

REVIEW

Open Access



Network-based multi-omics integrative analysis methods in drug discovery: a systematic review

Wei Jiang¹, Weicai Ye², Xiaoming Tan¹ and Yun-Juan Bao^{1,3*}

*Correspondence:

Yun-Juan Bao

yjbao@hubu.edu.cn

¹School of Life Sciences, Hubei University, Wuhan, China

²School of Computer Science and Engineering, Guangdong Province

Key Laboratory of Computational

Science, National Engineering

Laboratory for Big Data Analysis

and Application, Sun Yat-sen

University, Guangzhou, China

³No.368 Youyi Avenue,

Wuhan 430062, China

Abstract

The integration of multi-omics data from diverse high-throughput technologies has revolutionized drug discovery. While various network-based methods have been developed to integrate multi-omics data, systematic evaluation and comparison of these methods remain challenging. This review aims to analyze network-based approaches for multi-omics integration and evaluate their applications in drug discovery. We conducted a comprehensive review of literature (2015–2024) on network-based multi-omics integration methods in drug discovery, and categorized methods into four primary types: network propagation/diffusion, similarity-based approaches, graph neural networks, and network inference models. We also discussed the applications of the methods in three scenarios of drug discovery, including drug target identification, drug response prediction, and drug repurposing, and finally evaluated the performance of the methods by highlighting their advantages and limitations in specific applications. While network-based multi-omics integration has shown promise in drug discovery, challenges remain in computational scalability, data integration, and biological interpretation. Future developments should focus on incorporating temporal and spatial dynamics, improving model interpretability, and establishing standardized evaluation frameworks.

Keywords Multi-omics, Biological network, Drug discovery, Precision medicine, Network analysis, Data integration

Introduction

Human body experiences millions of signals transferred and complex interactions every second between cells, tissues, organs, and external environmental stimuli. This miraculous biological mechanism operating at an ultra-microscopic level that remains elusive to capture and unveil. Understanding complex biological systems has been an on-going quest for many researchers. The rapidly decreasing costs of high-throughput sequencing, development of massively parallel technologies, and new sensor technologies have enabled us to collect a large amount of biological data [1]. These data include tissue exome sequencing, copy number variation (CNV), DNA methylation, gene expression (Fig. 1), and microRNA (miRNA) expression, as well as some physiological and clinical data such as race, tumor stage, relapse, and treatment response [2–4]. Currently, there



© The Author(s) 2025. **Open Access** This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

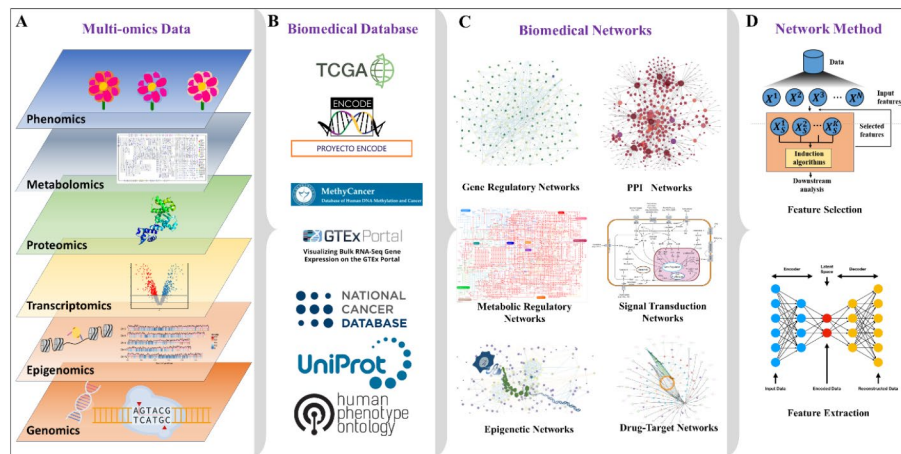


Fig. 1 A workflow diagram from omics data generation to network-based analysis. **(A)** Integration of multi-omics across the major omics layers (genomics, epigenomics, transcriptomics, proteomics, metabolomics, and phenomics); **(B)** The aggregation and archiving of multi-omics data into specialized databases, showcasing how these data repositories support standardization, preservation, and accessibility of large-scale biological data; **(C)** Multi-omics data extracted from databases are used to construct various biological networks (GRNs, PPI networks, MRNs, STNs, epigenetic networks, DTIs). The complex interactions in these networks reflect different biological activities; **(D)** The application of network-based analytical techniques to decipher the constructed networks. It includes methods like network feature selection methods and feature extraction methods

are numerous single omics approaches, which investigate how these data from distinct molecular layers contribute to the manifestation and progression of diverse biological mechanisms, medical challenges, and pharmacological applications [5–7].

However, a complete biological process always accompanies various regulations or chain-reactions, therefore, no single data type can capture the complexity of all the factors relevant to understanding a phenomenon such as a disease [8, 9]. The holistic understanding of the molecular and cellular bases of disease phenotypes and normal physiological processes requires integrated investigations of the contributions and associations between multiple (different but parallel) molecular layers driving the observed outcome. Interestingly, some studies revealed the wide complementarity and better understanding of biological feature in different omics. For instance, PRO-seq (Precision Run-On sequencing) for measuring nascent RNA can directly capture rapid transcriptional changes, identifying genes that are early transcriptionally activated or repressed by HIF1A. Combined with measurement of HIF1A chromatin binding and steady-state mRNA levels in multiple cell lines, allowing for an in-depth analysis of the transcriptional activation mechanisms mediated by HIF1A [10]. Huang et al. combined single-cell transcriptomics and metabolomics data to delineate how NNMT-mediated metabolic reprogramming, through modulating E-cadherin expression, drives lymph node metastasis in esophageal squamous cell carcinoma. Such cross-level, multidimensional molecular profiling provides novel insights into disease mechanisms and precision medicine [11]. Another notable example is the work by Liao et al., they integrated multi-omics data spanning genomics, transcriptomics, DNA methylation, and copy number variations of SARS-CoV-2 virus target genes across 33 cancer types. This comprehensive analysis elucidated the genetic alteration patterns, expression differences, and clinical prognostic associations of these genes [12].

These studies widely proved integrating multi-omics data can comprehensively delineate the connections and influences between different biological strata, possessing

significant advantages. However, another challenging issue is that multi-omics studies include data that differ in type, scale and source, with often thousands of variables and only few samples. Additionally, biological datasets are complex, noisy, biased, heterogeneous, with potential errors due to measurement mistakes or unknown biological deviations [7, 13–15]. Utilizing pertinent information and integrating the omics into a meaningful model are therefore difficult and a great number of methods and strategies have been developed in recent years to tackle this challenge. If the integration is not done correctly, adding more omics might not result in a significant increase of performance, but will increase the complexity of the problem along with computational time. Therefore, it is essential to select an appropriate approach for the accurate processing of multi-omics data.

It is widely acknowledged that biomolecules do not perform their functions alone, rather, they interact with one another to form biological networks. For example, genes interact in various pathways and protein complexes, and the cancerous potential of a cell is a consequence of the disruption of the pathway, but not necessarily the mutation of one specific gene within the pathway [2, 7]. Biological networks constitute the foundational framework of biological systems, that are derived from different sources and that cover different scales (Fig. 1). Prominent examples are co-expression networks, co-evolution networks, metabolic pathways, protein–protein interaction (PPI) networks or drug–target interaction (DTI) networks [16–18]. Within these networks, nodes represent individual molecules, such as genes, proteins, and DNA. The connections between nodes, termed edges, reflect the relationships among them. Studying these ubiquitous organizational features of networks across biological systems has yielded the promise of discovering universal fundamentals of multi-omics and shown the global patterns. Therefore, abstracting the interactions among various omics in biology into network models aligns with the principles of biological systems, and is one of the hot research topics in multi-omics data mining, especially in the fields of drug prediction and disease mechanisms [19].

Due to the complexity of biological systems and the high failure rates of traditional methodologies, drug discovery is increasingly challenged. Network-based multi-omics integration offers unique advantages for drug discovery, as these approaches can capture the complex interactions between drugs and their multiple targets. By integrating various molecular data types and performing network analyses, such methods can better predict drug responses, identify novel drug targets, and facilitate drug repurposing [20, 21]. Despite the proliferation of network-based integration methods, several critical challenges remain unaddressed. First, the field lacks standardized frameworks for evaluating and comparing different integration methods, making it difficult to select appropriate approaches for specific applications. Second, many current methods struggle with computational efficiency when handling large-scale multi-omics datasets. Third, maintaining biological interpretability while increasing model complexity remains a significant challenge. There have been several reviews surveying multi-omics integration methods and their applications in drug discovery [15, 22–24]. These reviews have provided valuable insights into the mathematical mechanisms of data integration, offering detailed discussions on the implementation of omics data fusion techniques such as network-based integration, dimensionality reduction, and machine learning approaches. They have also contributed significantly to understanding the benefits of integrating

diverse biological data types, such as genomics, transcriptomics, and proteomics, in drug discovery contexts. However, there are some limitations in those reviews. They primarily focus on either the mathematical aspects of data integration or specific applications, without comprehensively examining the interplay between network biology and multi-omics integration, which is essential for drug discovery. Our review aims to fill this gap by offering a holistic perspective on how network-based approaches can unify multi-omics data for more accurate predictions in drug discovery.

This review aims to address these gaps by providing a systematic analysis of network-based multi-omics integration methods and their applications in drug discovery, offers several distinctive contributions to the field: (1) A novel classification framework that systematically categorizes network-based integration methods according to their algorithmic principles and biological applications, thereby providing clearer guidance for method selection; (2) A comprehensive evaluation of the contributions of various network types (such as gene regulatory networks (GRNs), protein interaction networks (PINs), metabolic reaction networks (MRNs), etc.) to drug discovery, offering insights into the biological relevance of each network type; (3) An in-depth analysis of the technological evolution from traditional network analysis to modern AI-driven approaches, particularly in the context of drug discovery applications; (4) Explore the application and performance of different approaches using biological networks in various aspects of drug discovery, including target identification, drug response prediction, and drug repurposing; (5) A thorough assessment of method performance across different drug discovery tasks, supported by recent case studies spanning 2015–2024; and (6) A systematic examination of how network-based methods address the integration of diverse omics data types, tackling practical challenges related to data heterogeneity and dimensionality. This review serves as a valuable resource for researchers seeking to understand and apply network-based multi-omics integration methods in drug discovery, while also highlighting critical areas for future methodological development.

Methods

Literature search and collection strategy

Our comprehensive review focused on network-based multi-omics integration methods in drug discovery published between January 2015 and December 2024. The literature collection process employed a three-tiered strategy combining systematic database searching, expert recommendations, and citation network analysis. We conducted systematic searches across major scientific databases including PubMed, Web of Science, and IEEE Xplorer. The search strategy utilized carefully constructed combinations of key terms: (“multi-omics” OR “multiomics” OR “omics fusion”) AND (“network analysis” OR “biological network” OR “biological interaction”) AND (“drug discovery” OR “drug prediction” OR “drug development”). This initial search was supplemented by scanning proceedings from leading computational biology conferences such as ISMB, RECOMB, and PSB, where cutting-edge methodological developments were often first presented.

To ensure the quality and relevance of our analysis, we applied stringent inclusion and exclusion criteria. The inclusion criteria for the selected studies were as follows: (1) presented original research developing or significantly improving network-based methods for multi-omics integration; (2) demonstrated clear applications in drug discovery or drug response prediction; (3) provided performance evaluation using standard metrics;

(4) utilized multiple types of omics data (minimum of two omics layers); and (5) incorporated biological network information in their analytical framework.

Studies were excluded if they met any of the following criteria: (1) review articles, perspectives, and conference abstracts; (2) studies focusing solely on single-omics analysis; (3) methods without network components or biological validation; and (4) papers without publicly accessible implementation details.

To capture the full scope of methodological developments and their applications, we employed citation network analysis, involving both forward citations tracking of seminal papers and backward reference scanning of recent comprehensive reviews. We paid particular attention to works that bridged methodological innovation with practical applications in pharmaceutical research. Papers were initially screened based on titles and abstracts, followed by full-text review of potentially eligible studies.

