operate as a unified entity. This comprehensive methodology facilitates a more thorough comprehension of intricate biological systems and can lead to the discovery of new biomarkers and therapeutic targets. Systems biology and metabolomics are closely linked disciplines within the broader realm of biology, each offering distinct viewpoints and methodologies for interpreting biological systems at varying levels of complexity. Through the integration of these methodologies, researchers can acquire a more holistic understanding of biological processes, pathways, and networks at a functional level. This integration enables the identification of crucial metabolites, pathways, and regulatory mechanisms that hold significant roles in biological systems. Chemical and compound databases are pivotal in systems biology and metabolomics research as they grant access to information regarding the structure, characteristics, and interactions of diverse molecules. These databases serve as valuable assets for researchers to recognize and describe compounds, explore metabolic pathways, and devise experiments. Prominent chemical and compound databases utilized in these fields encompass PubChem, ChemSpider, and KEGG (Kyoto Encyclopedia of Genes and Genomes). In essence, the fusion of systems biology and metabolomics has the potential to transform our comprehension of biological systems and enhance our capacity to diagnose, treat, and prevent diseases.

Keywords: System biology; Metabolomics; Databases; Network analysis and pathway analysis

1. Introduction

Systems biology is an interdisciplinary approach that aims to understand complex biological systems by integrating various types of data, such as genomics, proteomics, and metabolomics [1]. Metabolomics is the study of the small molecules, or metabolites, that are produced during cellular processes. By analyzing the metabolome, researchers can gain insights into the biochemical processes occurring within an organism [2].

Metabolomics is particularly useful in systems biology as it provides a snapshot of the metabolic state of a biological system at a given time. By combining metabolomics data with other omics data, researchers can build comprehensive models to better understand how biological systems function as a whole [3]. The integration of metabolomics into systems biology has enabled researchers to study the metabolic pathways underlying various biological processes, such as disease development, drug response, and environmental stress. This holistic approach allows for a more complete understanding of complex biological systems and can lead to the identification of novel biomarkers and therapeutic targets. Overall, the combination of systems biology and metabolomics has the potential to revolutionize our understanding of biological systems and improve our ability to diagnose, treat, and prevent diseases. Systems biology and metabolomics are two closely related fields within the broader domain of biology, each offering unique perspectives and tools for understanding biological systems at various levels of complexity [4].

2. Systems biology

Systems biology is an interdisciplinary approach to understanding the complexity of biological systems, ranging from individual cells to entire organisms and even ecosystems. It seeks to integrate diverse biological data, such as genomics, transcriptomics, proteomics, and metabolomics, with computational and mathematical models to gain insights into the behavior of biological systems as a whole.

Key concepts and approaches in systems biology include:

(1) Network analysis: Systems biologists often represent biological components (e.g., genes, proteins, metabolites) as nodes in networks, with interactions between them represented as edges. This allows for the visualization and analysis of complex biological relationships.

(2) Mathematical modeling: Mathematical models are used to describe the dynamics of biological systems and predict their behavior under different conditions. These models can range from simple kinetic equations to more complex systems of differential equations.

(3) Data integration: Systems biology aims to integrate data from multiple sources to construct comprehensive models of biological processes. This may involve combining experimental data with computational simulations to refine our understanding of biological systems [5].

3. Metabolomics

Metabolomics is a field of study focused on the systematic analysis of small molecules, known as metabolites, within biological systems. These metabolites are the end products of cellular processes and can provide valuable insights into the physiological state of an organism.

Key concepts and techniques in metabolomics include:

(1) Metabolite profiling: Metabolomics involves the comprehensive analysis of metabolites present in a biological sample. This can be done using analytical techniques such as mass spectrometry (MS) and nuclear magnetic resonance (NMR) spectroscopy.

(2) Metabolic pathway analysis: Metabolomics data can be used to reconstruct and analyze metabolic pathways within cells and organisms. By identifying changes in metabolite levels under different conditions, researchers can gain insights into the underlying metabolic processes.

