



OPEN ACCESS

EDITED BY

Tianzhou Ma,
University of Maryland, United States

REVIEWED BY

Nektarios Valous,
German Cancer Research Center (DKFZ),
Germany
Menglu Liang,
University of Maryland, United States

*CORRESPONDENCE

Hugo Chenel,
✉ hugo.chenel@inserm.fr
Andrei Zinovyev,
✉ andrei.zinovyev@evotec.com
Vera Pancaldi,
✉ vera.pancaldi@inserm.fr

RECEIVED 28 December 2025

REVISED 09 April 2026

ACCEPTED 13 April 2026

PUBLISHED 04 June 2026

CITATION

Chenel H, Marku M, James T, Zinovyev A
and Pancaldi V (2026) Multilayer network
approaches to omics data integration in
digital twins for cancer research.
Front. Syst. Biol. 6:1776941.
doi: 10.3389/fsysb.2026.1776941

COPYRIGHT

© 2026 Chenel, Marku, James, Zinovyev
and Pancaldi. This is an open-access
article distributed under the terms of the
[Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/)
(CC BY). The use, distribution or
reproduction in other forums is permitted,
provided the original author(s) and the
copyright owner(s) are credited and that
the original publication in this journal is
cited, in accordance with accepted
academic practice. No use, distribution or
reproduction is permitted which does not
comply with these terms.

Multilayer network approaches to omics data integration in digital twins for cancer research

Hugo Chenel^{1,2,3*}, Malvina Marku¹, Tim James⁴, Andrei Zinovyev^{3*}
and Vera Pancaldi^{1,2*}

¹Univ Toulouse, INSERM, CNRS, CRCT, Toulouse, France, ²Equipe Labellisée Ligue Nationale Contre le Cancer, Paris, France, ³Evotec, Toulouse, France, ⁴Evotec, Abingdon, Oxfordshire, United Kingdom

How can we effectively integrate and represent heterogeneous multi-omics data to better understand complex biological systems and support the development of personalized digital models? The growing availability of high-dimensional omics data across different molecular scales presents both opportunities and challenges in biomedical research. Traditional approaches often treat each omics layer in isolation or rely on concatenation strategies that obscure the interactions between different regulatory layers. In this review, we discuss the multilayer network-based framework as an extensive representation of different omics data types, capturing the modularity, redundancy and cross-talk between layers, and providing a more faithful interpretable view of the biological system. We explore how this approach can be used as a basis for the construction of Digital Twins, computational replicas of individual biological systems capable of simulating disease progression and treatment outcomes. In contrast to existing multi-omics integration reviews), we emphasize the role of multilayer networks as a mechanistic and interpretable scaffold for Digital Twins development. We discuss key methodological considerations, benefits and potential applications of this approach, highlighting its promise for advancing both our understanding of biological complexity and our ability to design personalized interventions.

KEYWORDS

cancer systems biology, digital twins, multilayer networks, multi-omics integration, network medicine, precision oncology

Introduction

Precision oncology faces a major challenge: how to integrate heterogeneous multi-omics data into coherent and clinically actionable models of disease. The emergence of precision medicine warrants the development of tools that can exploit the vast and complex datasets generated by multi-omics approaches. These technologies, including (epi)genomics, transcriptomics, proteomics, metabolomics, metagenomics and other techniques that exhaustively characterize specific sets of biological entities, provide a comprehensive view of biological systems at multiple levels of regulation. Integrating these diverse types of datasets is essential to have an extensive and inclusive digital representation of patients and to identify novel biomarkers or therapeutic targets (Hasin et al., 2017), especially in cancer, which will be the focus of this review.

Digital Twins (DTs) have emerged as a transformative paradigm in healthcare, providing dynamic, *in silico* models of cells, organs, or entire organisms that are continuously updated with real-time data. These virtual counterparts enable predictive modeling and personalized interventions (Kamel Boulos and Zhang, 2021). DTs leverage

advanced computational techniques to integrate and analyze large-scale multi-omics data to implement precision oncology. However, current DT approaches often rely on predictive or data-driven models that do not explicitly capture interactions across multiple biological scales, limiting their ability to provide mechanistic insights and to represent the multi-scale nature of cancer systems. From a systems biology point of view, an important type of DTs is based on network representations that span multiple scales and layers of biological information, often realized through multilayer networks. Multilayer networks (Kivela et al., 2014) have gained popularity for their ability to model complex systems where entities interact across multiple layers or dimensions. Early applications of multiplex biological networks demonstrated the value of propagating information across interconnected molecular layers for improved disease module identification (Valdeolivas et al., 2019; Baptista et al., 2024). In the context of multi-omics data integration, these multilayer networks can represent different types of biological information as interconnected layers, to capture the relationships and interactions between genes, proteins or other molecular entities (Lee et al., 2020; Trimbouret et al., 2024), enabling the identification of crucial regulatory mechanisms and pathways that may remain elusive in single-layer network analyses, or using simple aggregation of these networks (De Domenico, 2023). Multilayer networks can be thought of as a mathematical foundation of knowledge graphs.

While the ongoing research in integrating multilayer networks in DTs shows promising outcomes (Vallée, 2026), several conceptual and practical challenges in their implementation for personalized medicine persist, including the intrinsic heterogeneous nature of multi-omics datasets, requiring elaborated algorithms (and potentially substantial computational resources). Ensuring data quality or consistency remains a significant obstacle. Finally, the development of reliable DTs requires robust validation methods to ensure that their predictive capabilities align with real-world outcomes (Hernandez-Boussard et al., 2021).

Despite these advances, there is still a lack of integrative frameworks that simultaneously preserve the structure of individual omics layers, model cross-layer interactions, and also support mechanistic interpretation required for DTs applications. In this review, we address both the importance and the challenges of integrating heterogeneous multi-omics datasets by adopting a multilayer network representation, which enables the modeling of complex biological interactions across molecular layers. This approach preserves the intrinsic structure of each omics layer while allowing modeling of inter-layer relationships, thus offering a more comprehensive and interpretable view of biological systems. We further discuss how this multilayer representation provides a foundational framework for the development of DTs in medicine that can simulate individual-specific disease progression and therapeutic responses.

Multi-omics data integration approaches

Integration of multi-omics data has been an active area of research for more than a decade, leading to several frameworks which can be categorized based on assumptions, computational complexity, interpretability, and the tasks for which they are

designed (clinical outcome prediction, mechanistic inference of cross-omics molecular interactions, or dynamic modeling of the temporal evolution of biological systems). Different classes of multi-omics data integration approaches have been proposed, mostly differing in the stage at which data modalities are integrated (Wang et al., 2022; Picard et al., 2021) (Figure 1a).