Method summary framework

Our summary framework for network-based multi-omics integration methods was designed to systematically evaluate how different approaches utilize biological networks and integrate multi-omics data for drug discovery applications. We structured our summary around three core components: biological network analysis, integration mechanism assessment, and algorithmic framework evaluation. Specifically, we first conducted a systematic evaluation of approaches for constructing various biological networks as foundational frameworks for data integration. Subsequently, we investigated the mechanisms by which different multi-omics methods incorporate and synthesize diverse omics data within the network context. Through comprehensive literature analysis, we then established a classification system for the predominant methodological categories in this field.

Results

Biological networks and their relationship with diseases

The integration of multi-omics data with network biology has led to the realization that diseases are rarely the result of a single molecular aberration; rather, they arise from perturbations in complex, interconnected biological networks. The network can be constructed based upon our current knowledge (i.e., genetic, physical, or biochemical interactions obtained from databases or previous studies). In recent years, comprehensive databases on various biological pathways and molecular interaction networks have been continuously developed, with representative databases including Reactome [25], Signor [26], OmniPath [27], and SPOKE [28]. Moreover, some important multi-relational drug-gene or disease database have also been established, such as CIVIC [29], Tri@DB [54], PharmGKB [30], TTD [31] and so on (Tables 1 and 2). Through the application of suitable methods to map multi-omics data onto these network structures, investigators can delineate the systemic changes that occur during disease progression, identify key molecular drivers, and uncover potential therapeutic targets (Fig. 1). This network-centric approach to disease biology not only captures the inherent complexity of biological systems but also provides a robust framework for integrating diverse types of molecular data. It is possible to explore how network-based multi-omics data analysis can reshape our understanding of complex diseases and guide innovative approaches to

Table 1 Key databases for network-based multi-omics integration in drug discovery

Category	Database Name	Description	Main Features	Reference
Biological Networks	STRING	Protein-protein interactions	Confidence scoring, comprehensive coverage, API access, functional enrichment	[32]
	BioGRID	Physical/genetic/protein-protein interactions	Literature-curated, experimental evidence, includes human and model organisms	[33]
	Reactome	Pathway database	Detailed reaction/pathway information, visualization tools, pathway enrichment	[25]
	KEGG	Pathway database	Metabolic and signaling pathways, disease pathways, integration with other omics data	[33]
	SIGNOR 3.0	Signaling network resource	Causal relationships in signal transduction, pathway visualization	[26]
	OmniPath	Molecular interaction database	Multi-source data integration, PTM relationships, protein complexes, intercellular communication	[27]
Multi-Omics Data	TCGA	Cancer multi-omics data	Genomics, transcriptomics, clinical data, comprehensive cancer types	[34]
	CCLC	Cancer cell line encyclopedia	Gene expression, mutation, drug response data, various cancer types	[35]
	CPTAC	Proteogenomics data	Proteomic and phosphoproteomic profiles, clinical correlation	[36]
	GTEx	Normal tissue expression	Tissue-specific expression profiles, integration with disease data	[37]
	SPOKE	Knowledge integration engine	Multi-layer network integration, biomedical knowledge graph	[28]
	Tri@DB	Cancer precision medicine	Gene-disease-therapy relationships, population carrier rates, pathway analysis, automated reporting system	[54]
Disease Associations	DisGeNET	Gene-disease associations	Evidence-based scoring, comprehensive coverage, cross-species data	[38]
	OMIM	Genetic disorders	Detailed phenotype information, inheritance patterns, curated genetic data	[39]
	CIVIC	Clinical interpretations	Clinical evidence summaries, drug-disease relationships, evidence quality scoring	[29]

drug discovery. The principal categories of biological networks, each with distinct characteristics and implications for disease etiology, are as follows:

Gene regulatory networks (GRNs)

GRNs encapsulate the complex regulatory mechanisms governing gene expression. They comprise a diverse array of elements, including transcription factors, cis-regulatory modules, and non-coding RNAs [45, 46]. The dynamic nature of GRNs underlies cellular differentiation, homeostasis, and adaptive responses to environmental stimuli [47]. Perturbations in GRNs have been implicated in various pathological states, notably in cancer, where dysregulation of key regulatory nodes can precipitate uncontrolled proliferation and metastasis [48]. Deciphering GRNs in cancerous cells can identify potential drug targets, laying the groundwork for precision medicine and tailored treatments.

Protein-Protein interaction networks (PINs)

PINs delineate the physical and functional associations between proteins, forming the backbone of cellular signal transduction and metabolic pathways [49, 50]. Characterized by scale-free and small-world properties, PINs exhibit a hierarchical organization with highly connected “hub” proteins serving as critical nodes. The disruption of these

Table 2 Drug-related databases for network-based integration

Database Name	Data Type	Key Features	Applications in Drug Discovery	Reference
DrugBank	Drug information	Chemical, pharmacological, and pharmaceutical data, FDA-approved drugs	Target identification, mechanism of action, drug repurposing	[40]
GDSC	Drug sensitivity	Cancer cell line drug response data, genetic profiles	Drug response prediction, precision oncology, biomarkers	[41]
PubChem	Chemical information	Chemical structures, bioactivity data, compound properties	Drug similarity analysis, virtual screening, compound library screening	[42]
PharmGKB	Pharmacogenomics	Genetic variation effects on drug response, drug interactions	Precision medicine, genetic-guided therapy, drug optimization	[30]
TTD	Therapeutic targets	Drug-target interactions, therapeutic indications	Target validation, drug development, target discovery	[31]
LINCS	Drug perturbation	Gene expression signatures induced by small molecule perturbations	Mechanism prediction, drug repurposing, response prediction	[43]
ChEMBL	Bioactivity data	Binding, functional, ADMET (Absorption, Distribution, Metabolism, Excretion, Toxicity) data	Target prediction, bioactivity profiling, lead optimization	[44]

hubs often has far-reaching consequences, potentially leading to systemic dysfunction and disease manifestation [50, 51]. Notably, pathogens frequently exploit host PINs to subvert cellular machinery, underscoring the networks' significance in host-pathogen interactions [52].

Metabolic regulatory networks (MRNs)

MRNs represent the intricate web of biochemical reactions governing cellular metabolism. These networks are orchestrated by enzymes and regulatory molecules, employing sophisticated control mechanisms such as allosteric regulation and feedback inhibition [53–55]. The plasticity of MRNs enables cellular adaptation to varying nutritional and environmental conditions. Dysregulation of MRNs is associated with metabolic disorders, exemplified by the intricate interplay between insulin signaling and glucose homeostasis in diabetes mellitus [56].

Epigenetic networks

These networks encompass the complex interplay of DNA methylation, histone modifications, and chromatin remodeling processes [57, 58]. Epigenetic networks mediate the interface between environmental stimuli and genomic responses, playing a crucial role in cellular memory and phenotypic plasticity. Aberrations in epigenetic networks have been implicated in a spectrum of diseases, from neurodevelopmental disorders to cancer, where they can lead to inappropriate silencing of tumor suppressor genes or activation of oncogenes [57–59].

Other networks

Disease-Gene Networks (DGNs) and Drug-Target Networks (DTNs) represent higher-order networks that integrate genetic, molecular, and pharmacological data [60–64]. DGNs map the complex relationships between genetic variants and disease phenotypes, while DTNs elucidate the interactions between pharmacological agents and their

molecular targets. The integration of these networks provides a systems-level perspective on disease mechanisms and therapeutic interventions, facilitating the identification of novel drug targets and repurposing opportunities [62–64]. The integration of DGNs and DTNs becomes particularly powerful in translational medicine and drug discovery. By overlaying these networks, researchers can identify which disease-associated genes are also drug targets, offering a dual perspective on how genetic alterations contribute to disease and how they might be therapeutically modulated [64].

The structural and functional properties of biological networks, as detailed above, provide the foundation for developing predictive algorithms. The following section evaluates how these network features are leveraged to address three critical challenges in drug prediction: target prioritization, patient stratification, and combination therapy design.

Network-based multi-omics integration methods

Multi-omics integration leverages data from genomics, transcriptomics, proteomics, metabolomics, and other omics layers to provide a holistic view of biological systems. The integration of multi-omics data using network approaches enables the identification of novel biomarkers and optimal treatment strategies by finding patterns and relationships that may not be evident when considering single omics datasets in isolation [22, 65, 66].

The overall concept of network-based integration approaches is to project multi-omics experimental data into our existing knowledge framework as distinct network representations, subsequently aiming to pinpoint critical interactions and modules to unravel the underlying systems biology. For example, many studies use a large number of genome-scale networks, including protein-protein and genetic interaction networks, which are now available for several organisms, to predict genes that cause a particular phenotype or have a particular function [67, 68]. Some integration methods build drug-target interaction networks or similarity networks between cell lines and between drug descriptors and then use different network analysis techniques to predict drug responses [69].

In these methods, which are combined with biological networks, we categorize them based on principal algorithms and analytical approaches into four main types: network propagation/diffusion (21%), similarity-based methods (19%), network-based inference methods (24%), and artificial neural networks (36%). These methods integrate data from various omics layers either into a single matrix or by receiving all data simultaneously for dimensionality reduction prior to analysis, subsequently utilized for downstream decision-making tasks (Fig. 2). We believe this focus is more practical compared to broader, methodology-oriented approaches. We aim to explore how these methods can be applied to both broad and specific research questions, illustrating specific cases and demonstrating how each method provides a unique analytical perspective.

Network propagation/diffusion methods

Network propagation/diffusion methods have emerged from theoretical studies in social sciences and mathematics, applied to analyze and comprehend how information, innovation, or behaviors are disseminated across populations through social and communication networks [70]. In biology, network propagation models are grounded in the concept that biological information does not operate in isolation but rather traverses a web of interconnected pathways and interactions [71]. By leveraging this principle, propagation

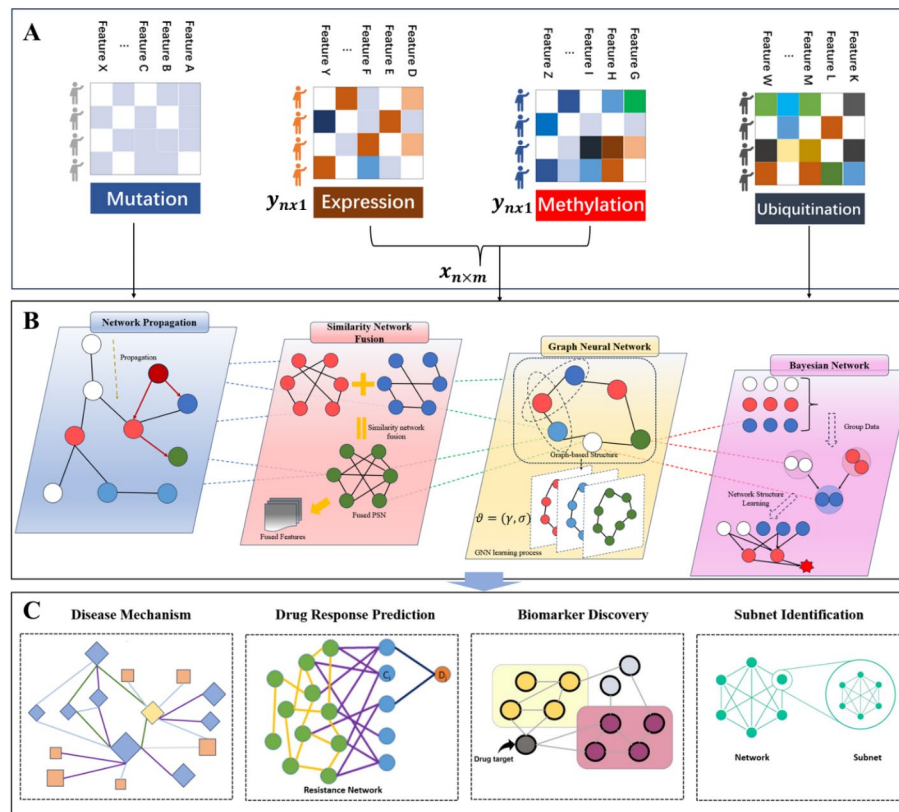


Fig. 2 Analysis and application of multi-omics data in biomedical research. **(A)** The collection of data matrices representing various types of omics data (mutations, gene expression, DNA methylation and Ubiquitination). Each matrix is organized to reflect the quantitative measurements of biological variables across multiple samples, illustrating the initial structured form of multi-omics data that serves as the input for further computational analyses; **(B)** The advanced computational strategies used to analyze the multi-omics data. Techniques emphasizing their role in integrating diverse omics data of biological inference. These methods facilitate the extraction of meaningful patterns and relationships from complex, high-dimensional datasets; **(C)** The practical applications of analyzed multi-omics data, focusing on predicting biomarkers and drug responses. It shows how integrated data analysis can lead to actionable insights in personalized medicine, such as identifying potential biomarkers for disease diagnosis or prognosis and predicting patient-specific drug responses

methods aim to disseminate omics data across a network, akin to the flow of information or the spread of a signal through a network of nodes and edges. These nodes typically represent biological entities such as genes, proteins, or metabolites, while the edges signify the interactions or functional associations between them. In the process of integrative analysis, network propagation methods transform omics data into score vectors, which are then mapped onto the nodes of the network [72, 73].