(3) Metabolite identification: One of the challenges in metabolomics is the identification of individual metabolites from complex biological samples. This often involves comparing experimental data to reference databases and using techniques such as tandem MS for structural elucidation [6].

4. Integration of systems biology and metabolomics

Systems biology and metabolomics are highly complementary approaches, with metabolomics providing valuable data for the construction and validation of systems-level models. By integrating metabolomics data with other omics data and computational models, researchers can gain a more holistic understanding of biological systems and their underlying mechanisms. Systems biology and metabolomics offer powerful tools for studying the complexity of biological systems and uncovering the molecular basis of physiological processes. By combining experimental data with computational modeling, researchers can gain insights into how biological systems function and respond to changes in their environment. The integration of systems biology and metabolomics is a powerful approach that combines the study of biological systems as a whole with the analysis of small molecules present in a biological system. Systems biology aims to understand complex biological systems by considering the interactions and relationships between components, while metabolomics focuses on the comprehensive analysis of metabolites within a biological system. By integrating systems biology and metabolomics, researchers can gain a more holistic understanding of biological processes, pathways, and networks at a functional level. This integration allows for the identification of key metabolites, pathways, and regulatory mechanisms that play important roles in biological systems. Through the use of computational modeling, data integration, and advanced analytical techniques, researchers can generate comprehensive datasets that capture the dynamic changes in metabolite profiles and their interactions within a biological system. This integrated approach can provide valuable insights into disease mechanisms, biomarker discovery, drug development, and personalized medicine. Overall, the integration of systems biology and metabolomics holds great promise for advancing our understanding of complex biological systems and accelerating discoveries in various fields, including biomedicine, agriculture, and environmental science [7].

5. Chemical and compound databases in systems biology and metabolomics

Chemical and compound databases play a crucial role in systems biology and metabolomics research by providing access to information on the structure, properties, and interactions of various molecules. These databases serve as valuable resources for researchers to identify and characterize compounds, study metabolic pathways, and design experiments. In systems biology, these databases contain information on biological molecules such as proteins, genes, metabolites, and pathways, which are essential for understanding complex biological systems. By utilizing these databases, researchers can model and analyze the interactions within biological systems to gain insights into cellular processes and disease mechanisms. In metabolomics, chemical databases provide detailed information on small molecules such as metabolites, drugs, and compounds found in biological samples. These databases help researchers identify and quantify metabolites, annotate mass spectra, and link metabolites to metabolic pathways. Utilizing these databases enables the interpretation of metabolomics data, leading to the discovery of biomarkers, understanding of metabolic changes, and insight into disease mechanisms. Overall, chemical and compound databases are indispensable tools in systems biology and metabolomics research, facilitating data analysis, interpretation, and the advancement of scientific knowledge in these fields.

In systems biology and metabolomics, access to comprehensive chemical and compound databases is crucial for the identification, annotation, and interpretation of experimental data. These databases serve as repositories of information about chemical structures, properties, and biological activities, enabling researchers to link experimental observations to known compounds and biological pathways. Here are some widely used chemical and compound databases in these fields:

(1) PubChem

PubChem is a comprehensive database maintained by the National Center for Biotechnology Information (NCBI). It provides information on the biological activities of small molecules, including chemical structures, properties, bioassay data, and references. PubChem is freely accessible and widely used by researchers in various fields, including systems biology and metabolomics.

(2) ChemSpider

ChemSpider, operated by the Royal Society of Chemistry, is another extensive chemical database that provides information on chemical structures, properties, and associated data. It offers tools for compound searching, structure drawing, and property prediction, making it valuable for researchers involved in compound identification and annotation.

(3) KEGG (Kyoto Encyclopedia of Genes and Genomes)

KEGG is a comprehensive resource for understanding biological pathways and functional annotations of genes and compounds. It contains information on metabolic pathways, signaling pathways, and diseases, along with associated chemical structures and reactions. KEGG is widely used for pathway analysis and annotation in systems biology and metabolomics research.