Early integration techniques typically combine different data types by concatenating features in a single matrix before analysis (Bersanelli et al., 2016). Such an approach is easy to understand and can be easily implemented with many different machine learning (ML) techniques, like regression analysis, clustering or deep learning. This simple approach is often criticized for excessive feature dimensionality and weak interpretability (Hauptmann and Kramer, 2023). Most importantly, it fails to capture the different biological semantics of each data layer and does not explicitly model cross-layer interactions.

Intermediate integration methods, which rely on latent variables or factors, are designed to learn common and/or modality-specific representations across different omics layers. Some examples include matrix factorization, canonical correlation analysis, Bayesian latent variable models and multi-omics factor analysis (MOFA) (Argelaguet et al., 2018). These methods reduce dimensionality while preserving shared sources of variation between datasets, making them particularly useful for patient stratification or biomarker discovery. However, latent variable approaches may trade interpretability for compactness, failing to preserve relationships between molecular entities across different regulatory layers, which is essential for mechanistic models and simulation (Mardinoglu et al., 2025). In DT applications, representations of cross-layer biological interactions and mechanistic detail are essential to simulate system dynamics.

Recent advancements in representation learning further highlight this tension. For example, Joint Embedding Predictive Architectures (JEPA) prioritize the learning of latent representations that are predictive without relying on autoencoder-style reconstruction of the data input (LeCun, 2022), improving robustness but compromising biological interpretability (mathematically useful, not biologically meaningful). Related efforts in single-cell analysis attempt to impose additional structure on latent representations to improve interpretability (Bahrami et al., 2025).

As a result, while techniques of intermediate integration are effective in discovery-oriented and predictive tasks, the abstraction of the latent space makes them less appropriate for building DTs, where interpretability, interaction at different scales, and simulation are critical.

In late integration methods, each omics layer is analyzed independently, typically using modality-specific frameworks, with predictions or learned representations combined only at a later stage, for example, through deep neural networks (Sharifi-Noghabi et al., 2019). This method is advantageous because of its robustness, especially when dealing with structurally different omics layers, but it lacks the ability to capture cross-layer interactions, making it less suitable for DTs.

Integration strategies across modeling paradigms

Besides the timing of integration, multi-omics approaches can also be classified according to the modeling paradigm that they

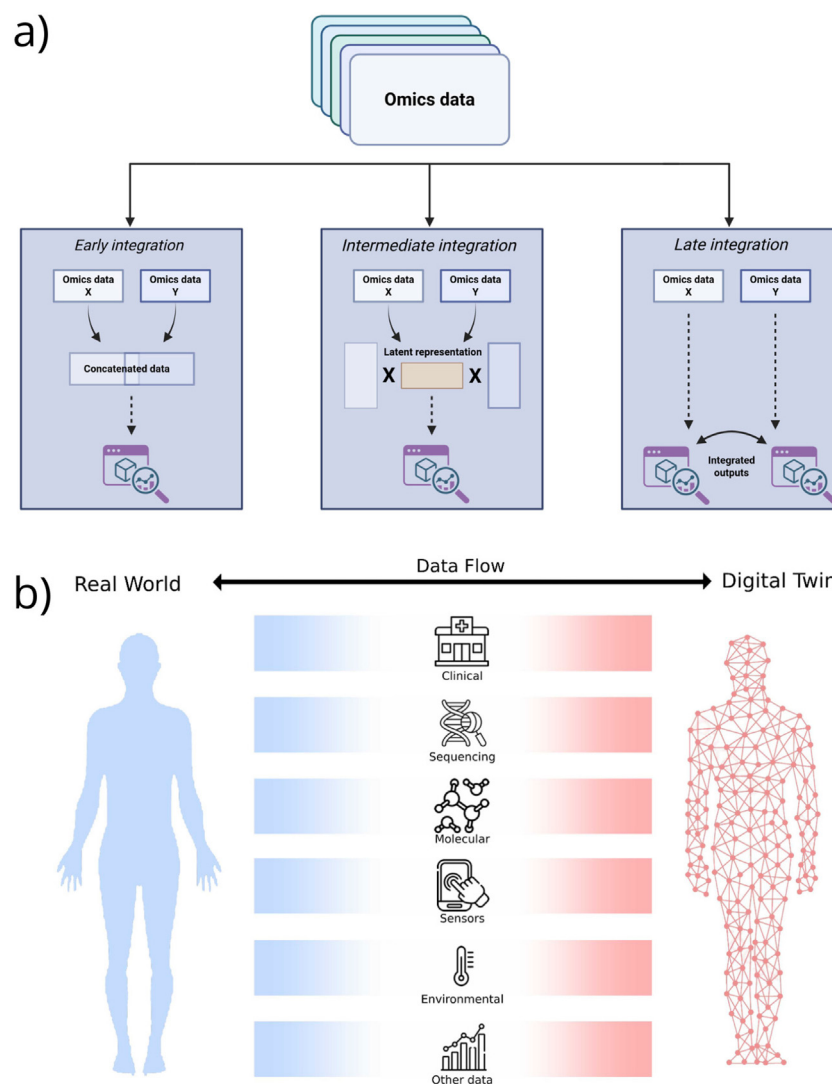


FIGURE 1
 Frameworks for multi-omics integration and DT modeling in cancer research. **(a)** Schematic representation of different strategies for multi-omics integration. **(b)** Conceptual diagram of DTs in the context of cancer research.

adopt, namely, linear, deep learning-based and network-based (Table 1).

Most linear techniques include the early stages of feature concatenation coupled with regression or PCA, intermediate stages that are represented by common latent factor models that employ CCA, MOFA (Argelaguet et al., 2018), DIABLO (Singh et al., 2019) from mixOmics (Rohart et al., 2017) or Mowgli (Huizing et al., 2023), and the late stages that are represented by feature ensembling techniques.

Deep learning-based methods span from early input-level fusion using neural networks to intermediate latent-space fusion via multi-modal autoencoders, variational autoencoders (VAEs), graph neural networks, or contrastive learning frameworks (Ballard et al., 2024), to late prediction-level ensembles like MOLI (Sharifi-Noghabi et al., 2019). These approaches enable cross-modal representation learning but typically operate in abstract latent spaces with limited mechanistic transparency.

In parallel, the field of multi-omics analysis has also been impacted by recent advances in representation learning and foundation models.

Specifically focusing on single-cell omics, models such as scVI (Lopez et al., 2018) learn probabilistic latent representations that can be used for alignment, denoising and transfer learning, while the related extension totalVI (Gayoso et al., 2021) extends this framework to multimodal single-cell integration. More generally, predictive representation learning architectures (JEPA) have been proposed to focus on learning abstract latent embeddings that do not require input reconstruction (LeCun, 2022). More recent models, notably scGPT (Cui et al., 2024), have continued this line of research by developing transformer-based models that are pre-trained on large datasets for flexible downstream tasks. Similar foundation-model approaches have also been proposed for biological data, including scPRINT (Kalfon et al., 2025a) and related large-scale biological foundation models (Kalfon et al., 2025b). Although these models are not necessarily developed for multi-omics integration, a trend towards highly abstract latent representations that focus on predictive performance and generalization rather than interpretability is clear throughout the literature.