The diffusion of information across a network quantifies the proximity between entities in a global way and allows for the incorporation of diverse omics data layers, endowing the capacity to capture the emergent properties of biological systems. For example, a genotype can influence a phenotype not merely through a direct connection but via a cascade of intermediary molecular events, a nuance that network propagation methods are well-equipped to model [74].

Common network propagation methods include random walks, heat diffusion processes, or label propagation, which iteratively estimate the proximity between features associated with different data types by considering all possible paths [74, 75]. This proximity often reveals biologically relevant entities and interactions that are informed by

the initial input data and the underlying network structure. Although the fundamental concepts underlying various network propagation algorithms are similar, there are distinctions in the implementation and applications (Table 3). These can be applied to predict drug responses by propagating information across drug-target interaction networks. By incorporating multi-omics data, these methods allow for the identification of potential drug candidates and biomarkers associated with treatment outcomes.

Additionally, each of these methods has distinct algorithms, applications, and limitations, making them suitable for different types of problems. Random Walks are versatile but computationally intense and sensitive to parameters. Heat Diffusion Processes are excellent for continuous data and network analysis but can be resource-intensive and suffer from over-smoothing issues. Label Propagation is efficient and easy to implement but may be less reliable in noisy or poorly structured networks.

Network similarity-based methods

Network similarity-based methods are derived from collaborative-filtering algorithms in recommender systems, play a pivotal role in the integration and analysis of multi-omics data, particularly in addressing complex biological questions such as cancer subtype detection [76, 77]. Typically, these approaches first build similarity networks for each

Table 3 The comparison of common network propagation algorithms

Method	Computational Efficiency	Representative Algorithms	Limitations	Main Applications
Random Walks	Random walks often entail a higher computational cost due to the need for numerous iterations to reach equilibrium, especially in large and complex networks.	Random Walk with Restart (RWR) [62], PRINCE [63]	(1) Results can be sensitive to parameter choices, such as restart probability. (2) May over-emphasize local network structure at the expense of global features.	(1) Disease gene prioritization. (2) Protein function prediction. (3) Drug-target interaction prediction. (4) Network-based stratification of cancer subtypes.
Heat Diffusion Processes	While ensuring thorough propagation, heat diffusion processes can become computationally intensive as they require updates across all nodes uniformly until reaching a stable state. But heat diffusion processes typically converge faster and can theoretically guarantee reaching a stable uniform state.	HotNet [64], TieDIE [65]	Results are sensitive to initial heat allocation, noise and erroneous connections in the network.	(1) Identification of cancer driver genes. (2) Protein complex detection. (3) Gene module discovery. (4) Network-based disease gene prediction.
Label Propagation	Label propagation boasts lower computational complexity and efficiency, as it primarily uses local neighborhood information to rapidly converge in a few iterations.	GeneMANIA [66]	(1) Convergence issues: May face convergence difficulties in certain network structures. (2) Network quality dependence: Heavily reliant on the quality and completeness of the underlying network. (3) Heterogeneous network challenges: May struggle with heterogeneous biological networks.	(1) Gene function prediction. (2) Protein-protein interaction prediction. (3) Disease-gene association.

omics data type, these networks are crucial for drug response prediction, as they allow researchers to identify patient subtypes that are more likely to respond to certain treatments. By comparing the similarity between drug profiles and disease profiles, these methods can predict which drugs may be effective for specific disease subtypes.

Similarity Network Fusion (SNF) stands as a seminal method in this field [77]. SNF integrates multiple data types to construct similarity networks through an iterative process, creating a comprehensive patient similarity network. It first acts as an individual network by constructing a sample-by-sample similarity matrix for each data type. Then SNF fuses different similarity matrices and networks using the non-linear method of message pass theory (KNN and graph diffusion) until they converge to a single network, which is then partitioned using spectral clustering [78]. Its advantage is that the weak connections (noise) disappear with iterations, whereas the strong connections are propagated till convergence [77]. This method not only integrates data from various omics levels including genomics, transcriptomics, and proteomics but also proves valuable in disease subtype classification and drug response prediction. Recent years have seen several improvements to SNF, such as the Joint-SNF proposed by Li et al. [79], which significantly enhances strong similarities and weakens some spurious associations between samples while reducing the noise by extracting the joint structure between omics data types.

Network Alignment is another crucial method in the network similarity-based approach. This technique aims to identify corresponding nodes or subgraphs between different omics networks, facilitating the transfer of information across omics layers. For instance, woo et al. proposed MONACO, a method for highly accurate pairwise and multiple protein-protein interaction networks [80]. In the application of multi-omics integration context, network alignment can help identify conserved modules or pathways across different omics layers, providing insights into cross-omics relationships.

Multilayer Network Integration methods focus on representing multi-omics data as interconnected layers of networks, where each layer corresponds to a different omics type. These methods preserve the unique properties of each omics layer while capturing inter-layer relationships. For example, Didier et al. introduced MuxViz, a framework for multilayer network analysis and visualization, which has been applied to multi-omics data integration [81]. More recently, Dimitrova et al. developed MONET (Multi-Omics NETWORKS), a method that constructs a multilayer network from multi-omics data and applies community detection algorithms to identify cross-omics modules [78].

Graph neural network-based methods

Graph neural networks (GNNs) have emerged as a powerful tool for analyzing and integrating multi-omics data in the context of biological networks. GNNs extend the capabilities of traditional neural networks to graph-structured data, making them particularly well-suited for capturing complex relationships in biological systems. By analyzing the relationships between drugs and diseases or patient characteristics, these systems can recommend drugs that are likely to be effective based on similarities in molecular profiles or treatment histories. The fundamental principle of GNNs in multi-omics integration is to represent different omics layers as interconnected graphs, where nodes typically represent biological entities (e.g., genes, proteins, metabolites) and edges represent relationships between these entities. GNNs then learn to aggregate information

from neighboring nodes, effectively propagating signals across the graph and capturing both local and global network structures. One of the pioneering works in this area is the Graph Convolutional Network (GCN) proposed by Kipf and Welling [82]. While not specifically designed for multi-omics data, GCNs laid the groundwork for many subsequent developments in the field. Building upon this foundation, Zitnik et al. introduced Decagon, a GNN-based method for predicting polypharmacy side effects by integrating protein-protein interaction networks with drug-target interactions [83]. Decagon demonstrated superior performance compared to traditional machine learning methods, highlighting the potential of GNNs in pharmaceutical applications [83].

More recently, Gao et al. proposed MGNN, a multi-view graph neural network for integrating multi-omics data in cancer prognosis prediction. MGNN constructs separate graphs for different omics data types and uses a novel mechanism to fuse information across these graphs, achieving state-of-the-art performance in survival prediction tasks [84]. In the realm of drug discovery, Peng et al. introduced MOFGCN, a GNN-based framework for integrating multi-omics data to predict drug responses. This method constructs a heterogeneous graph that incorporates gene expression, copy number variation and somatic mutation data, along with known drug-target interactions [85]. By leveraging the power of GNNs to capture complex patterns in this heterogeneous graph, MOFGCN demonstrates improved accuracy in predicting drug sensitivity across various cancer types [85]. One of the key advantages of GNN-based methods is their ability to naturally handle the hierarchical and modular nature of biological systems. For instance, Yan et al. developed a hierarchically designed deep learning algorithm to integrate multi-omics data at different biological scales, from molecular interactions to cellular pathways [86]. This approach employed a graph neural network to extract gene features and used learnable linear layer to aggregate gene-level features into pathway-level features. Their results suggested that multiple omics data achieves superior performance than on single omics data and graph neural networks demonstrate significant potential in processing multi-omics data and regulatory networks.

Despite their successes, GNN-based methods for multi-omics integration face several challenges. One key issue is the interpretability of the learned models. To address this, Schulte-Sasse et al. introduced a GNN framework EMOGI [87]. EMOGI employs graph convolutional networks to integrate multi-omics data and protein-protein interaction networks for cancer gene prediction, while achieving model interpretability through layer-wise relevance propagation. This approach enables the identification of key features driving each gene's classification, stratification of genes based on their driving factors, detection of important modules within the network, and analysis of the contributions of various molecular alterations to predictions [86, 87].

In addition, the field of GNN-based multi-omics integration is rapidly evolving. Emerging trends include the development of self-supervised learning approaches to leverage unlabeled multi-omics data [88], the integration of GNNs with other deep learning architectures such as transformers for improved performance [89], and the adaptation of GNN-based methods to single-cell multi-omics data [90].

Network-based inference methods

Causality-based network inference methods have emerged as powerful methods for unraveling complex biological relationships and predicting drug responses. These

methods aim to go beyond mere correlation, striving to infer causal relationships between molecular entities across different omics layers and reconstruct the underlying biological networks that drive cellular processes and drug responses [91, 92].

Probabilistic Graphical Models (PGMs) serve as the theoretical foundation for this field, combining graph theory and probability theory to provide a flexible framework for modeling complex systems. PGMs represent probabilistic relationships among a set of variables using graph structures, where nodes typically represent molecular entities (e.g., genes, proteins, metabolites) and edges denote probabilistic dependencies between these entities [93]. The versatility of PGMs lies in their ability to represent both directed (as in Bayesian Networks) and undirected (as in Markov Random Fields) relationships [94]. Directed models require pre-defined directionality or capture conditional (in)dependencies to assert an influence on features, undirected models are limited to inference tasks because they fail to capture the influence of nodes on neighbouring nodes.

Bayesian Networks (BNs) is a subclass of PGMs that can add prior knowledge (such as distributions of the input) or encode the domain knowledge to improve the learning or training phases. In BNs, edge directions are often interpreted as causal relationships or information flow, allowing for efficient representation and well capture linear, non-linear, combinatorial, and stochastic types of relationships among variables across multiple levels of biological organization [95, 96]. Dynamic Bayesian methods can unfold dependencies over time and thus are extensively applied in the research of temporal processes involving various hierarchical levels of omics due to their capacity for modeling time-series data. For instance, Kourou et al. employed DBNs to model time series gene expression data for classification, revealed differences and similarities between the regulatory networks of cancer and normal cells [96]. In addition, Bayesian Networks are particularly useful in predicting drug responses by modeling the underlying biological processes that drive drug efficacy. For instance, using Bayesian networks to infer gene-drug interactions allows for the identification of key regulatory genes that influence drug sensitivity, facilitating personalized treatment strategies.

Closely related to BNs are Causal Inference Networks, which place a more explicit focus on inferring causal relationships. While BNs can represent probabilistic dependencies, CINs introduce concepts of intervention and counterfactual reasoning to distinguish true causal relationships from mere correlations. The integration of BN and CIN principles can infer temporal causal relationships across multiple molecular layers, capturing not only static dependencies but also dynamic causal effects over time.

Furthermore, the integration and innovative application of these methods are pushing the boundaries of multi-omics data analysis. A notable development is the integration of these traditional methods with modern statistical techniques. NetMIM addresses the challenges of missing multi-omics data and interpretability in feature selection by incorporating gene pathway information and employing the Dirac spike-and-slab variable selection method, thereby enhancing the predictive accuracy of the model [97]. Gong et al. introduced Coblot [98], an innovative approach that combines integrates Bayesian probabilistic modeling, deep learning (via Variational Autoencoders), and concepts of transfer learning, can analyze multi-modality sequencing data, whether separately or integrated with single-modality data. This integration harnesses the strengths of multiple methodologies. Bayesian models provide interpretability, elucidating the data generation process. Deep learning methods, like VAEs, offer flexibility in handling

complex, high-dimensional data. Transfer learning concepts enable Cobolt to leverage joint-modality data insights to enhance single-modality data analysis. This combination creates a powerful, versatile tool for single-cell multi-omics data analysis [98]. Moreover, the integration of Structural Equation Models (SEMs) and BNs, which captures the underlying relationships between different omics platforms (e.g., RNA-seq, WGS) by introducing latent variables. This method is further integrated with the accelerated failure time model, providing a comprehensive framework that not only considers the biological interconnections among data but also facilitates the simultaneous handling of multiple genes. It provides a novel perspective for investigating causal relationships within complex biological systems [99].