(4) HMDB (Human Metabolome Database)

HMDB is a specialized database focused on human metabolites and metabolic pathways. It provides detailed information on metabolite structures, properties, concentrations in different biological fluids and tissues, as well as associated enzymes and pathways. HMDB is particularly useful for researchers studying human metabolism and disease.

(5) MetaboLights

MetaboLights is a database for metabolomics experiments and associated metadata. It allows researchers to submit, search, and analyze metabolomics data, including raw data files, processed data, and experimental metadata. MetaboLights facilitates data sharing and collaboration in the metabolomics community.

(6) METLIN

METLIN is a metabolite database maintained by the Scripps Center for Metabolomics. It contains information on metabolite structures, masses, and associated data, including tandem mass spectrometry (MS/MS) spectra. METLIN is widely used for metabolite identification and annotation in mass spectrometry-based metabolomics studies.

(7) ChEBI (Chemical Entities of Biological Interest)

ChEBI is a database of chemical entities with biological interest, curated and maintained by the European Bioinformatics Institute (EBI). It contains information on chemical structures, names, classifications, and biological activities of compounds. ChEBI is useful for annotating compounds and linking them to biological pathways and processes. These databases play a crucial role in systems biology and metabolomics research by providing researchers with access to curated information on chemical compounds, metabolites, and biological pathways, facilitating data interpretation, hypothesis generation, and discovery [8].

6. Metabolomics raw files databases

Metabolomics raw files contain the primary data generated from analytical instruments such as mass spectrometers or nuclear magnetic resonance (NMR) spectrometers. These files typically include spectral data,

chromatographic data, and other raw output from the instruments. While there are no dedicated databases specifically for raw metabolomics data, there are repositories where researchers can deposit and access raw data along with associated metadata. Here are some of the key repositories for metabolomics raw files:

(1) Metabolomics Workbench

Metabolomics Workbench is a comprehensive repository for metabolomics data, including raw data files, metadata, and processed data. It accepts data from various analytical platforms and provides tools for data analysis and visualization. Researchers can deposit and access metabolomics data in a standardized format, facilitating data sharing and collaboration.

(2) MetaboLights

MetaboLights, in addition to processed metabolomics data, also accepts raw data files from mass spectrometry and NMR experiments. Researchers can submit raw data along with associated metadata, making it a valuable resource for accessing and sharing metabolomics raw files.

(3) GNPS (Global Natural Products Social Molecular Networking)

GNPS is a platform for sharing mass spectrometry data, particularly focused on natural products and metabolomics. It allows researchers to upload raw mass spectrometry data, which is then processed and analyzed using molecular networking approaches to identify metabolites and annotate spectral features.

(4) MassIVE

MassIVE (Mass Spectrometry Interactive Virtual Environment) is a repository for mass spectrometry data, including metabolomics raw files. It accepts data from a wide range of mass spectrometry experiments and provides tools for data analysis and visualization. Researchers can deposit and access metabolomics data in a standardized format, promoting data sharing and reproducibility.

(5) PRIDE Archive

While primarily focused on proteomics data, PRIDE Archive also accepts mass spectrometry data from metabolomics experiments. Researchers can deposit raw mass spectrometry data along with associated metadata, enabling data sharing and integration with other omics datasets.

(6) BioStudies

BioStudies is a database for storing and sharing biological data, including metabolomics raw files. It accepts data from various omics experiments and provides tools for data submission, curation, and retrieval. Researchers can deposit metabolomics raw data along with detailed experimental descriptions and annotations.

These repositories play a crucial role in promoting data sharing, reproducibility, and collaboration in the metabolomics community by providing a platform for depositing and accessing raw metabolomics data. Researchers can utilize these resources to disseminate their data, facilitate data integration and analysis, and accelerate scientific discoveries in metabolomics research [9,10].