TABLE 1 Taxonomy of integration strategies across modeling paradigms.

Modeling paradigm	Type	Early integration	Intermediate integration	Late integration downstream analysis
Linear methods	Concepts	Feature concatenation PCA on concatenated omics, linear/logistic regression, elastic net, linear SVM	Shared latent factors CCA, joint matrix factorization, multi-block PLS, tensor ICA	Model ensembling Separate regressions per omics, weighted averaging, voting/stacking classifiers
	Tools	None specific	DIABLO (mixOmics); MOFA/MOFA+; RGCCA; Mowgli	None specific
Deep learning-based	Concepts	Input-level fusion Fully connected neural networks, deep autoencoders	Latent-space fusion Multimodal autoencoders, VAEs, contrastive learning models, JEPA-style representations	Prediction-level ensembles Omics stacking, stacked neural networks, DeepEnsemble
	Tools	autoOmics; DeepOmix; flexynesis	MOGONET; MAUI; totalVI; OmiEmbed; OmiVAE; scPairing	MOLI; Super.FELT
Network-based	Concepts	Aggregated network construction Correlation networks from concatenated data	Hybrid latent-network models Multilayer networks, similarity-based fusion, graph-regularized factor models	Separate networks + result integration Independent PPI, GRN, pathway networks with downstream score aggregation
	Tools	None specific	SNF; OmicsIntegrator; netDx	clusterProfiler; EnrichmentMap; WGCNA, ARACNe, GENIE3, CollecTRI, VIPER

Approaches based on networks explicitly model relationships, either between samples or between biological entities. Methods such as similarity network fusion (SNF) construct networks of samples for each omics layer and integrate them to improve tasks such as cancer subtyping (Wang et al., 2014). In contrast, multilayer biological network approaches represent interactions between molecular entities across multiple omics layers. Inter-layer edges in such networks can be defined based on different sources of biological evidence, including known molecular relationships (gene-protein mappings), regulatory interactions, or data-driven associations inferred from multi-omics datasets. However, the definition of these edges is often subject to uncertainty, particularly when inferred from noisy or heterogeneous data, which often affect the robustness of the resulting network models. Although these models preserve cross-omics structure and provide a natural framework for mechanistic interpretation, they remain comparatively underexplored in practical multi-omics integration workflows, with only a limited number of recent approaches explicitly implementing such architectures (Trimbour et al., 2024).

Finally, integration strategies also vary based on the structural relationship between the datasets. In vertical integration, multiple modalities are measured on the same samples, while mosaic integration is based on partially overlapping sample sets. Diagonal integration, on the other hand, refers to datasets with non-overlapping samples. These structural differences play a significant role in data integration, especially when integrating heterogeneous datasets (Argelaguet et al., 2020; Cantini et al., 2021). These distinctions are critical for selecting integration methods suitable for DT development.

Beyond multi-omics and multimodal integration, there has been growing interest in the combination of bulk and single-cell transcriptomic data, which is a cross-resolution problem rather than a cross-omics challenge. In this case, the same type of

modality is assessed at different biological scales; there is a need for methods that can align signals to reconcile population-level and cell-level variation.

Implications for digital twins

The development of meaningful biological DTs involves the application of multi-omics integration strategies that extend beyond prediction and specifically capture cross-scale interactions, ensure mechanistic interpretability, and facilitate longitudinal modeling (Mardinoglu et al., 2025). Recent studies highlight the need for DTs to integrate heterogeneous molecular layers with clinical, imaging and phenotypic data in a temporally resolved manner, thus enabling continuous updates of patient-specific models as new data become available (Hsu et al., 2025). Within this setting, emerging multi-omics time series integration frameworks underscore the need to explicitly model longitudinal dynamics rather than relying on the application of cross-sectional latent representations, thus closely aligning with the simulation and prediction objectives of DTs (Sherwani et al., 2025).

In addition, a number of recent conceptual models further suggest that DTs should, in fact, be computational avatars that integrate various omics-based variables with clinical variables, facilitating identification of new biomarkers that can be used for early detection, improved tumor diagnosis and therapy optimization (Lococo et al., 2023). Digital Human Avatars models stress the importance of a structured model, such as hybrid mechanistic/statistical models or networks, that is biologically meaningful but also computationally tractable. In a complementary direction, a number of recent perspectives in cancer systems biology stress that, in order to be biologically plausible, robust and trustworthy, DTs should integrate data-driven learning with mechanistic constraints (Asghar and Chung, 2025).

Taken together, these perspectives indicate that integration methods optimized for predictive performance alone, such as latent variable or black-box deep learning models, are not sufficient for DT implementations, which require mechanistic and longitudinal approaches to simulate interventions (Mardinoglu et al., 2025). Network models and hybrid models, especially in their time-evolving and patient-specific variants, provide a crucial methodological foundation for the development of the next-generation of clinically actionable DTs.

Digital twins in precision health

Definition and concept

By offering opportunities for individualized healthcare through virtual models, DTs represent a paradigm shift in the field of precision health and personalized medicine. In healthcare, DTs can be defined as highly detailed and dynamic virtual replicas of physical entities (cells, tissues, organs, patients and health systems), reproducing the structure, behavior and context of their physical counterpart (Qi et al., 2021). These virtual replicas are continuously updated with real-time data to simulate and predict health outcomes, thereby optimizing clinical decision-making (Fuller et al., 2020). DTs integrate data from multiple sources in real-time to simulate health outcomes and can be adapted to each patient, potentially increasing proactive healthcare management by enabling more precise, timely and effective clinical interventions, ultimately improving patient treatment efficiency.

Beyond computational models, DTs are not limited to simple static digital models; they involve a bidirectional flow of information between the physical entity and its digital counterpart, ensuring that the virtual model evolves at the same time as its real-world counterpart (Figure 1b). This continuous synchronization enables DTs to reflect the actual current state of the physical entity accurately, as well as predict future states under various scenarios (Kamel Boulos and Zhang, 2021). For instance, in a clinical setting, DTs of a patient would integrate data from various sources (clinical records, genetic data, lifestyle information, real-time monitoring from wearable devices, etc.) to provide a comprehensive dynamic representation of the patient's health status and guide recommendations for therapeutic interventions or increased surveillance.