Application of multi-omics integration in drug discovery based on biological network analysis

The key developments and main tasks in drug prediction

Drug prediction is a data-driven computational process of predicting how a drug will affect a sample of specific molecular profile. It can involve predicting how a drug will respond in a new cell line, or how a drug will interact with a specific target. Traditional methods for predicting new drugs involve measuring various indices (i.e., IC₅₀, EC₅₀) between drugs and their targets through biological experiments. However, these methods are time-consuming and resource-intensive. As research of biological systems has deepened, the emergence of network-based perspectives in pharmacology has revolutionized our understanding of the intricate relationships between drugs, targets, and diseases. This new paradigm, often referred to as the “multiple drugs → multiple targets → multiple diseases” model, acknowledges the fact that most drugs interact with multiple targets, and that these interactions can have far-reaching effects across diverse biological pathways and systems. The integration of multi-omics data within the framework of biological network analysis has emerged as a powerful approach for drug prediction, offering unprecedented insights into complex biological systems and their responses to pharmaceutical interventions. This combinations the molecular profiling capabilities of multi-omics with the systemic perspective provided by biological network analysis, enabling more accurate and mechanistically informed predictions of drug efficacy, potential repurposing opportunities and precision medicine [100, 101]. The journey in drug discovery begins with the collection of diverse omics data types, such as genomics, transcriptomics, and proteomics. These data are then analyzed using advanced computational strategies. These methods allow for the integration and analysis of complex, high-dimensional datasets within the context of biological networks. Finally, the integrated analysis leads to practical applications in drug discovery, such as identifying novel drug targets, predicting drug responses, and designing personalized treatment strategies (Fig. 3).

Several key factors underpin the rationale for applying multi-omics integration and network analysis in drug prediction. First, this holistic approach transcends the limitations of traditional single-target or single-pathway analyses, recognizing that drugs typically engage with multiple targets and perturb diverse biological processes simultaneously. Second, by integrating multi-omics data into the framework of biological networks, these methods provide a more physiologically relevant and interconnected context for drug response predictions. Third, the synergistic integration of multiple

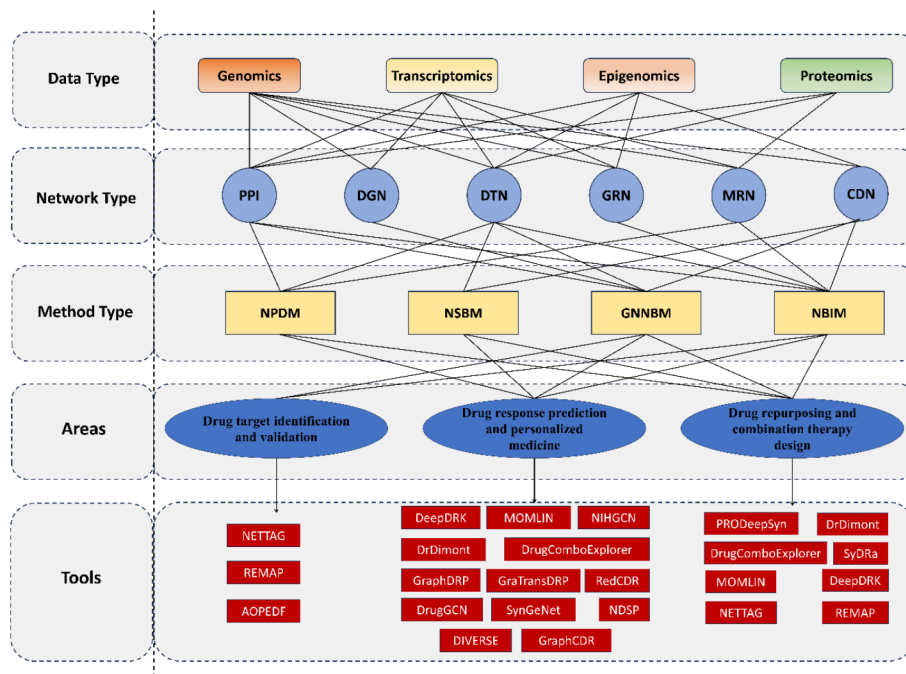


Fig. 3 Overview of network-based multi-omics integration methods/tools in drug discovery. PPI, Protein-Protein Interaction Network; DGN, Disease-Gene Network; DTN, Drug-Target Network; GRN, Gene Regulatory Network; MRN, Metabolic Regulatory Network; CDN, Cell Line-Drug Network; NPDM, Network Propagation/Diffusion Method; NSBM, Network Similarity-Based Method; GNNBM, Graph Neural Network-Based Method; NBIM, Network-Based Inference Method

omics layers enhances signal detection and mechanistic insight, facilitating the development of targeted, personalized therapeutic strategies that account for individual molecular variability.

Through our comprehensive analysis of literature published between 2015 and 2024, we identified and evaluated 21 methods/tools that employed network-based multi-omics integration in drug discovery from a total of 265 candidates. We categorized three primary application domains: drug target identification and validation (19%), drug response prediction and personalized medicine (58%), and drug repurposing and combination therapy design (23%). Each of these categories employs distinct methodologies while sharing the fundamental principle of utilizing integrated multi-omics data within a network context to enhance predictive power and biological interpretability (Table 4).

Drug target identification and validation

This category focuses on leveraging multi-omics data and biological networks to identify and validate potential drug targets. They typically integrate genomics, transcriptomics, and proteomics data with protein-protein interaction networks to capture complex molecular relationships. REMAP [106], through the integration of diverse drug-target interaction data, constructs a comprehensive drug-protein interaction network. Building upon this foundation, it innovatively applies matrix factorization and collaborative filtering techniques to accurately predict novel interactions from a vast array of candidate drug-protein pairs, substantially enhancing both the efficiency and precision of drug target discovery. AOPEDF [110] advances this concept further by integrating 15 distinct biological networks, including protein-protein interaction networks and disease-gene

Table 4 Summary of network-based multi-omics integration tools for drug prediction applications

Model/ Tool Name	Primary Focus	Input Data (Data Type)	Network Type	Method Type	Published Time
SynGeNet [88]	Drug combination prediction	Gene expression, mutation, copy number variation, RPPA expression, metabolomics data from TCGA and GDSC	Protein-protein interaction network	Network centrality analysis, Connectivity mapping, Belief propagation	2019
- [89]	Drug response prediction, Survival prediction	Gene expression, copy number variation, gene mutations, RPPA expression, methylation, miRNA data from TCGA and GDSC	-	Deep neural networks, NCA feature selection, Unsupervised clustering	2021
NETTAG [90]	Drug repurposing, Target identification	GWAS data, multi-omics data (gene expression, proteomics, etc.), drug-target networks, patient EHR data	Protein-protein interaction network	Network propagation, Bayesian inference, Network proximity	2022
- [91]	Biomarker identification	GWAS, gene expression, DNA methylation data	Protein-protein interaction network	Network propagation, Functional enrichment analysis	2020
REMAP [92]	Drug target identification	Drug-target interaction data, Chemical structures, Protein sequences	Drug-target interaction network	Matrix factorization, Collaborative filtering	2019
NDSP [93]	Drug response prediction	Gene expression, Copy number aberration, DNA methylation	Sample similarity networks	Deep learning, Similarity network fusion, Sparse PCA	2023
DrugGCN [94]	Drug response prediction	Gene expression, PPI network	PPI network	Graph convolutional network	2021
GraTrans-DRP [95]	Drug response prediction	Drug molecular graphs, Gene expression, Mutation, DNA methylation	Molecular graph	Graph transformer, Convolutional neural network, KernelPCA	2022
Cancer-OmicsNet [96]	Drug response prediction	PPI network, Differential gene expression, Disease-gene association, Kinase inhibitor profiling	Cancer-specific PPI network	Graph neural network, Attention mechanism	2022
DeepDRK [97]	Drug repurposing	Gene expression, Somatic mutations, Copy number variations, Drug chemical properties	Multi-omics similarity network	Deep learning, Kernel-based integration, Graph convolutions	2022
AOPEDF [98]	Drug target identification, Drug repurposing	Drug chemical structures, drug targets, side effects, gene expression data, monotherapy data	Heterogeneous network of drugs, targets, and diseases	Network embedding, Deep forest	2020
DrugComboExplorer [99]	Drug combination prediction	DNA-seq, gene copy number, DNA methylation, RNA-seq data	Driver signaling networks	Network propagation, Bayesian factor regression	2019
MOMLIN [100]	Drug response prediction	Clinical data, mutation data, gene expression, tumor microenvironment cells, molecular pathways	Driver signaling networks	Sparse correlation analysis, Class-specific feature selection	2024
DrDimont [101]	Drug response prediction	Gene expression, proteomics, phosphosite, metabolomics data	Heterogeneous multi-layer molecular networks	Network integration, Differential network analysis	2022
NIHGCN [102]	Drug response prediction	Gene expression, DNA mutation, tumor microenvironment, pathway activity data	Heterogeneous bipartite network of drugs and cell lines	Graph Convolutional Network, Neighborhood Interaction	2022

Table 4 (continued)

Model/ Tool Name	Primary Focus	Input Data (Data Type)	Network Type	Method Type	Pub- lished Time
SyDRa [103]	Drug combina- tion therapy design	Drug chemical structure, drug-target interactions, protein-protein interactions, drug induced gene expression profiles	PPI network, drug-target interaction network	Network-based inference methods, Machine learning (Random Forest)	2017
DIVERSE [104]	Drug response prediction	Drug similarity, gene expres- sion, protein-protein interac- tion, drug-target interaction, cell line-drug interaction	PPI network, drug-target interaction network	Bayesian matrix factorization	2021
PRODeep- Syn [105]	Drug combina- tion therapy design	Gene expression data, gene mutation data, drug molecular fingerprints and descriptors	PPI network	Graph neural net- work (GCN), Deep neural network	2022
GraphDRP [106]	Drug response prediction	Drug molecular graphs, ge- nomic aberration of cell lines	Drug molecu- lar graph	Graph convolutional networks	2020
RedCDR [107]	Drug response prediction	Cell line multi-omics data, drug information	Cell line-drug, cell line-cell line, drug-drug heterogeneous graphs	Low-rank global attention for graph representation, Bilinear predictor	2022
GraphCDR [108]	Drug response prediction	Cell line multi-omics data, drug molecular structures, cell line- drug response data	Cell line-drug heterogeneous graph	Graph neural network, Contrastive learning	2022

association networks, to construct an information-rich heterogeneous network. Within this framework, AOPEDF ingeniously employs advanced machine learning algorithms such as network embedding and cascade forests to predict and prioritize drug-target interactions from multiple dimensions. Moreover, it leverages network topology information to elucidate potential mechanisms of drug action. These methods share a common theme of leveraging network structures to capture complex relationships between molecular entities, enabling more comprehensive target identification than traditional single-omics approaches.

Drug response prediction and personalized medicine

This category aims to predict individual patient responses to drugs, supporting personalized treatment decisions. These approaches often integrate patient-specific multi-omics data with drug information and biological networks, considering inter-individual differences and complex molecular interactions. For example, MOMLIN [112] integrates clinical data, mutation data, gene expression, tumor microenvironment cells, and molecular pathway information to construct driver signaling networks. MOMLIN uses sparse correlation analysis to identify response-specific sparse components and develops drug response predictors using latent components. In predicting paclitaxel response, MOMLIN not only considered known microtubule dynamics-related genes but also discovered the importance of immune-related pathways, demonstrating its ability to capture complex biological relationships that traditional single-omics methods might overlook. CancerOmicsNet [109] employs a graph neural network with attention propagation mechanisms to predict the therapeutic effects of kinase inhibitors across various tumors. It constructs cancer-specific networks from PPI, differential gene expression, and drug inhibition data. By integrating multiple heterogeneous data types, CancerOmicsNet achieves high prediction accuracy and generalizability, especially in predicting drug

responses for new cell lines and new drugs. NIHGCN [114] predicts drug response by constructing a heterogeneous network of drugs and cell lines. It incorporates both node-wise and element-wise interactions, considering the heterogeneity of cell lines and drugs. NIHGCN's parallel graph convolution and neighborhood interaction layers allow it to capture complex patterns in multi-omics data and regulatory networks, leading to improved prediction performance. These methods demonstrate the power of network-based approaches in integrating diverse omics data types and capturing complex biological interactions for more accurate drug response predictions.