7. Metaboanalyst web server for metabolomics analysis

MetaboAnalyst is a web-based platform designed for comprehensive metabolomics data analysis. It offers a wide range of tools and algorithms to process, analyze, and interpret metabolomics data, making it a valuable resource for researchers in the metabolomics community. MetaboAnalyst is a popular web server for metabolomics data analysis. It provides a user-friendly platform for a wide range of tools and functions to analyze and interpret metabolomics data. Users can upload their data and perform tasks such as statistical analysis, pathway analysis, metabolic network analysis, and data visualization. MetaboAnalyst offers various

statistical methods for data normalization, transformation, and statistical testing to identify significant metabolites. It also provides tools for pathway analysis to explore the biological relevance of the identified metabolites. Users can visualize their data through interactive plots, heatmaps, and metabolic network diagrams to gain insights into their metabolomics experiments. Additionally, MetaboAnalyst supports the integration of multiple omics data types, enabling researchers to perform integrative analysis. The web server is regularly updated with new features and improvements, making it a valuable tool for researchers in the field of metabolomics. Overall, MetaboAnalyst is a comprehensive and user-friendly platform for metabolomics analysis, making it a valuable resource for researchers looking to extract meaningful biological insights from their metabolomics data.

Here's an overview of the key features and functionalities of the MetaboAnalyst web server:

(1) Data pre-processing

MetaboAnalyst supports the pre-processing of metabolomics data from various analytical platforms, including MS and NMR spectroscopy. It provides tools for data normalization, scaling, filtering, and missing value imputation to prepare the data for downstream analysis.

(2) Statistical analysis

MetaboAnalyst offers a wide range of statistical analysis tools for identifying significant differences between experimental groups, such as *t*-tests, ANOVA, and non-parametric tests. It also supports multivariate analysis techniques, including principal component analysis (PCA), partial least squares discriminant analysis (PLS-DA), and hierarchical clustering analysis (HCA), to explore patterns and relationships in the data.

(3) Pathway analysis

MetaboAnalyst integrates pathway analysis tools to identify metabolic pathways that are significantly affected by experimental conditions. It employs pathway enrichment analysis and pathway topology analysis to prioritize biologically relevant pathways and visualize metabolic networks.

(4) Metabolite Set Enrichment Analysis (MSEA)

MetaboAnalyst supports metabolite set enrichment analysis, allowing users to identify sets of metabolites that are significantly enriched in specific biological pathways, chemical classes, or other user-defined categories. It helps uncover functional insights and biological interpretation from metabolomics data.

(5) Time-series analysis

For metabolomics datasets with temporal or sequential measurements, MetaboAnalyst provides tools for time-series analysis to identify dynamic changes in metabolite concentrations over time. It includes methods for trend analysis, clustering, and visualization of temporal profiles.

(6) Integration with other omics data

MetaboAnalyst supports integrative analysis of metabolomics data with other omics datasets, such as transcriptomics and proteomics. It enables researchers to explore multi-omics relationships and identify coordinated changes across different molecular layers.

(7) User-friendly interface and visualization

MetaboAnalyst features a user-friendly web interface with interactive visualization tools for data exploration and interpretation. It provides customizable plots, heatmaps, and pathway diagrams to visualize metabolomics data and analysis results.

(8) Educational resources and workflows

MetaboAnalyst offers educational resources, tutorials, and workflow templates to guide users through the data analysis process. It provides step-by-step instructions and examples to help researchers

perform metabolomics analysis effectively.

MetaboAnalyst is a powerful and user-friendly platform for metabolomics data analysis, offering a comprehensive suite of tools and functionalities to support a wide range of research applications in metabolomics and systems biology [11,12].

8. Metabolomics analysis using MzMine 3 software

MzMine 2 and MzMine 3 are open-source software tools specifically designed for the pre-processing, analysis, and visualization of MS data, including metabolomics data. MzMine enables users to perform various data processing steps, such as peak detection, alignment, normalization, and feature extraction, to extract meaningful information from raw mass spectrometry data. Here's an overview of how MzMine 3 can be used for metabolomics analysis:

(1) Data import

MzMine 3 allows users to import raw mass spectrometry data files from different vendors and formats, including mzML, mzXML, mzData, and proprietary formats. Users can specify the parameters for data import, including scan range, mass range, and data processing settings.