The effectiveness of DTs depends on their predictive capability (Fuller et al., 2020). Leveraging real-time data and historical health information helps DTs forecast various factors such as disease progression, outcomes of different treatment options, or early signs of health deterioration. This predictive capability is particularly valuable in the management of chronic diseases, where early intervention can significantly influence disease progression. DTs can be used not only for optimizing disease treatments, but also to suggest strategies for preventing disease onset and mitigating its impact. Their predictive power comes from modeling disease trajectories, i.e., mapping a disease's progression over time, including its onset, development and chronicity, when relevant, starting from clinical records, multi-omics data and real-time health metrics. Similar data-driven

approaches have also been applied to predict disease risk and evolution using longitudinal health data and ML (Lasko et al., 2025). Therefore, they offer the potential for medical professionals to proactively adjust treatment plans, and anticipate significant turning points in the disease's course. They can also be used to provide alerts for early indicators of health deterioration for chronic diseases, whose trajectory frequently includes periods of remission or relapse.

Moreover, the DTs framework can help in the creation of a learning healthcare system where continuous feedback from DTs informs clinical decisions and enhances healthcare delivery. This iterative process not only improves individual patient care, but also contributes to the broader knowledge base, driving key advancements in medical research and clinical practice (Kamel Boulos and Zhang, 2021).

Potential and existing applications of DTs or DT-like models in healthcare

There are three main uses that DTs could have in healthcare, namely, disease management and treatment optimization, virtual clinical trials, and patient education.

Personalized medicine and disease management are a few of the areas in which DTs are gaining recognition for their transformative potential in the healthcare industry. Several existing systems integrate multi-omics data with clinical and lifestyle information to create a more complete and individualized health profile that can serve as a basis for *in silico* disease modeling. These models can form the basis for clinical decision support systems, providing predictions of responses to the different treatment options, a highly valuable tool for clinicians who can tailor interventions based on the unique characteristics of each patient. In addition, DTs enable customized interventions by utilizing real-time data to offer precise and dynamic insights into patient health; this involves continuous mapping of the physical/digital counterparts.

Virtual representations play an important role in the prevention and management of chronic disease. The constant monitoring of physiological data allows DTs to identify early signs of disease onset but also aggravation, which is crucial for timely intervention.

In the context of diabetes management, DTs have demonstrated the capacity to predict glucose fluctuations by integrating data on dietary intake, physical activity or medication adherence, making them a powerful tool to help clinicians with better management of the disease and preventing complications (Shamanna et al., 2021). Recently, this has led to the design of DT systems that administer insulin through real-time monitoring of patient glucose levels (Thamotharan et al., 2023). Other studies in cardiovascular health demonstrate how DTs can monitor vital or relevant cardiovascular signs (resting heart rate, heart rate variability, blood pressure) to detect abnormalities at an early stage and suggest preventive measures (Coorey et al., 2022).

Specifically in oncology, advanced computational frameworks have long been used for modeling tumor growth and predicting how a tumor will respond to different therapeutic interventions (Benzekry et al., 2014), leading to optimization of cancer patient treatment plans by improving drug effectiveness and minimizing side effects (Bruynseels, Santoni de Sio and van den Hoven, 2018). Recent work at the interface of mathematical oncology and AI

further emphasizes the role of computational models in survival prediction (Rockne et al., 2026; Taieb et al., 2026).

However, only a few real DTs for oncology have been developed so far. This is due to both the difficulty in obtaining longitudinal data, and the near-impossibility of dynamically adapting treatments in response to model predictions. As a result, the systems currently described in the literature often lack bidirectional feedback between the real system and the digital counterpart, suggesting that what we discuss here could be referred to as Digital Shadows, as discussed in (Pellegrino et al., 2025).

Some studies demonstrated a significant increase in treatment efficacy that shows the potential of mathematical models that are similar to DTs in adapting cancer therapies. Simulations predicting a trajectory for a patient's cancer under one of the treatment options can be used to adapt treatment (Hernandez-Boussard et al., 2021). Additional examples of oncological DTs are reviewed in (Stahlberg et al., 2022), illustrating the limits but also the potential of DTs to improve early diagnosis of cancers that are generally detected at advanced stages. Some projects on the detection of pancreatic cancer found subtle changes in biomarker profiles that could be detected at an early stage, potentially indicative of the initial stages of pancreatic cancer (Boyd et al., 2023). However, matching real patients in cohorts with simulated patients from the model did not result in a definitive mapping between real-world data and model parameters (Junet et al., 2023). Another project involved using simulations of cellular interactions in the tumor to personalize melanoma treatment, showing that computational models integrating tumor growth or immune response could predict patient-specific outcomes in melanoma therapy (Perlstein et al., 2019; Hodgkinson et al., 2022).

Another remarkable application predicts drug efficacy in lung cancer using DTs (Wang et al., 2024). DT models allow the simulation of different delivery methods and dosages under different therapeutic scenarios to establish treatment efficacy. Understanding these mechanisms is essential for clinicians to be able to adjust treatment plans as a therapeutic or preventive measure (Vallée, 2024). Some studies also underscored DTs' value in the anticipation of treatment response in cancer therapy (Venkatesh et al., 2024).

On the other side, further from real patients, testing new drugs or treatments via virtual platforms is an application of DTs that is transforming drug discovery. The simulation of biological processes and disease progression is implemented in DTs to predict the interaction of new compounds with different molecular targets, observing how these processes or the disease will evolve in simulated diverse patient populations constituting virtual cohorts (Wang et al., 2024). This approach reduces the need for extensive *in vivo* testing, improving the drug development process.

Finally, DTs could also play a major role in patient education, by improving their own awareness and encouraging them to be proactive about their health. If an interactive representation of health status via Patient Reported Outcomes and the impact of different lifestyle choices is provided to patients, DTs could indirectly enhance motivation for self-management. For instance, DTs can show some patients with obesity the impact of weight loss on some health metrics, encouraging changes in diet and exercise regimens (Gkouskou et al., 2020).

These examples show that the applications of DTs in healthcare are broad: personalized medicine, disease prevention/management, drug discovery and patient engagement. The integration of real-time data by DTs offers dynamic insights that ameliorate personalized healthcare. As the technology continues to evolve, the potential of these DTs to improve healthcare outcomes should continue to grow, highlighting their importance in the future of precision health (Kamel Boulos and Zhang, 2021). Examples from the literature reveal the potential of dynamic and predictive models to improve early detection, personalize treatment strategies, or anticipate therapeutic resistance. However, the development of tools that enable feedback between the DT and real patients is still in progress (Fuller et al., 2020), holding promise in cancer research.

Integrating multi-omics data in DTs of cancer patients: challenges and opportunities

A full comprehension of molecular mechanisms driving cancer progression is made possible by the integration of multi-omics data, which has become an essential component in oncology. The word "omics" refers to a group of scientific areas, each of which focuses on specific biological entities (Srivastava et al., 2024). Genomics, epigenomics, transcriptomics, proteomics, metabolomics, metagenomics and even radiomics (medical imaging such as MRI, CT scans, etc.) are the main categories of data that are important for cancer research. We give an overview of different types of omics data, their general characteristics and relevance to cancer research in Table 2.