Drug repurposing and combination therapy design

This category is dedicated to discovering new therapeutic uses for existing pharmaceuticals and developing synergistic drug combination strategies. The goal is to maximize the use of current pharmacological resources while reducing the costs and risks linked with developing new drugs from scratch. To achieve this, the methods employed typically integrate cheminformatics and systems biology, considering interactions between drugs and the effects of targeting multiple biological pathways. DrDimont [113] pioneered a novel drug repurposing methodology predicated on differential network analysis. This approach initially integrates multi-omics data to construct disease-specific molecular networks, subsequently employing differential analysis to predict drug action disparities across diverse disease contexts, thereby uncovering new therapeutic indications. This network-level differential analysis method overcomes the limitations of traditional approaches, capturing the dynamic changes and context-specificity of drug actions, thus demonstrating unique advantages in drug repurposing. PRODeepSyn [116], in contrast, focuses on the challenge of drug combination therapy, proposing an innovative graph neural network framework. This method utilizes protein-protein interaction networks as a scaffold, integrating multi-omics data including gene expression and mutation profiles. It incorporates an ingenious drug embedding method based on fingerprints and descriptors, enabling accurate characterization of both chemical properties and biological activities of drugs. Building upon this foundation, PRODeepSyn employs an end-to-end deep learning model to predict synergistic effects of drug combinations, achieving remarkable performance. The methodologies exemplified by DrDimont and PRODeepSyn represent significant advancements in leveraging the complexity of biological systems for drug discovery.

Representative case studies of network-based multi-omics integration in drug discovery

To highlight the practical applications and translational potential of network-based multi-omics integration, we examine two cutting-edge case studies that represent different but complementary approach strategies in drug discovery.

Multi-modal network integration for precision cancer treatment

The SynGeNet framework demonstrates how network analysis can be used to integrate multi-omics data to improve precision oncology, particularly in predicting the efficacy of drug combinations for various genomic subtypes of melanoma. The approach innovatively utilizes network analysis to create subtype-specific protein sub-networks that integrate genomic variants and transcriptomic features of matched samples. It incorporates protein-protein interaction (PPI) networks and utilizes a belief propagation method to

map network flows of key driver or “root” genes. Network edges are weighted according to biological evidence and expression levels, providing a comprehensive characterization of biological systems.

Drug combination prediction in SynGeNet follows a two-step process: the first step is connectivity mapping to identify gene expression signature reversals; the second step is network centrality analysis to locate key nodes in the network. This dual approach integrates local and global network features to enhance the robustness of the prediction. The framework was rigorously validated by multiple methods, including strong cross-validation performance across all genomic subtypes, confirming the robustness of the network to individual gene variants.

The most compelling validation came from experimental testing of the highest predictive combination of BRAF-mutant melanomas. The combination of vemurafenib and tretinoin showed a significant synergistic effect with CIs of $ED_{50}=0.385$ and $ED_{75}=0.308$. In vitro studies showed enhanced cytotoxicity and apoptosis induction, while an in vivo xenograft model demonstrated a significant shrinkage of the tumors ($p=0.029$). RNA-seq analyses further validated the molecular mechanisms underlying the predictions, providing valuable mechanistic insights into the effectiveness of the combination. provided valuable mechanistic insights.

Deep learning enhanced heterogeneous network integration

AOPEDF is a novel deep learning approach for systematic drug-target interaction prediction by integrating different biological networks. AOPEDF is unique in that it integrates 15 different biological networks, including drug-centered networks (e.g., drug-drug interactions, drug-disease associations, and various similarity measures) and protein-centered networks (e.g., protein-protein interactions, disease associations, and functional networks). AOPEDF uniquely integrates 15 different biological networks, including both drug-centric networks (such as drug-drug interactions, drug-disease associations, and various similarity measures) and protein-centric networks (such as protein-protein interactions, disease associations, and functional networks). AOPEDF introduces several technological innovations:

Network embedding Utilizing arbitrary-order proximity preservation to capture complex network topologies.

Model architecture Applying deep forest technique with fewer hyperparameters than traditional neural networks.

Prediction mechanism Generating interpretable decision rules through a tree-based model.

The method demonstrates robust performance in multiple validation scenarios. In the internal validation with 5-fold cross-validation, AOPEDF achieves excellent results with $AUROC=0.985$ and $AUPR=0.985$. More impressively, it maintains strong performance on the external validation set, achieving $AUROC=0.868$ and $AUPR=0.869$ on the DrugCentral dataset, $AUROC=0.868$ and $AUPR=0.869$ on the ChEMBL dataset, and $AUROC=0.868$ and $AUPR=0.869$ on the DrugCentral dataset. on the DrugCentral

dataset and AUROC = 0.768 and AUPR = 0.764 on the ChEMBL dataset, outperforming current state-of-the-art methods.

A particularly notable application of AOPEDF is the identification of novel drug-target interactions for substance abuse disorders. The method successfully predicted several validated interactions, including the interaction between aripiprazole and HRH3, highlighting its potential to generate clinically relevant and mechanistically sound predictions.

Performance comparison

In the rapidly evolving landscape of network-based multi-omics integration for drug discovery, several key methodologies have emerged, each offering unique advantages and facing distinct challenges. We will conduct a thorough analysis of the four principal methods previously mentioned: network propagation/diffusion methods, network similarity-based methods, graph neural network-based methods, and network-based inference methods. We will examine their underlying principles, strengths, limitations, and applications in various drug discovery contexts by using some methods in Table 4.

Network propagation/diffusion methods simulate the flow of information across complex molecular networks, typically starting from a set of seed nodes and diffusing their influence to neighboring nodes based on network topology [74]. These approaches excel in uncovering hidden associations not apparent from direct interactions and are capable of prioritizing candidate genes or drug targets in a biologically meaningful context. A prime example of this approach is the DrugComboExplorer [111] tool, which employs a non-parametric, bootstrapping-based simulated annealing algorithm to identify robust dysregulated signaling networks from multi-omics data. This tool then uses these networks to predict synergistic drug combinations, demonstrating the power of network propagation in integrating complex data types for practical drug discovery applications. The strength of network propagation lies in its ability to uncover latent connections and leverage the topological properties of biological networks. This is particularly valuable when dealing with incomplete or noisy data, as the method can infer missing links and reduce the impact of false positives. For instance, in the context of rare diseases or limited sample sizes, network propagation can amplify weak signals by considering the broader network context [119]. However, these methods are not without limitations. They can be computationally intensive for large-scale networks, potentially limiting their applicability to genome-wide studies, such as DrugComboExplorer may mitigate this by using a Bayesian factor regression model to decompose drug treatment signatures, balancing computational efficiency with predictive power. Additionally, the results in network propagation methods can be sensitive to the quality and completeness of the underlying network structure. Despite these challenges, network propagation methods remain a cornerstone of multi-omics integration, particularly in exploratory analysis and hypothesis generation.

Network similarity-based methods construct and analyze similarity networks across different omics layers, often using techniques like Similarity Network Fusion or kernel-based approaches to integrate diverse data types [77]. These methods are particularly effective in scenarios where the relationships between different omics layers are not well understood or are expected to be highly complex, such as patient stratification and disease subtype identification, capable of capturing global similarities across multiple omics

layers. They are often more interpretable than black-box machine learning approaches. However, they may overlook local, fine-grained interactions crucial in certain biological contexts and can be sensitive to the choice of similarity metric [120]. The challenges of network similarity-based methods may struggle with high-dimensional data, potentially leading to the “curse of dimensionality” where the number of features far exceeds the number of samples. By using Kernel PCA to reduce the dimensionality of gene expression data and constructing similarity networks, NDSP [107] addresses the curse of dimensionality while preserving important biological information. This approach allows for the adaptive integration of multi-omics data, taking into account the importance of different omics in drug response prediction. However, similarity-based methods can be sensitive to the choice of similarity metric and may struggle with interpretability, especially when dealing with complex biological systems.

Graph Neural Network (GNN)-based methods leverage the power of deep learning on graph-structured data, learning complex, non-linear relationships in multi-omics networks through message passing between nodes [121]. These methods are highly effective in handling large-scale, heterogeneous biological networks and can capture both local and global network features. CancerOmicsNet [109] exemplifies the potential of GNN-based approaches in drug response prediction. By constructing cancer-specific networks and employing graph neural networks with sophisticated attention mechanisms, CancerOmicsNet achieves high prediction accuracy for the therapeutic effects of kinase inhibitors across various tumors. Its ability to generalize well to unseen data demonstrates the power of GNNs in capturing complex biological interactions.

Network-based inference methods, including Bayesian networks and causal inference approaches, are powerful tools for uncovering causal relationships and temporal dynamics in biological systems [92]. These methods excel at handling uncertainty and incorporating prior knowledge, making them particularly suitable for mechanistic studies and hypothesis generation. The DIVERSE [115] framework exemplifies the power of network-based inference methods. By using Bayesian matrix factorization to integrate multiple data types, DIVERSE enables precise drug response prediction. This approach demonstrates how network-based inference methods can effectively handle the uncertainty inherent in biological data while incorporating diverse data types. Moreover, the strength of these methods lies in their ability to model complex, probabilistic relationships and to incorporate prior biological knowledge. This makes them particularly valuable in scenarios where the underlying biological mechanisms are partially known but incomplete. For instance, in studying drug-induced gene expression changes, these methods can help infer the causal chain of events leading from drug administration to observed phenotypic changes. However, network-based inference methods can be computationally expensive for large networks and may struggle with cyclic relationships common in biological systems. Additionally, the quality of inferences is heavily dependent on the accuracy of the prior knowledge incorporated into the model.

While each of these methodologies offers unique strengths, the future of multi-omics integration lies in synergistic approaches that combine multiple methods to leverage their complementary strengths. For example, NETTAG [104] exemplifies this trend by integrating aspects of network propagation, Bayesian inference, and network proximity analysis. It uses a comprehensive framework that combines a Bayesian model selection method to infer Alzheimer’s disease risk genes, belief propagation to construct protein

subnetworks, and a network proximity approach to prioritize repurposable drugs. This hybrid approach allows NETTAG to capture both the genetic risk factors of Alzheimer's disease (through Bayesian inference and network propagation) and the potential therapeutic interventions (through network proximity analysis), resulting in improved drug repurposing predictions for Alzheimer's disease. The trend towards hybrid approaches is likely to continue, with future models potentially incorporating aspects of all four method categories discussed here. For instance, one could envision a model that uses network propagation to identify key modules in a biological network, similarity-based methods to integrate diverse omics data types, graph neural networks to learn representations of these integrated data, and network-based inference methods to uncover causal relationships within the identified modules.

The choice of method often depends on the specific drug discovery task, the types and scale of available omics data, and the desired balance between predictive power and biological interpretability. As the field continues to evolve, we anticipate the development of increasingly sophisticated models that can fully leverage the strengths of multiple approaches, pushing the boundaries of what's possible in multi-omics data integration and its applications in drug discovery. These advancements will undoubtedly play a crucial role in unraveling complex disease mechanisms and developing more effective, personalized therapeutic strategies.

Discussion

Network-based multi-omics integration: transforming drug discovery

The integration of multi-omics data using network-based approaches has emerged as a cornerstone in the development of systems biology and precision medicine [122]. Unlike traditional single-omics or linear methods, network-based integration leverages the interconnected nature of molecular systems, enabling the modeling of complex biological processes that underlie diseases. By integrating data such as genomics, transcriptomics, proteomics, and epigenomics, network models have revealed novel insights into molecular mechanisms, therapeutic targets, and biomarkers while providing a robust foundation for drug discovery [122–124].