(2) Peak detection

MzMine 3 offers various algorithms for peak detection, which identify peaks corresponding to metabolites in the mass spectra. Users can choose from different peak picking methods, such as centroid, wavelet, or matched filtering, and adjust parameters like signal-to-noise ratio and peak width to optimize peak detection.

(3) Chromatographic alignment

MzMine 3 provides tools for chromatographic alignment, which aligns peaks across multiple samples to account for retention time shifts and improve data consistency. Users can perform alignment based on retention time or m/z values, using algorithms like dynamic time warping (DTW) or correlation-based alignment.

(4) Data filtering and normalization

MzMine 3 enables users to filter and normalize the data to remove noise and systematic variations. It offers options for baseline correction, noise filtering, peak area normalization, and intensity scaling to ensure data quality and comparability across samples.

(5) Feature extraction and annotation

MzMine 3 allows users to extract features from the pre-processed data, which represent metabolites or molecular entities detected in the mass spectra. Users can annotate features using databases such as HMDB or METLIN to identify metabolites based on accurate mass and fragmentation patterns.

(6) Statistical analysis

MzMine 3 supports statistical analysis tools for comparing metabolomics data between experimental groups and identifying significant differences. Users can perform univariate analysis (e.g., *t*-tests, ANOVA) and multivariate analysis (e.g., PCA, PLS-DA) to identify metabolite biomarkers and patterns associated with experimental conditions or phenotypes.

(7) Visualization

MzMine 3 provides visualization tools for exploring and visualizing metabolomics data. Users can generate chromatograms, mass spectra, heatmaps, volcano plots, and pathway maps to visualize and interpret the results of data analysis.

(8) Workflow management

MzMine 3 allows users to create and save data analysis workflows, enabling reproducibility and automation of data processing steps. Users can save workflow templates and parameter settings for future use or sharing with collaborators.

MzMine 3 is a versatile software tool for metabolomics data analysis, offering a wide range of functionalities for processing, analyzing, and visualizing mass spectrometry data. It provides a user-friendly interface and customizable workflows to facilitate data analysis and interpretation in metabolomics research [13].

9. Overview of pathway databases

Pathway databases are repositories of curated biological pathways, which represent interactions among genes, proteins, metabolites, and other biomolecules involved in specific cellular processes or functions. These databases play a crucial role in systems biology, bioinformatics, and computational biology by providing structured knowledge. Pathway databases are crucial resources in biological research, providing organized collections of information about molecular pathways and networks involved in various cellular processes. Here are some notable pathway databases:

(1) KEGG (Kyoto Encyclopedia of Genes and Genomes)

KEGG is one of the most comprehensive pathway databases, providing information on metabolic pathways, regulatory networks, and other biological processes. It includes curated pathway maps, molecular diagrams, and associated gene/protein annotations for various organisms. KEGG is widely used for pathway analysis and functional annotation in genomics, systems biology, and drug discovery.

(2) Reactome

Reactome is an open-access pathway database focusing on human biology and disease. It provides detailed pathway diagrams, molecular interactions, and functional annotations for a wide range of biological processes, including metabolism, signaling, and immune response. Reactome supports pathway enrichment analysis and pathway visualization tools for data interpretation.

(3) WikiPathways

WikiPathways is a community-curated pathway database that allows researchers to collaboratively contribute, edit, and annotate pathway information. It covers a broad range of biological pathways, including metabolism, cellular processes, and diseases. WikiPathways emphasizes data accessibility and openness, making it a valuable resource for both researchers and educators.

(4) BioCyc

BioCyc is a collection of pathway/genome databases (PGDBs) covering a wide range of organisms, from bacteria to eukaryotes. Each BioCyc database contains curated pathway information, metabolic models, and genome annotations for a specific organism. BioCyc databases are extensively used for metabolic pathway analysis, systems biology modeling, and comparative genomics studies.

(5) MetaCyc

MetaCyc is a reference database of metabolic pathways and enzymes, containing information curated from the scientific literature. It includes experimentally determined metabolic pathways, enzyme reactions, and enzyme function annotations for a diverse set of organisms. MetaCyc is widely used for metabolic pathway reconstruction and analysis in systems biology and bioinformatics.