The integration of multi-omics data in oncology offers significant potential to advance the understanding of cancer biology. However, several challenges must be addressed to fully realize its promise. These challenges can be technical, computational and biological in nature, also including issues with data administration policies, each representing unique opportunities for innovation.

One of the main issues in multi-omics integration is the inherent heterogeneity of the data, stemming from both its nature and its method of generation. Different omics data sources are generated using diverse technologies, each having its own data format (Cao et al., 2021; Ren et al., 2023), indicating the necessity for a deeper investigation of the crosstalk between different levels of data granularity (Cirillo et al., 2021). This inconsistency adds complexity to the integration process, because the harmonization and standardization of these datasets require advanced methodologies (Hasin et al., 2017).

The high volume of multi-omics data requires efficient algorithms for extracting useful and interpretable information. Also, this high-dimensional data integration involves handling large-scale datasets that demand high computational power as well as scalable storage capacity (Nikolski et al., 2024). Moreover, the development of efficient algorithms for the different steps of data processing (preprocessing, normalization and integration) is crucial to deal with this complexity (Santiago-Rodriguez and Hollister, 2021). ML or AI techniques are increasingly employed to address these challenges (Rockne et al., 2026), but their implementation requires substantial expertise and resources (Angermueller et al.,

TABLE 2 Different types of omics data, their characteristics and relevance to cancer research.

Omics type	Description	Technologies	Applications to cancer research
Genomics	Study of DNA sequence, identifying mutations and genetic alterations driving disease	Whole-genome sequencing (WGS), Whole-exome sequencing (WES), liquid biopsy (ctDNA sequencing)	Identification of mutations in oncogenes, tumor suppressor genes, and genetic alterations, non-invasive tumor profiling and monitoring through circulating tumor DNA
Epigenomics	Study of changes in chromatin that do not involve changes in the DNA sequence, including DNA methylation, histone modifications, chromatin accessibility and organization	ChIP-seq, Cut&Run, ATAC-seq, Hi-C	Understanding epigenetic regulation and gene silencing
Transcriptomics	Study of RNA transcripts, providing insights into gene expression	RNA sequencing (RNA-seq), single-cell RNA-seq (scRNA-seq)	Identification of dysregulated genes, tumor heterogeneity, molecular subtypes, and patient stratification
Proteomics	Large-scale study of proteins, including their expression, modifications, and interactions	Mass spectrometry (LC-MS/MS), protein arrays	Identification of altered protein expression, protein biomarkers, and targets for therapy in cancer
Metabolomics	Study of metabolites, small molecules that are intermediate products of metabolism, offering a functional readout of biochemical activities	Nuclear magnetic resonance (NMR), Mass spectrometry (MS)	Characterization of metabolic alterations in cancer, such as the Warburg effect, and understanding of their role in tumor development
Metagenomics	Analysis of the genetic material from microbial communities, in the gut/saliva/skin or at tumor sites, offering insights into the microbiome's role in cancer	Shotgun metagenomic sequencing, 16S rRNA sequencing	Exploration of microbiome influence on tumorigenesis and potential microbiota-based cancer therapies
Spatial omics	Study of the spatial organization of cellular components within the TME, preserving tissue architecture during analysis	Spatial transcriptomics, spatial proteomics	Visualization of cellular heterogeneity and spatial patterns in tumors
Exposomics	Study of lifetime environmental exposures and their interaction with the genome, focusing on external and internal factors such as chemicals, diet, and lifestyle	Data mostly from environmental sensors, wearable devices, or questionnaires	Understanding of environmental factors contributions to cancer risk through gene-environment interactions

2016; Libbrecht and Noble, 2015; Cai et al., 2022; Acharya and Mukhopadhyay, 2024).

Ensuring data quality and reproducibility is another significant challenge. Omics data can be affected by various technical issues at different steps of the experiments or analysis pipeline (technical biases, batch effects, sample-specific variations, etc.), which can hide biological signals and lead to false conclusions (Leek et al., 2010). Therefore, rigorous quality control measures, standardized protocols and reproducible workflows are essential to mitigate all these issues. Additionally, the integration process must account for missing data and varying sample sizes across different omics layers (Kim et al., 2005).

On another level, complex datasets are generated by the integration of multi-omics data, which requires careful interpretation to derive biologically meaningful insights. For instance, the identification of key drivers in cancer and the elucidation of their functional roles in the context of multi-omics data is an important and challenging task. Distinguishing true biological signals from noise and correlating multi-omics findings with clinical outcomes is non-trivial. Collaborative efforts between researchers and clinicians are essential to contextualize the results

and then translate them into actionable knowledge (Demirel et al., 2022).

Despite all these challenges, the integration of multi-omics data currently offers substantial opportunities to deepen our understanding of cancer biology at the tumor site and across the whole organism. The combination of information from several types of omics data can bring an extensive view of the molecular mechanisms that drive tumor growth via different types of data integration (Huang et al., 2017) (Figure 2a). This global approach helps to identify new biomarkers that may not be apparent from single-omics studies alone, with some methods taking into consideration data type diversity and big data volume (Wu et al., 2015).

Great progress can be achieved in personalized medicine thanks to multi-omics integration. The detailed molecular profiling of individual tumors can reveal patient-specific vulnerabilities, to enable designing customized therapeutic strategies for the patients. For instance, integrating genomic and transcriptomic data is interesting to identify actionable mutations or dysregulated pathways, while transcriptomic, proteomic and metabolomic analyses can improve the understanding of tumor