This review categorized these methods into four main types: network propagation/diffusion, similarity-based approaches, graph neural networks (GNNs), and network inference models. Each approach has demonstrated distinct advantages in specific applications. For example, propagation methods uncover latent relationships in biological data, similarity-based methods excel in clustering and stratifying patient subtypes, GNNs provide a powerful framework for learning hierarchical and modular relationships, and network inference models capture probabilistic dependencies and causality.

The impact of these methods is evident in their ability to predict drug responses, identify potential therapeutic targets, and repurpose existing drugs. However, the growing complexity of multi-omics datasets necessitates continuous refinement of these methodologies to address their inherent challenges, which are elaborated below.

Tackling data heterogeneity and complexity

The inherent heterogeneity and complexity of multi-omics data remain significant obstacles in the field of network-based integration. Multi-omics datasets encompass diverse data types such as DNA mutations, gene expression profiles, protein-protein

interactions, and epigenetic modifications, each collected using different technologies and scales. These datasets are often noisy, sparse, and high-dimensional, posing challenges for meaningful integration and interpretation [110, 125].

One of the primary challenges is reconciling differences in data resolution and scale. Genomic data often involve static mutations or variations, while transcriptomics and proteomics capture dynamic changes in gene and protein expression under various conditions. Epigenomic data add another layer of complexity by reflecting environmental influences on gene regulation. The integration of these diverse data types requires advanced algorithms capable of capturing cross-level interactions without introducing significant biases or overfitting. Current methods, such as Similarity Network Fusion (SNF), attempt to address these challenges by constructing individual similarity networks for each data type and iteratively fusing them into a unified network [79]. While effective, these methods often struggle with scalability and are sensitive to the choice of similarity metrics.

The advent of single-cell and spatial omics technologies has further increased the complexity of integration [98, 126, 127]. Single-cell data reveal cellular heterogeneity by capturing the transcriptomic or proteomic profiles of individual cells, while spatial omics provide insights into the tissue architecture and cell-cell interactions within their native spatial contexts. These technologies highlight the need for scalable algorithms capable of handling millions of cells and preserving spatial relationships. Despite recent efforts, such as developing graph-based models tailored for single-cell data, the field still lacks generalizable frameworks to integrate these emerging data types into multi-omics networks.

Additionally, data noise and sparsity further complicate integration. Missing data, technical artifacts, and biological variability introduce uncertainties that can distort network construction and downstream analyses. Recent approaches, such as self-supervised learning and probabilistic inference, have shown promise in imputing missing values and denoising multi-omics datasets [88]. However, more robust methods are required to ensure accurate and biologically meaningful integration.

Balancing model complexity with biological interpretability

The growing sophistication of network-based methods, particularly those employing machine learning, has significantly improved predictive accuracy in multi-omics applications. However, this progress often comes at the cost of biological interpretability. Advanced models such as graph neural networks (GNNs) and deep learning-based architectures operate as “black boxes,” obscuring the underlying biological mechanisms that drive their predictions. This lack of interpretability undermines their utility in biomedical research, where understanding the rationale behind predictions is critical for experimental validation and clinical translation.

For example, GNNs, such as NIHGCN [114], have demonstrated remarkable performance in drug response prediction by learning hierarchical relationships in biological networks. However, their predictions often lack transparency, making it difficult to identify the specific genes, pathways, or network modules responsible for the observed outcomes. Similarly, network inference models, such as Bayesian networks, offer probabilistic insights into causal relationships but require extensive prior knowledge, which is not always available or reliable.

Recent advances in model interpretability provide a potential solution. Techniques such as layer-wise relevance propagation (LRP) and attention mechanisms in GNNs enable researchers to trace predictions back to specific network components, such as genes, proteins, or pathways [87, 109]. These methods enhance the interpretability of complex models, allowing researchers to generate hypotheses that are biologically testable. Moreover, hybrid approaches that combine interpretable probabilistic models with deep learning architectures hold promise for balancing predictive power with transparency.

To fully address the interpretability challenge, future efforts should prioritize the development of explainable AI frameworks tailored for multi-omics data. These frameworks should not only provide accurate predictions but also generate insights into the underlying biological processes, bridging the gap between computational models and experimental validation.

Standardization and benchmarking in method development

The lack of standardized evaluation frameworks remains a critical barrier to progress in network-based multi-omics integration. Current methods are often evaluated using disparate datasets, performance metrics, and experimental designs, making it difficult to compare their effectiveness and generalizability. This fragmentation not only hampers reproducibility but also creates challenges in translating computational methods into practical applications. A standardized benchmarking framework should include:

Diverse datasets

Incorporating datasets that represent a wide range of diseases, omics types, and experimental conditions to ensure the generalizability of methods.

Comprehensive metrics

Moving beyond traditional metrics such as accuracy and precision to include measures of interpretability, robustness, and computational efficiency.

Open-Access tools and resources

Providing public access to benchmarking datasets, pre-trained models, and evaluation pipelines to promote transparency and collaboration.

Emerging platforms

Offering resources for network-based multi-omics analysis. Expanding these platforms to include standardized pipelines, interactive visualization tools, and community-driven benchmarking initiatives will further accelerate the development and adoption of network-based methods.

Emerging trends and future directions

The field of network-based multi-omics integration is evolving rapidly, with several emerging trends poised to redefine its landscape. Among these, the integration of temporal and spatial dynamics into network models stands out as a key area of innovation. Current methods primarily focus on static networks, which fail to capture the dynamic nature of biological processes such as disease progression and treatment responses. Incorporating time-series

data into network models could provide valuable insights into the temporal dynamics of gene regulation, protein interactions, and metabolic pathways [128]. Similarly, spatial omics data offer opportunities to map tissue-specific molecular interactions, enabling a more comprehensive understanding of disease biology [129].

Another promising trend is the adoption of self-supervised learning and transfer learning [130, 131]. Self-supervised learning, which leverages unlabeled data to learn generalizable features, is particularly well-suited for multi-omics datasets, where labeled data are often scarce. Transfer learning, in which models trained on one dataset are adapted to another, has the potential to enhance the generalizability of network-based methods across diverse diseases and populations.

Finally, hybrid approaches that combine the strengths of multiple methodologies hold significant promise. For instance, integrating network propagation, GNNs, and probabilistic inference into a unified framework could overcome the limitations of individual methods while leveraging their complementary strengths. Such hybrid models could provide a holistic perspective on multi-omics integration, enabling more accurate predictions and deeper biological insights.

The ultimate success of these innovations will depend on their ability to translate computational findings into actionable outcomes in clinical and pharmaceutical settings. This requires interdisciplinary collaboration among computational biologists, clinicians, and industry stakeholders, as well as the development of user-friendly tools that facilitate the application of network-based methods in real-world scenarios.

Conclusion

Network-based multi-omics integration represents a powerful paradigm in drug discovery, enabling more accurate predictions of drug responses and identification of novel therapeutic targets. Our review has revealed the complementary strengths of different methodological approaches, from network propagation and similarity-based methods to advanced graph neural networks and network inference models. While significant progress has been made in improving prediction accuracy and biological interpretability, critical challenges persist in handling data heterogeneity, computational scalability, and standardization of evaluation frameworks [101, 132]. The integration of emerging single-cell and spatial omics technologies presents both opportunities and challenges, requiring novel computational strategies to handle increased data complexity while preserving biological meaning [126]. Furthermore, the field needs to address the balance between model sophistication and interpretability, particularly in the context of advanced machine learning approaches. Future developments should prioritize several key areas: the integration of temporal and spatial dynamics into network models, enhancement of model interpretability without sacrificing predictive power, development of scalable algorithms for massive-scale multi-omics datasets, and establishment of standardized evaluation frameworks. Success in addressing these challenges will be crucial for translating computational insights into actionable clinical applications, ultimately advancing the field of precision medicine and drug development.

Acknowledgements

The work was supported by the researchers themselves who performed this study.

Author contributions

WJ: Methodology, Analysis, Writing – original draft, Writing – review & editing; WY: Conceptualization, Methodology, Writing – review & editing; XT: Conceptualization, Methodology, Analysis, Writing – review & editing; YJB: Conceptualization, Methodology, Analysis, Writing – review & editing.

Data availability

No datasets were generated or analysed during the current study.

Declarations**Ethical approval**

The current study does not involve ethical issues.

Competing interests

The authors declare no competing interests.

Received: 25 September 2024 / Accepted: 17 March 2025

Published online: 28 March 2025

References

1. Stephens PJ, Tarpey PS, Davies H, Van Loo P, Greenman C, Wedge DC, et al. The landscape of cancer genes and mutational processes in breast cancer. *Nature*. 2012;486:400–4.
2. Cancer Genome Atlas Network. Comprehensive molecular portraits of human breast tumours. *Nature*. 2012;490:61–70.
3. Ng CKY, Bidard F-C, Piscuoglio S, Geyer FC, Lim RS, de Bruijn I, et al. Genetic heterogeneity in Therapy-Naive synchronous primary breast cancers and their metastases. *Clin Cancer Res Off J Am Assoc Cancer Res*. 2017;23:4402–15.
4. Hovelson DH, Udager AM, McDaniel AS, Grivas P, Palmboos P, Tamura S, et al. Targeted DNA and RNA sequencing of paired urothelial and squamous bladder cancers reveals discordant genomic and transcriptomic events and unique therapeutic implications. *Eur Urol*. 2018;74:741–53.
5. Cui J, Chen Y, Chou W-C, Sun L, Chen L, Suo J, et al. An integrated transcriptomic and computational analysis for biomarker identification in gastric cancer. *Nucleic Acids Res*. 2011;39:1197–207.
6. Rexach J, Lee H, Martinez-Agosto JA, Németh AH, Fogel BL. Clinical application of next-generation sequencing to the practice of neurology. *Lancet Neurol*. 2019;18:492–503.
7. Dagogo-Jack I, Shaw AT. Tumour heterogeneity and resistance to cancer therapies. *Nat Rev Clin Oncol*. 2018;15:81–94.
8. Massard C, Michiels S, Féré C, Le Deley M-C, Lacroix L, Hollebecque A, et al. High-Throughput genomics and clinical outcome in Hard-to-Treat advanced cancers: results of the MOSCATO 01 trial. *Cancer Discov*. 2017;7:586–95.
9. Rodon J, Soria J-C, Berger R, Miller WH, Rubin E, Kugel A, et al. Genomic and transcriptomic profiling expands precision cancer medicine: the WINHER trial. *Nat Med*. 2019;25:751–8.
10. Andrysik Z, Bender H, Galbraith MD, Espinosa JM. Multi-omics analysis reveals contextual tumor suppressive and oncogenic gene modules within the acute hypoxic response. *Nat Commun*. 2021;12:1375.
11. Huang Q, Chen H, Yin D, Wang J, Wang S, Yang F, et al. Multi-omics analysis reveals NNMT as a master metabolic regulator of metastasis in esophageal squamous cell carcinoma. *NPJ Precis Oncol*. 2024;8:24.
12. Liao Y, Wang J, Zou J, Liu Y, Liu Z, Huang Z. Multi-omics analysis reveals genomic, clinical and immunological features of SARS-CoV-2 virus target genes in pan-cancer. *Front Immunol*. 2023;14:1112704.
13. Tini G, Marchetti L, Priami C, Scott-Boyer M-P. Multi-omics integration—a comparison of unsupervised clustering methodologies. *Brief Bioinform*. 2019;20:1269–79.
14. Song M, Greenbaum J, Luttrell J, Zhou W, Wu C, Shen H, et al. A review of integrative imputation for Multi-Omics datasets. *Front Genet*. 2020;11:570255.
15. Bersanelli M, Mosca E, Remondini D, Giampieri E, Sala C, Castellani G, et al. Methods for the integration of multi-omics data: mathematical aspects. *BMC Bioinformatics*. 2016;17:S15.
16. Liu K, Chen X, Ren Y, Liu C, Lv T, Liu Y, et al. Multi-target-based polypharmacology prediction (mTPP): an approach using virtual screening and machine learning for multi-target drug discovery. *Chem Biol Interact*. 2022;368:110239.
17. Kumari B, Dholaniya PS. Parkinson's disease gene prioritising using an efficient and biologically appropriate network-based consensus strategy. *J Comput Sci*. 2022;65:101879.
18. Amanatidou AI, Dedoussis GV. Construction and analysis of protein-protein interaction network of non-alcoholic fatty liver disease. *Comput Biol Med*. 2021;131:104243.
19. Zhou G, Li S, Xia J. Network-Based approaches for Multi-omics integration. *Methods Mol Biol Clifton NJ*. 2020;2104:469–87.
20. Arnold M, the Alzheimer's Disease Metabolomics Consortium. Integrating multi-omics data for target and biomarker discovery. *Alzheimers Dement*. 2024;20:e086331.
21. Bagherian M, Sabeti E, Wang K, Sartor MA, Nikolovska-Coleska Z, Najarian K. Machine learning approaches and databases for prediction of drug-target interaction: A survey paper. *Brief Bioinform*. 2021;22:247–69.
22. Turanlı B, Karagoz K, Gulfidan G, Sinha R, Mardinoglu A, Arga KY. A Network-Based cancer drug discovery: from integrated Multi-Omics approaches to precision medicine. *Curr Pharm Des*. 2018;24:3778–90.
23. Momeni Z, Hassanzadeh E, Saniee Abadeh M, Bellazzi R. A survey on single and multi omics data mining methods in cancer data classification. *J Biomed Inf*. 2020;107:103466.
24. Zitnik M, Nguyen F, Wang B, Leskovec J, Goldenberg A, Hoffman MM. Machine learning for integrating data in biology and medicine: principles, practice, and opportunities. *Inf Fusion*. 2019;50:71–91.
25. Vastrik I, D'Eustachio P, Schmidt E, Gopinath G, Croft D, de Bono B, et al. Reactome: a knowledge base of biologic pathways and processes. *Genome Biol*. 2007;8:R39.
26. Lo Surdo P, Iannuccelli M, Contino S, Castagnoli L, Licata L, Cesareni G, et al. SIGNOR 3.0, the signaling network open resource 3.0: 2022 update. *Nucleic Acids Res*. 2023;51:D631–7.
27. Türei D, Korcsmáros T, Saez-Rodriguez J, OmniPath. Guidelines and gateway for literature-curated signaling pathway resources. *Nat Methods*. 2016;13:966–7.
28. Morris JH, Soman K, Akbas RE, Zhou X, Smith B, Meng EC, et al. The scalable precision medicine open knowledge engine (SPOKE): a massive knowledge graph of biomedical information. *Bioinformatics*. 2023;39:btad080.