(6) PathBank

PathBank is a comprehensive pathway database focusing on human metabolism. It integrates data from various sources, including KEGG, Reactome, and other pathway databases, to provide a unified

resource for metabolic pathway analysis. PathBank includes metabolite-centric pathway maps, pathway enrichment analysis tools, and data visualization features for exploring metabolic networks.

(7) Panther Pathways

Panther Pathways is a pathway database developed by the Protein Analysis Through Evolutionary Relationships (PANTHER) classification system. It provides curated pathway diagrams, gene/protein annotations, and functional classifications for various biological pathways and processes. Panther Pathways integrates pathway data with protein family information, enabling comprehensive functional analysis of genes and proteins.

These pathway databases serve as valuable resources for understanding the molecular basis of biological processes, exploring gene/protein functions, and interpreting high-throughput omics data in the context of biological pathways and networks. Researchers often utilize these databases for pathway enrichment analysis, pathway visualization, and hypothesis generation in various fields of biological and biomedical research [14,15].

10. Pathway and correlation-based systems biology approaches

Pathway and correlation-based systems biology approaches are two distinct but complementary strategies used to study complex biological systems at the molecular level. Pathway and correlation-based systems biology approaches are two key methods used to study complex biological systems. Pathway-based approaches focus on the interactions and relationships between biological molecules within specific pathways or networks. By analyzing these pathways, researchers can gain insights into how different molecules work together to regulate various biological processes. On the other hand, correlation-based approaches examine the statistical relationships between different molecular features or variables. By identifying correlations between molecules, researchers can uncover potential functional relationships or regulatory interactions within a biological system.

Both pathway and correlation-based approaches have their strengths and limitations. Pathway-based approaches provide a structured framework for understanding biological processes but may overlook novel interactions or pathways. In contrast, correlation-based approaches can reveal unexpected associations but may not always infer causality. Integrating both pathway and correlation-based approaches can provide a more comprehensive understanding of complex biological systems. By combining these methods, researchers can uncover new insights into the underlying mechanisms driving biological processes and diseases. Let's delve into each approach!

10.1. Pathway-based systems biology

Pathway-based systems biology focuses on understanding biological systems through the analysis of molecular pathways, which are networks of interacting genes, proteins, metabolites, and other biomolecules involved in specific biological processes. Here's how this approach works:

(1) Pathway identification: Researchers identify relevant biological pathways associated with a particular cellular process, disease, or experimental condition using pathway databases such as KEGG, Reactome, or WikiPathways.\

(2) Pathway analysis: Once pathways of interest are identified, researchers analyze omics data (e.g., transcriptomics, proteomics, metabolomics) to determine how genes, proteins, or metabolites within these pathways are modulated under different conditions. This analysis may involve pathway enrichment analysis to identify overrepresented pathways in the dataset.

(3) Pathway visualization: Pathway maps are used to visualize the interactions and relationships between molecules within a pathway. Visualization tools such as Cytoscape or PathVisio allow researchers to

create graphical representations of pathways and overlay experimental data for interpretation.

(4) Pathway modeling: Computational modeling techniques, such as kinetic models or constraint-based models, can be used to simulate the dynamics of pathway activity and predict how changes in pathway components affect cellular phenotypes.

10.2. Correlation-based systems biology

Correlation-based systems biology aims to uncover relationships and interactions between molecular components (e.g., genes, proteins, metabolites) based on statistical correlations observed in high-dimensional omics datasets. Here's an overview of this approach:

(1) Data integration: Multiple omics datasets (e.g., gene expression, protein abundance, metabolite concentrations) are collected from the same biological samples under different conditions or treatments.

(2) Correlation analysis: Statistical methods, such as Pearson correlation coefficient, Spearman rank correlation, or mutual information, are used to quantify the degree of association between pairs of molecular entities across samples. Positive correlations suggest co-regulation or co-expression, while negative correlations indicate potential regulatory relationships.