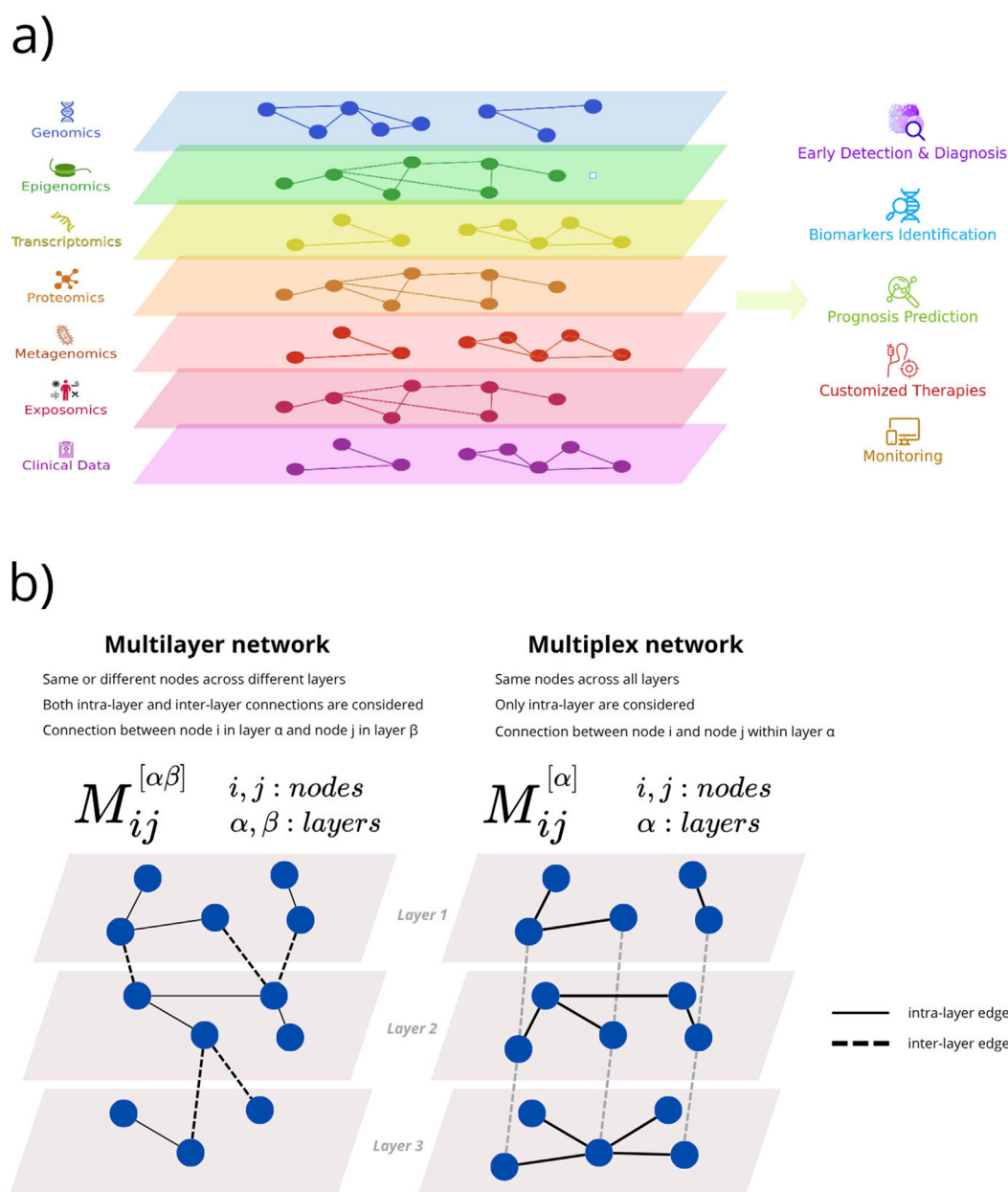


FIGURE 2

Multi-omics interactions and their network-based representation in therapeutic modeling. **(a)** An overview of different multi-omics layers and their potential applications in clinically relevant tasks. **(b)** Schematic representation of multilayer and multiplex networks. A multilayer network showing intra- and inter-layer connections (left) and a special case of a multilayer, a multiplex network, in which the nodes are the same in all layers (right).

metabolism alterations and cellular phenotypes (Zhang and Plevritis, 2018; Gegner et al., 2023).

Developing novel algorithms or ML models to handle high-dimensional, heterogeneous data is another emergent area of research (Wu et al., 2019). These advancements not only improve the performance (efficiency and accuracy) of data integration, but also contribute in a broader way to the field of computational biology. The complexity of the integration makes collaborative efforts and data sharing among researchers and institutions absolutely necessary (Gómez-López et al., 2019). Some initiatives, such as The Cancer Genome Atlas (TCGA)

(Weinstein et al., 2013), International Cancer Genome Consortium (ICGC) (Zhang et al., 2011) and Human Cell Atlas (HCA) (Rozenblatt-Rosen et al., 2017), provide highly valuable resources and infrastructure for multi-omics research. Some recent resources are specific to certain types of omics data, such as the Human Protein Atlas (Pontén et al., 2008) (proteomics), International Human Epigenome Consortium (IHEC) (Stunnenberg et al., 2016) (epigenomics), Human Metabolome Database (HMDB) (Wishart et al., 2007) (metabolomics), or the Virtual Metabolic Human (VMH) database (Noronha et al., 2019), which provides curated metabolic knowledge and genome-scale

models linking metabolism and microbiome interactions. These collaborative platforms are excellent for the exchange of data, tools or expertise, accelerating scientific discoveries and clinical translation.

Multilayer network approaches for multi-omics integration

Fundamentals of multilayer networks

Multilayer networks are a powerful framework to model complex systems characterized by multiple types of interactions (or relationships). Unlike traditional single-layer networks, which represent a single type of interaction among molecular entities, multilayer networks incorporate multiple interconnected layers, where each of them represents a different type of relationship, capturing the multifaceted nature of real-world interactions (Kivela et al., 2014; De Domenico, 2022). More specifically, a multilayer network is formally defined by a set of nodes and layers, where each node can exist in one or more layers, while edges represent the connections between nodes within the same layer (intra-layer edges) or across different layers (inter-layer edges) (Figure 2b). The interconnections between the different layers (inter-layer edges) are key elements in the network to understand the influence of different types of interactions and the impact on the overall dynamics of the system.

One of the main advantages of multilayer networks, compared to single-layer networks, comes from their ability to model interdependencies between different types of data. In systems biology, a multilayer network can simultaneously represent genomic, transcriptomic, proteomic and clinical data, each as a distinct layer. The interactions within each layer capture relationships specific to that data type, such as transcription factors (TFs) - target genes (TGs) regulatory interactions in the transcriptomic layer or protein-protein interactions (PPI) in the proteomic layer. On the other hand, inter-layer edges capture the relationships between different data types. The construction of multilayer networks involves several methodological steps:

1. Data preprocessing to ensure compatibility: Ensures consistency across omics datasets, involving quality control, noise reduction, and standardization. Batch effect removal and data imputation address technical biases and missing values to accurately reflect biological variability (Leek et al., 2010; Lazar et al., 2013).
2. Single-layer network reconstruction: For each omics type, individual single-layer networks are constructed, based on annotated interactions such as TF-target interactions, PPI, or metabolic pathways. Algorithms tailored to each data type can establish connections indicating correlation, causality, or interaction strength, with edge direction or weight based on interaction type and confidence.
3. Integration into the multilayer network: Single-layer networks are integrated into a multilayer network via inter-layer edges, representing cross-omics interactions, i.e., regulatory relationships between genes and proteins,

or other molecular entities. Several approaches can be used to define inter-layer edges, including data-driven methods that infer relationships based on statistical correlations or ML models that predict interactions based on prior knowledge of biological principles. A special case of multilayer networks is obtained when all nodes are present in all layers, denoted as multiplex networks, where distinct layers represent different interaction types for the same set of nodes (Figure 2b), preserving the individual network topologies (Kivela et al., 2014; Valdeolivas et al., 2019).

The constructed network can be visualized using *ad hoc* developed tools, or extensions of other tools initially built for single-layer network visualization, which we summarize in Table 3.