29. Griffith M, Spies NC, Krysiak K, McMichael JF, Coffman AC, Danos AM, et al. CIViC is a community knowledgebase for expert crowdsourcing the clinical interpretation of variants in cancer. *Nat Genet.* 2017;49:170–4.
30. Gong L, Whirl-Carrillo M, Klein TE. PharmGKB, an integrated resource of Pharmacogenomic knowledge. *Curr Protoc.* 2021;1:e226.
31. Zhou Y, Zhang Y, Zhao D, Yu X, Shen X, Zhou Y, et al. TTD: therapeutic target database describing target druggability information. *Nucleic Acids Res.* 2024;52:D1465–77.
32. Szklarczyk D, Kirsch R, Koutrouli M, Nastou K, Mehryary F, Hachilif R, et al. The STRING database in 2023: protein–protein association networks and functional enrichment analyses for any sequenced genome of interest. *Nucleic Acids Res.* 2023;51:D638–46.
33. Kanehisa M, Furumichi M, Sato Y, Ishiguro-Watanabe M, Tanabe M. KEGG: integrating viruses and cellular organisms. *Nucleic Acids Res.* 2021;49:D545–51.
34. Weinstein JN, Collisson EA, Mills GB, Shaw KM, Ozenberger BA, Ellrott K, et al. The cancer genome atlas Pan-Cancer analysis project. *Nat Genet.* 2013;45:1113–20.
35. Ghandi M, Huang FW, Jané-Valbuena J, Kryukov GV, Lo CC, McDonald ER, et al. Next-generation characterization of the cancer cell line encyclopedia. *Nature.* 2019;569:503–8.
36. Edwards NJ, Oberti M, Thangudu RR, Cai S, McGarvey PB, Jacob S, et al. The CPTAC data portal: A resource for cancer proteomics research. *J Proteome Res.* 2015;14:2707–13.
37. Lonsdale J, Thomas J, Salvatore M, Phillips R, Lo E, Shad S, et al. *Nat Genet.* 2013;45:580–5. The Genotype-Tissue Expression (GTEx) project.
38. Piñero J, Bravo À, Queralt-Rosinach N, Gutiérrez-Sacristán A, Deu-Pons J, Centeno E, et al. DisGeNET: a comprehensive platform integrating information on human disease-associated genes and variants. *Nucleic Acids Res.* 2017;45:D833–9.
39. Hamosh A, Scott AF, Amberger JS, Bocchini CA, McKusick VA. Online Mendelian inheritance in man (OMIM), a knowledge-base of human genes and genetic disorders. *Nucleic Acids Res.* 2005;33 suppl1:D514–7.
40. Wishart DS, Knox C, Guo AC, Cheng D, Shrivastava S, Tzur D et al. DrugBank: a knowledgebase for drugs, drug actions and drug targets. *Nucleic Acids Res.* 2008;36 Database issue:D901–6.
41. Yang W, Soares J, Greninger P, Edelman EJ, Lightfoot H, Forbes S, et al. Genomics of drug sensitivity in cancer (GDSC): a resource for therapeutic biomarker discovery in cancer cells. *Nucleic Acids Res.* 2013;41:D955–61.
42. Kim S, Chen J, Cheng T, Gindulyte A, He J, He S, et al. PubChem in 2021: new data content and improved web interfaces. *Nucleic Acids Res.* 2021;49:D1388–95.
43. Koletti A, Terryn R, Stathias V, Chung C, Cooper DJ, Turner JP, et al. Data portal for the library of integrated Network-based cellular signatures (LINCS) program: integrated access to diverse large-scale cellular perturbation response data. *Nucleic Acids Res.* 2018;46:D558–66.
44. Zdrzil B, Felix E, Hunter F, Manners EJ, Blackshaw J, Corbett S, et al. The ChEMBL database in 2023: a drug discovery platform spanning multiple bioactivity data types and time periods. *Nucleic Acids Res.* 2024;52:D1180–92.
45. Hecker M, Lambeck S, Toepfer S, Van Someren E, Guthke R. Gene regulatory network inference: data integration in dynamic models—A review. *BioSystems.* 2009;96:86–103.
46. Feng K, Jiang H, Yin C, Sun H. Gene regulatory network inference based on causal discovery integrating with graph neural network. *Quant Biol.* 2023;11:434–50.
47. Willsey AJ, Sanders SJ, Li M, Dong S, Tebbenkamp AT, Muhle RA, et al. Coexpression networks implicate human midfetal deep cortical projection neurons in the pathogenesis of autism. *Cell.* 2013;155:997–1007.
48. Chen K, Wang Q, Li M, Guo H, Liu W, Wang F, et al. Single-cell RNA-seq reveals dynamic change in tumor microenvironment during pancreatic ductal adenocarcinoma malignant progression. *EBioMedicine.* 2021;66:103315.
49. Vinayagam A, Zirin J, Roesel C, Hu Y, Yilmazel B, Samsonova AA, et al. Integrating protein-protein interaction networks with phenotypes reveals signs of interactions. *Nat Methods.* 2014;11:94–9.
50. Cheng F, Desai RJ, Handy DE, Wang R, Schneeweiss S, Barabási A-L, et al. Network-based approach to prediction and population-based validation of in Silico drug repurposing. *Nat Commun.* 2018;9:2691.
51. Hao T, Peng W, Wang Q, Wang B, Sun J. Reconstruction and application of Protein–Protein interaction network. *Int J Mol Sci.* 2016;17:907.
52. Rothenburg S, Brennan G. Species-Specific host-Virus interactions: implications for viral host range and virulence. *Trends Microbiol.* 2020;28:46–56.
53. Basler G, Nikoloski Z, Larhlimi A, Barabási A-L, Liu Y-Y. Control of fluxes in metabolic networks. *Genome Res.* 2016;26:956–68.
54. Ogretmen B. Sphingolipid metabolism in cancer signalling and therapy. *Nat Rev Cancer.* 2018;18:33–50.
55. Valvezan AJ, Manning BD. Molecular logic of mTORC1 signalling as a metabolic rheostat. *Nat Metab.* 2019;1:321–33.
56. Campbell JE, Newgard CB. Mechanisms controlling pancreatic islet cell function in insulin secretion. *Nat Rev Mol Cell Biol.* 2021;22:142–58.
57. Hogg SJ, Beavis PA, Dawson MA, Johnstone RW. Targeting the epigenetic regulation of antitumour immunity. *Nat Rev Drug Discov.* 2020;19:776–800.
58. Rodriguez RM, Saiz ML, Suarez-Álvarez B, López-Larrea C. Epigenetic networks driving T cell identity and plasticity during Immunosenescence. *Trends Genet TIG.* 2022;38:120–3.
59. Turner AP, Caves LSD, Stepney S, Tyrrell AM, Lones MA. Artificial epigenetic networks: automatic decomposition of dynamical control tasks using topological Self-Modification. *IEEE Trans Neural Netw Learn Syst.* 2017;28:218–30.
60. Vinayagam A, Gibson TE, Lee H-J, Yilmazel B, Roesel C, Hu Y, et al. Controllability analysis of the directed human protein interaction network identifies disease genes and drug targets. *Proc Natl Acad Sci U S A.* 2016;113:4976–81.
61. Huang JK, Carlin DE, Yu MK, Zhang W, Kreisberg JF, Tamayo P, et al. Systematic evaluation of molecular networks for discovery of disease genes. *Cell Syst.* 2018;6:484–e4955.
62. Zheng S, Li Y, Chen S, Xu J, Yang Y. Predicting drug–protein interaction using quasi-visual question answering system. *Nat Mach Intell.* 2020;2:134–40.
63. Wu C-C, Wang YA, Livingston JA, Zhang J, Futreal PA. Prediction of biomarkers and therapeutic combinations for anti-PD-1 immunotherapy using the global gene network association. *Nat Commun.* 2022;13:42.
64. Barrio-Hernandez I, Schwartzentruber J, Shrivastava A, Del-Toro N, Gonzalez A, Zhang Q, et al. Network expansion of genetic associations defines a Pleiotropy map of human cell biology. *Nat Genet.* 2023;55:389–98.