(3) Network inference: Correlation matrices obtained from correlation analysis can be used to construct molecular interaction networks, where nodes represent molecular entities and edges represent significant correlations between them. Network inference algorithms, such as weighted gene co-expression network analysis (WGCNA) or partial correlation-based methods, are used to identify modules or clusters of highly correlated molecules.

(4) Network visualization and analysis: Once the molecular interaction network is constructed, visualization tools like Cytoscape or Gephi are employed to visualize the network structure and identify central nodes (hubs) or densely connected subnetworks (modules). Functional enrichment analysis may be performed to elucidate the biological functions and pathways associated with network modules.

(5) Validation and functional interpretation: Correlation-based findings are validated using experimental techniques such as knockout experiments, RNA interference (RNAi), or targeted metabolomics/proteomics assays. Functional interpretation of network modules involves identifying enriched biological processes, pathways, or protein-protein interaction motifs.

In summary, pathway-based and correlation-based systems biology approaches offer complementary strategies for deciphering the complexity of biological systems, from the level of individual pathways to the global network of molecular interactions. Integrating these approaches allows for a more comprehensive understanding of biological processes and the identification of key regulatory mechanisms underlying cellular function and disease [16,17].

10.3. Network construction and analysis in Cytoscape

Cytoscape is a powerful open-source software platform for visualizing and analyzing complex networks, particularly biological networks such as protein-protein interaction networks, gene regulatory networks, and metabolic pathways. Cytoscape is a powerful software platform for visualizing and analyzing biological networks.

By following these steps, you can effectively construct, visualize, and analyze biological networks in Cytoscape to gain insights into complex biological systems. Here's a basic overview of how to construct and analyze networks in Cytoscape:

(1) Data import: Start by importing your network data into Cytoscape. This can be done by loading files in

various formats such as edge lists, and adjacency matrices, or through direct interaction with databases. Cytoscape supports a wide range of standard network data formats.

(2) Network construction: Once your data is imported, you can construct your network by adding nodes and edges. Nodes represent entities (e.g., genes, proteins) and edges represent interactions or relationships between them. You can manually add nodes and edges, or use Cytoscape's built-in algorithms for automatic layout and edge creation.

(3) Visual styling: Customize the visual appearance of your network to better convey the information it contains. You can adjust node and edge colors, sizes, shapes, and labels, as well as apply different layout algorithms to arrange the network in a visually pleasing way.

(4) Network analysis: Cytoscape provides a wide range of analysis tools to explore and analyze the structure and properties of your network. This includes calculating network metrics such as degree distribution, clustering coefficient, and betweenness centrality, as well as performing more advanced analyses like network motif detection and community detection.

(5) Data integration: Cytoscape allows you to integrate additional data layers onto your network, such as gene expression data or functional annotations. This can help you gain deeper insights into the biological significance of different network components.

(6) Visualization and exploration: Use Cytoscape's interactive visualization tools to explore your network in detail. You can zoom, pan, and navigate through large networks, as well as apply filters and search functions to focus on specific subsets of nodes and edges.

(7) Data export: Once you've analyzed and visualized your network, you can export your results in various formats for further analysis or publication. Cytoscape supports exporting networks as images, PDFs, and various data formats compatible with other network analysis software.

Cytoscape's user-friendly interface, extensive documentation, and active community support make it a popular choice for researchers working with complex networks in various domains. Whether you're studying biological systems, social networks, or any other type of networked data, Cytoscape provides powerful tools for network construction and analysis [18,19].

11. Conclusion

Systems biology is an interdisciplinary approach that combines various data sources, including genomics, proteomics, and metabolomics, to understand complex biological systems. Metabolomics, which examines small molecules produced during cellular processes, provides insights into biochemical activities within an organism. By combining metabolomics data with other omics data, researchers can create comprehensive models that enhance their understanding of biological systems. This approach can lead to the discovery of new biomarkers and therapeutic targets. Systems biology and metabolomics are closely linked disciplines, offering different perspectives and methodologies for interpreting biological systems. By integrating these methodologies, researchers can gain a holistic understanding of biological processes, pathways, and networks. Chemical and compound databases, such as PubChem, ChemSpider, and KEGG, are crucial in these fields. The fusion of systems biology and metabolomics has the potential to transform our understanding of biological systems and improve disease diagnosis, treatment, and prevention.