Analysis of multilayer networks

The advancement of computational tools, algorithms and graph-based methods is essential to analyze multilayer networks and interpret their features. One approach is the use of centrality measures to determine the importance of nodes within the multilayer network. Frequently used centrality measures for single-layer networks (degree, betweenness centrality, alpha centrality, etc.) are extended to consider the multilayer context, offering the opportunity to find nodes that play a major role across multiple layers, or, even further, to identify functional modules (De Domenico and Lancichinetti, 2015). Another essential technique is community detection, an analytical technique which involves node clustering via community identification, by identifying sets of nodes that are densely connected amongst each other within and across layers compared to the connections they have with other nodes not in the community. This approach reveals functional modules of nodes that work together in a coordinated way in several contexts (Mucha et al., 2010; Didier et al., 2015), particularly in complex diseases (Choobdar et al., 2019; Halu et al., 2019).

Additionally, one can examine multilayer motifs: small recurring patterns of interconnections that occur within and between network layers. These motifs establish the fundamental building blocks of the network's structure (Boccaletti et al., 2014). More advanced concepts, such as the Von Neumann entropy of a multilayer network, can be utilized to characterize global connectivity patterns across layers (De Domenico et al., 2015a).

Studying network dynamics is also relevant to understand how information propagates through the multilayer network (a phenomenon called flow) via modeling the spread of perturbations and to investigate how interactions in one layer affect dynamics in other ones (Gómez et al., 2013; De Domenico et al., 2016). One of the most used approaches for this purpose is Random Walk with Restart, which identifies relationships between nodes, reconstructs communities and predicts key interactions (Cowen et al., 2017; Baptista et al., 2024).

Machine learning and predictive modeling

ML techniques are increasingly applied to multilayer networks to develop predictive models. Supervised learning algorithms, like support vector machines (SVM) or random forests (RF), can be trained on multilayer network features to predict clinical outcomes

TABLE 3 Tools used to reconstruct, visualize and analyze multilayer networks.

Tools	Description
MUXViz (De Domenico et al., 2015b)	Visualization of multiplex networks, allowing the exploration of multiple layers of interactions in the same system
Omics Integrator (Tuncbag et al., 2016)	Incorporation of different omics data into coherent networks, optimized for gene, protein and pathway interactions
NetZoo (Ben Guebila et al., 2023)	Collection of methods for inference and analysis of gene regulatory and multi-omic networks (like PANDA, LIONESS, DRAGON), for construction and downstream analysis of complex multilayer biological networks
MultiXrank (Baptista et al., 2022)	Random walk with restart on heterogeneous multilayer networks, used for node prioritization and exploration of information flow across layers
Cytoscape (Shannon et al., 2003); NetworkX (Hagberg et al., 2008); iGraph (Csárdi et al., 2006); Gephi (Bastian et al., 2009)	General-purpose network visualization tools, can be adapted for multilayer network exploration and custom multilayer implementations. Often used for custom multilayer implementations

or response to therapy. Unsupervised learning methods (clustering or dimensionality reduction) can help to discover new patterns or possibly some associations within the integrated data (Wu et al., 2019). Predictive modeling using multilayer networks enhances the ability to make data-driven decisions in personalized cancer treatment.

Network embedding approaches are also effective in analyzing a wide range of networks, even if they were not originally developed for biological networks (Parvizi et al., 2022). These methods have shown good performance in tasks like community detection, node classification and link prediction. However, most embedding methods are not designed to be applied to multiplex networks. To address this, advanced methods like MultiVERSE were developed and shown to have superior performance in tasks such as link prediction and network reconstruction, particularly in multiplex networks (Pio-Lopez et al., 2021).

With their ability to learn complex relationships and inherent structures in multilayer networks, Graph Neural Networks (GNNs) are highly effective at analyzing graph-structured data (Corso et al., 2024). They encode both the node features and the graph topology into low-dimensional embeddings, which can then be used for a variety of tasks, such as node classification and link prediction (Zhang, 2022). Latest advancements, like Graph Convolutional Networks (GCNs) and Graph Attention Networks (GATs) enhance the capacity to represent heterogeneous data and facilitate tasks such as drug target prediction (Grassia et al., 2021; Chatzianastasis et al., 2023; Chen et al., 2024).

Applications of multilayer networks in data integration

Data integration based on multilayer networks is performed by the inference or construction of integrated networks that combine multi-omics data. This all-around approach is ideal to map the interactions between different molecular entities across several biological layers, leading to a multidimensional view of cellular processes. Genomic data can reveal mutations in specific genes, while transcriptomic data can show how these mutations affect gene expression levels. Integrating these datasets into a multilayer network using approaches like HuMMuS (Trimbour et al., 2024) enables the reconstruction of molecular regulatory mechanisms from single-cell multi-omics data by capturing both intra-omics

(peak-peak, gene-gene, TF-TF) and inter-omics (peak-gene, TF-peak) interactions; this could be a way to establish the influence of genetic alterations on transcriptional changes and possibly the propagation of these changes through proteomic and metabolomic layers.

A few diverse multilayer networks have been constructed, for instance integrating disease-perturbed proteins, drug targets and biological functions into a multi-level interactome network for a better definition of disease mechanisms (Ruiz et al., 2021), while similar approaches employing multilayers have also been applied to the study of specific pathologies, such as infection by SARS-CoV-2 (Messina et al., 2020; Verstraete et al., 2020).

To conclude, in cancer research, significant benefits emerge from the application of multilayer network theory, especially the integration of diverse types of biological data that enhances the view of the molecular mechanisms driving tumors. This integrative approach can help to identify key regulatory pathways and potential therapeutic targets that might be overlooked in single-layer analyses. Overall, multilayer networks can help to interpret the complex interplay between genetics and environment (epigenetic modifications, etc.) in cancer evolution.

Digital twins for cancer research: a multilayer network perspective

The conceptual framework for utilizing DTs in oncology through a multilayer network perspective covers a systematic approach that integrates diverse types of biological data to create predictive models of tumors. This framework leverages the positive aspects of multilayer network theory to deliver a more complete representation of the biological processes and molecular functions involved in cancer progression (Laubenbacher et al., 2022). In the following sections, we consider the multilayer network as the scaffold of a patient DT and give a roadmap of how to build it by integrating multi-omics data, as well as how to use it in cancer research.

The integration of multi-omics data is the core of this conceptual framework, combining diverse molecular layers, and eventually complementary patient-level information such as clinical records and host-derived data (blood readouts as in liquid biopsies, metagenomics of the microbiome, exposomics, etc.). Each of

these layers brings unique and complementary insights into the different aspects of cancer.