65. Zhou G, Pang Z, Lu Y, Ewald J, Xia J. OmicsNet 2.0: a web-based platform for multi-omics integration and network visual analytics. *Nucleic Acids Res.* 2022;50:W527–33.
66. Zhao W, Gu X, Chen S, Wu J, Zhou Z. MODIG: integrating multi-omics and multi-dimensional gene network for cancer driver gene identification based on graph attention network model. *Bioinformatics.* 2022;38:4901–7.
67. Niu R, Guo Y, Shang X. GLIMS: A two-stage gradual-learning method for cancer genes prediction using multi-omics data and co-splicing network. *iScience.* 2024;27:109387.
68. Kim WK, Krumpelman C, Marcotte EM. Inferring mouse gene functions from genomic-scale data using a combined functional network/classification strategy. *Genome Biol.* 2008;9(1 Suppl 1):S5.
69. Zhao Y, Zheng K, Guan B, Guo M, Song L, Gao J, et al. DLDT: a learning-based framework for drug-target interaction identification using neural networks and network representation. *J Transl Med.* 2020;18:434.
70. Hobaiter C, Poisot T, Zuberbühler K, Hoppitt W, Gruber T. Social network analysis shows direct evidence for social transmission of tool use in wild chimpanzees. *PLoS Biol.* 2014;12:e1001960.
71. Di Nanni N, Bersanelli M, Milanese L, Mosca E. Network diffusion promotes the integrative analysis of multiple omics. *Front Genet.* 2020;11:106.
72. Li H, Li T, Quang D, Guan Y. Network propagation predicts drug synergy in cancers. *Cancer Res.* 2018;78:5446–57.
73. Altuntas V, Gök M, Kahveci T. Stability analysis of biological networks' diffusion state. *IEEE/ACM Trans Comput Biol Bioinform.* 2020;17:1406–18.
74. Cowen L, Ideker T, Raphael BJ, Sharan R. Network propagation: a universal amplifier of genetic associations. *Nat Rev Genet.* 2017;18:551–62.
75. Mohsen H, Gunasekharan V, Qing T, Seay M, Surovtseva Y, Negahban S, et al. Network propagation-based prioritization of long tail genes in 17 cancer types. *Genome Biol.* 2021;22:287.
76. Huang Z, Zeng D, Chen H. A comparison of Collaborative-Filtering recommendation algorithms for E-commerce. *IEEE Intell Syst.* 2007;22:68–78.
77. Wang B, Mezlini AM, Demir F, Fiume M, Tu Z, Brudno M, et al. Similarity network fusion for aggregating data types on a genomic scale. *Nat Methods.* 2014;11:333–7.
78. Rappoport N, Safra R, Shamir R. MONET: Multi-omic module discovery by omic selection. *PLoS Comput Biol.* 2020;16:e1008182.
79. Li L, Wei Y, Shi G, Yang H, Li Z, Fang R, et al. Multi-omics data integration for subtype identification of Chinese lower-grade gliomas: A joint similarity network fusion approach. *Comput Struct Biotechnol J.* 2022;20:3482–92.
80. Woo H-M, Yoon B-J. MONACO: accurate biological network alignment through optimal neighborhood matching between focal nodes. *Bioinforma Oxf Engl.* 2021;37:1401–10.
81. De Domenico M, Porter MA, Arenas A. MuxViz: a tool for multilayer analysis and visualization of networks. *J Complex Netw.* 2015;3:159–76.
82. Kipf TN, Welling M. Semi-supervised classification with graph convolutional networks. In: 5th international conference on learning representations, ICLR 2017, toulon, france, april 24–26, 2017, conference track proceedings. OpenReview.net; 2017.
83. Zitnik M, Agrawal M, Leskovec J. Modeling polypharmacy side effects with graph convolutional networks. *Bioinforma Oxf Engl.* 2018;34:i457–66.
84. Gao J, Lyu T, Xiong F, Wang J, Ke W, Li Z. MGNN: A Multimodal Graph Neural Network for Predicting the Survival of Cancer Patients. In: Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval. New York, NY, USA: Association for Computing Machinery; 2020. pp. 1697–700.
85. Peng W, Chen T, Dai W. Predicting drug response based on Multi-Omics fusion and graph Convolution. *IEEE J Biomed Health Inf.* 2022;26:1384–93.
86. Yan H, Weng D, Li D, Gu Y, Ma W, Liu Q. Prior knowledge-guided multilevel graph neural network for tumor risk prediction and interpretation via multi-omics data integration. *Brief Bioinform.* 2024;25:bbae184.
87. Schulte-Sasse R, Budach S, Hnisz D, Marsico A. Integration of multiomics data with graph convolutional networks to identify new cancer genes and their associated molecular mechanisms. *Nat Mach Intell.* 2021;3:513–26.
88. Hashim S, Nandakumar K, Yaqub M. Self-omics: A Self-supervised learning framework for Multi-omics cancer data. *Bio-computing 2023.* World Scientific; 2022. pp. 263–74.
89. Wang J, Liao N, Du X, Chen Q, Wei B. A semi-supervised approach for the integration of multi-omics data based on transformer multi-head self-attention mechanism and graph convolutional networks. *BMC Genomics.* 2024;25:86.
90. Ma A, Wang X, Li J, Wang C, Xiao T, Liu Y, et al. Single-cell biological network inference using a heterogeneous graph transformer. *Nat Commun.* 2023;14:964.
91. Kelly J, Berzuini C, Keavney B, Tomaszewski M, Guo H. A review of causal discovery methods for molecular network analysis. *Mol Genet Genomic Med.* 2022;10:e2055.
92. Michael T, Zhang JD. Causal inference in drug discovery and development. *Drug Discov Today.* 2023;28:103737.
93. Kotiang S, Eslami A. A probabilistic graphical model for system-wide analysis of gene regulatory networks. *Bioinforma Oxf Engl.* 2020;36:3192–9.
94. Mezlini AM, Goldenberg A. Incorporating networks in a probabilistic graphical model to find drivers for complex human diseases. *PLoS Comput Biol.* 2017;13:e1005580.
95. Liang Y, Kelemen A. Computational dynamic approaches for Temporal omics data with applications to systems medicine. *BioData Min.* 2017;10:20.
96. Zhang Y, Zhu L, Wang X, NEM-Tar: A probabilistic graphical model for cancer regulatory network inference and prioritization of potential therapeutic targets from Multi-Omics data. *Front Genet.* 2021;12.
97. B Z, Z Z, Sy L. X F. NetMIM: network-based multi-omics integration with block missingness for biomarker selection and disease outcome prediction. *Brief Bioinform.* 2024;25.
98. Gong B, Zhou Y, Purdom E. Cobolt: integrative analysis of multimodal single-cell sequencing data. *Genome Biol.* 2021;22:351.
99. Maity AK, Lee SC, Mallick BK, Sarkar TR. Bayesian structural equation modeling in multiple omics data with application to circadian genes. *Bioinformatics.* 2020;36:3951–8.
100. Bontha SV, Maluf DG, Mueller TF, Mas VR. Systems biology in kidney transplantation: the application of Multi-Omics to a complex model. *Am J Transpl Off J Am Soc Transpl Am Soc Transpl Surg.* 2017;17:11–21.

101. Yan J, Risacher SL, Shen L, Saykin AJ. Network approaches to systems biology analysis of complex disease: integrative methods for multi-omics data. *Brief Bioinform.* 2018;19:1370–81.
102. Regan-Fendt KE, Xu J, DiVincenzo M, Duggan MC, Shakya R, Na R, et al. Synergy from gene expression and network mining (SynGeNet) method predicts synergistic drug combinations for diverse melanoma genomic subtypes. *Npj Syst Biol Appl.* 2019;5:6.
103. Malik V, Kalakoti Y, Sundar D. Deep learning assisted multi-omics integration for survival and drug-response prediction in breast cancer. *BMC Genomics.* 2021;22:214.
104. Xu J, Mao C, Hou Y, Luo Y, Binder JL, Zhou Y, et al. Interpretable deep learning translation of GWAS and multi-omics findings to identify pathobiology and drug repurposing in Alzheimer's disease. *Cell Rep.* 2022;41:111717.
105. Qiu C, Yu F, Su K, Zhao Q, Zhang L, Xu C, et al. Multi-omics data integration for identifying osteoporosis biomarkers and their biological interaction and causal mechanisms. *iScience.* 2020;23:100847.
106. Lim H, Xie L. Omics data integration and analysis for systems Pharmacology. In: Larson RS, Oprea TI, editors. *Bioinformatics and drug discovery.* New York, NY: Springer New York; 2019. pp. 199–214.
107. Liu X-Y, Mei X-Y. Prediction of drug sensitivity based on multi-omics data using deep learning and similarity network fusion approaches. *Front Bioeng Biotechnol.* 2023;11.
108. Kim S, Bae S, Piao Y, Jo K. Graph convolutional network for drug response prediction using gene expression data. *Mathematics.* 2021;9:772.
109. Pu L, Singha M, Ramanujam J, Brylinski M. CancerOmicsNet: a multi-omics network-based approach to anti-cancer drug profiling. *Oncotarget.* 2022;13:695–706.
110. Zeng X, Zhu S, Hou Y, Zhang P, Li L, Li J, et al. Network-based prediction of drug–target interactions using an arbitrary-order proximity embedded deep forest. *Bioinformatics.* 2020;36:2805–12.
111. Huang L, Brunell D, Stephan C, Mancuso J, Yu X, He B, et al. Driver network as a biomarker: systematic integration and network modeling of multi-omics data to derive driver signaling pathways for drug combination prediction. *Bioinformatics.* 2019;35:3709–17.
112. Rashid MM, Selvarajoo K. Advancing drug-response prediction using multi-modal and -omics machine learning integration (MOMLIN): a case study on breast cancer clinical data. *Brief Bioinform.* 2024;25:bbab300.
113. Hjordt P, Hugo J, Zeinert J, Müller N, Kashyap S, Rajapakse JC, et al. DrDimont: explainable drug response prediction from differential analysis of multi-omics networks. *Bioinformatics.* 2022;38 Supplement2:ii13–9.
114. Peng W, Liu H, Dai W, Yu N, Wang J. Predicting cancer drug response using parallel heterogeneous graph convolutional networks with neighborhood interactions. *Bioinformatics.* 2022;38:4546–53.
115. Paltun BG, Kaski S, Mamitsuka H. DIVERSE: bayesian data integrative learning for precise drug response prediction. *IEEE/ACM Trans Comput Biol Bioinform.* 2022;19:2197–207.
116. Wang X, Zhu H, Jiang Y, Li Y, Tang C, Chen X, et al. PRODeepSyn: predicting anticancer synergistic drug combinations by embedding cell lines with protein–protein interaction network. *Brief Bioinform.* 2022;23:bbab587.
117. Nguyen T, Nguyen GTT, Nguyen T, Le D-H. Graph convolutional networks for drug response prediction. *IEEE/ACM Trans Comput Biol Bioinform.* 2022;19:146–54.
118. Xu M, Zhu Z, Zhao Y, He K, Huang Q, Zhao Y. RedCDR: dual relation distillation for cancer drug response prediction. *IEEE/ACM Trans Comput Biol Bioinform.* 2024;1–12.
119. Leiserson MDM, Vandin F, Wu H-T, Dobson JR, Eldridge JV, Thomas JL, et al. Pan-cancer network analysis identifies combinations of rare somatic mutations across pathways and protein complexes. *Nat Genet.* 2015;47:106–14.
120. Ritchie MD, Holzinger ER, Li R, Pendergrass SA, Kim D. Methods of integrating data to uncover genotype-phenotype interactions. *Nat Rev Genet.* 2015;16:85–97.
121. Zhou J, Cui G, Hu S, Zhang Z, Yang C, Liu Z, et al. Graph neural networks: A review of methods and applications. *AI Open.* 2020;1:57–81.
122. Aldea M, Friboulet L, Apcher S, Jaulin F, Mosele F, Sourisseau T, et al. Precision medicine in the era of multi-omics: can the data tsunami guide rational treatment decision? *ESMO Open.* 2023;8:101642.
123. Creixell P, Reimand J, Haider S, Wu G, Shibata T, Vazquez M, et al. Pathway and network analysis of cancer genomes. *Nat Methods.* 2015;12:615–21.
124. Sonawane AR, Weiss ST, Glass K, Sharma A. Network medicine in the age of biomedical big data. *Front Genet.* 2019;10:294.
125. Adossa N, Khan S, Rytönen KT, Elo LL. Computational strategies for single-cell multi-omics integration. *Comput Struct Biotechnol J.* 2021;19:2588–96.
126. Miao Z, Humphreys BD, McMahon AP, Kim J. Multi-omics integration in the age of million single-cell data. *Nat Rev Nephrol.* 2021;17:710–24.
127. Jia Q, Chu H, Jin Z, Long H, Zhu B. High-throughput single-cell sequencing in cancer research. *Signal Transduct Target Ther.* 2022;7:145.
128. Jin M, Koh HY, Wen Q, Zambon D, Alippi C, Webb GI, et al. A survey on graph neural networks for time series: forecasting, classification, imputation, and anomaly detection. *IEEE Trans Pattern Anal Mach Intell.* 2024;46:10466–85.
129. Zhang Y, Boninsegna L, Yang M, Misteli T, Alber F, Ma J. Computational methods for analysing multiscale 3D genome organization. *Nat Rev Genet.* 2024;25:123–41.
130. He Z, Hu S, Chen Y, An S, Zhou J, Liu R, et al. Mosaic integration and knowledge transfer of single-cell multimodal data with MIDAS. *Nat Biotechnol.* 2024;42:1594–605.
131. Shen R, Liu L, Wu Z, Zhang Y, Yuan Z, Guo J, et al. Spatial-ID: A cell typing method for Spatially resolved transcriptomics via transfer learning and Spatial embedding. *Nat Commun.* 2022;13:7640.
132. Subramanian I, Verma S, Kumar S, Jere A, Anamika K. Multi-omics data integration, interpretation, and its application. *Bioinforma Biol Insights.* 2020;14:117793221989905.

Publisher's note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.