Disclosure statement:

The author declares that there is no conflict of interest in this paper.

References

[1] Dominguez PG, Abreu IN, Moritz T, et al., 2024, Systems Biology as a Tool to Uncover Interdisciplinary Links within the Complex Forest Tree System. Monitoring Forest Damage with Metabolomics Methods, 71–108.

[2] Dey A, Ghosh S, Majhi S, et al., 2024, Systems Biology Approaches as a Tool in Understanding Neurodevelopmental Disorders: Some Case Studies. Systems Biology Approaches: Prevention, Diagnosis, and Understanding Mechanisms of Complex Diseases, 511–536.

[3] Jasmine R, Soundararajan S, Rosita AS, et al., 2024, Systems Biology and Human Diseases. Systems Biology Approaches: Prevention, Diagnosis, and Understanding Mechanisms of Complex Diseases, 27–54.

[4] Wishart, DS, 2019, Metabolomics for Investigating Physiological and Pathophysiological Processes. Physiological Reviews, 99(4): 1819–1875.

[5] Bruggeman FJ, Westerhoff HV, 2007, The Nature of Systems Biology. Trends in Microbiology, 15(1): 45–50.

[6] Wishart, DS, 2009, Computational Strategies for Metabolite Identification in Metabolomics. Bioanalysis, 1(9): 1579–1596.

[7] Wang JH, Byun J, Pennathur S, 2010, Analytical Approaches to Metabolomics and Applications to Systems Biology. Seminars in Nephrology, 30(5): 500–511.

[8] Kell DB, 2006, Systems Biology, Metabolic Modelling and Metabolomics in Drug Discovery and Development. Drug Discovery Today, 11(23–24): 1085–1092.

[9] Misra BB, 2021, New Software Tools, Databases, and Resources in Metabolomics: Updates from 2020. Metabolomics, 17(5): 49.

[10] O'Shea K, Misra BB, 2020, Software Tools, Databases and Resources in Metabolomics: Updates from 2018 to 2019. Metabolomics, 16(3): 36.

[11] Xia J, Psychogios N, Young N, et al., 2009. MetaboAnalyst: A Web Server for Metabolomic Data Analysis and Interpretation. Nucleic Acids Research, 37(suppl_2): 652–660.

[12] Xia J, Wishart DS, 2016, Using MetaboAnalyst 3.0 for Comprehensive Metabolomics Data Analysis. Current Protocols in Bioinformatics, 55(1): 14–10.

[13] Pluskal T, Korf A, Smirnov A, et al., 2020, Metabolomics Data Analysis Using MZmine. Processing Metabolomics and Proteomics Data with Open Software: A Practical Guide, 232–254.

[14] Go EP, 2010, Database Resources in Metabolomics: An Overview. Journal of Neuroimmune Pharmacology, 2010(5): 18–30.

[15] García-Campos MA, Espinal-Enríquez J, Hernández-Lemus E, 2015, Pathway Analysis: State of the Art. Frontiers in Physiology, 2015(6): 683.

[16] Armitage EG, Kotze HL, Williams KJ, 2014, Correlation-based Network Analysis of Cancer Metabolism. A New Systems Biology Approach in Metabolomics, 1–61.

[17] Zhu J, Zhang B, Schadt EE, 2008, A Systems Biology Approach to Drug Discovery. Advances in Genetics, 2008(60): 603–635.

[18] Li M, Yang J, Wu FX, et al., 2018, DyNetViewer: A Cytoscape App for Dynamic Network Construction, Analysis and Visualization. Bioinformatics, 34(9): 1597–1599.

[19] Doncheva NT, Morris JH, Gorodkin J, et al., 2018, Cytoscape StringApp: Network Analysis and Visualization of Proteomics Data. Journal of Proteome Research, 18(2): 623–632.