The final step in the framework corresponds to the clinical implementation and validation of the DT, a challenging but major task for the technology to gain credibility. In this context, two different concepts exist. The first one corresponds to the integration of DTs into clinical workflows, where they can assist oncologists in the decision-making process with data-driven choices regarding diagnosis, prognosis or treatment (clinical decision support systems). The second one is based on the use of DTs to perform virtual clinical trials, testing the efficacy of new compounds or even drug associations in a virtual environment before applying them to patients to limit the risk (Wang et al., 2024). However, the validation of DTs remains a significant challenge. Retrospective validation can be done using held-out patient cohorts, whereas rigorous strategies in cross-validation, including nested cross-validation, should be used to reduce the risk of overfitting in a high-dimensional multi-omics environment. External validation across independent cohorts is essential in the overall process. Furthermore, the methodology is challenged by the potential issue of circularity in using the same omics data in network construction and downstream predictive modeling. Validation of DTs' predictive capabilities is indispensable to ascertain their reliability. Some retrospective analysis using historical patient data, as well as prospective validation in clinical trials like treatment timing optimization can be performed. Continuous monitoring and updating of DTs with dynamic, up-to-date patient data would reinforce the accuracy and confirm the effectiveness of these tools for cancer management (Hernandez-Boussard et al., 2021).

Practical challenges in DT implementation

Research gaps

The scarce availability of longitudinal data to capture the dynamic nature of tumor progression is a significant challenge. This causes a limited understanding of temporal dynamics in cancer progression or even treatment response. Continuous monitoring and data acquisition in real-time from patients involves the integration of clinical, imaging, liquid biopsy or wearable device data, as well as potentially patient-reported outcomes, with multi-omics profiles (Chow et al., 2024). Achieving this level of integration requires advanced data management systems that are capable of handling diverse data flows, while also maintaining temporal consistency, such as those employed in personalized nutrition applications (Romero-Tapiador et al., 2026).

A very important issue that should not be neglected is the use of sensitive personal health information because it raises important ethical and legal concerns related to privacy, consent, and data ownership. Among them, protecting patient data from unauthorized access is critical as data breaches can lead to potentially severe consequences, like identity theft or loss of trust in healthcare systems (Rothstein, 2015; Thapa and Camtepe, 2021). Data privacy should be made a priority by implementing stringent data governance policies to protect patient information. Techniques such as differential privacy, which add noise to datasets to prevent the identification of individuals, are increasingly employed to protect patient confidentiality while allowing data analysis (Dwork and Roth,

2014). Recently, promising solutions have been suggested, such as blockchain, which offers a decentralized and immutable ledger that can be used to securely store/share patient data (Pournaghi et al., 2020), or homomorphic encryption, allowing computations on encrypted data without needing to decrypt it first (Munjal and Bhatia, 2023). Security measures must address the risk of cyber-attacks with the implementation of robust cybersecurity protocols (intrusion detection systems, secure access controls, etc.), which is essential to protect multi-omics data and DT models from hostile attacks, and must also comply with evolving regulatory frameworks governing the use of health data. In Europe, the General Data Protection Regulation (GDPR) and the EU AI Act impose requirements on data privacy, transparency and accountability, particularly for high-risk AI systems in healthcare. In the US, the Health Insurance Portability and Accountability Act (HIPAA) governs the use of protected health information, while guidance from the Food and Drug Administration (FDA) applies to AI/ML-based software as a medical device. These regulatory considerations can represent important challenges for the clinical deployment of DTs.

The biological interpretation of integrated multi-omics data is another research gap. While multilayer networks can reveal complex interactions, translating these findings into concrete biological insights remains a challenge. There is a need for new tools to interpret the results of multilayer network analyses in a technically and biologically meaningful way. In addition, experimental validation of the predictions made by DTs is necessary to confirm different parameters, like accuracy or utility in clinical settings.

Data collection and AI

The adoption of DTs in precision medicine is increasingly operational, driven by extensive data collection alongside traditional biomedical methodologies. However, the reliance on black-box predictive models based on large datasets presents limitations that could potentially slow down the broader application of DTs in clinical settings. In the literature, it has been argued that hypothesis-driven generative models, which generate data based on a set of underlying assumptions of the processes that generate the observed data, more particularly multi-scale modeling, are essential to boost the clinical accuracy of DTs. The transformative potential of DTs in healthcare has been explored by emphasizing their capability to simulate complex interdependent biological processes across multiple scales. The integration of generative models with extensive datasets can deliver scenario-based approaches to explore diverse therapeutic strategies. It is an excellent strategy to support dynamic clinical decision-making. This method not only leverages advancements in data science to improve disease management, but it also incorporates insights from complex systems, quantitative biology and digital medicine to improve patient care (De Domenico et al., 2025).

Perspectives

In the context of cancer, multilayer networks have been used to capture complex interactions, leading to the identification of

novel therapeutic targets (Liu et al., 2020). The integration of gene expression data with chromatin accessibility and protein-protein interaction networks was shown to uncover key regulatory proteins that are central to tumor progression, which might not be clearly identified when examining single-layer networks, highlighting the added value of a multilayer approach to propose new biomarkers.

Given the importance of the tumor microenvironment (TME) in affecting tumor progression and therapy response (Quail and Joyce, 2013), it is imperative to explore the integration of information regarding all cellular populations within a tumor sample into predictive models. The TME refers to the complex ecosystem surrounding a tumor composed of immune cells, fibroblasts, endothelial cells, extracellular matrix, blood vessels, and signaling molecules, all of which influence tumor growth. This process requires the non-trivial development of new methods that capture salient information from either estimation of cell type proportions from bulk RNA-seq (Merotto et al., 2024), or from single-cell spatial (Visium/Xenium) (Stahl et al., 2016) and non-spatial datasets (10x Genomics scRNA-seq/scATAC-seq) (Zheng et al., 2017). It is likely that new modes of dimensionality reduction will have to be applied to these types of data to produce correspondences between these single-cell level descriptions and more traditional layers, in which nodes are molecular entities. In this context, node identity shifts from molecular entities to cells or cell types, and multilayer networks can incorporate additional layers capturing cell-cell communication using methods such as NicheNet, CellChat, LIANA (Sang-aram et al., 2025; Jin et al., 2025; Dimitrov et al., 2024). Recent tools for gene regulatory inference (SCENIC+, ArchR) (Bravo González-Blas et al., 2023; Granja et al., 2021), multimodal integration (Seurat WNN), and spatial analysis (Squidpy) (Palla et al., 2022) further illustrate these developments.

Multi-level and hybrid models will become necessary to use the multilayer structures beyond simple data integration and towards dynamic modeling (Pancaldi, 2026). Despite the obvious interest in analyzing the integrated datasets on this multilayer structure, looking for communities spanning several layers, for example, the real power of this framework will come from the construction of dynamic models of these complex systems. A first approximation of dynamics can be provided by flow analyses on the network, which simulate how the information will spread within and across layers, providing an idea of perturbations that can move the DT towards a target state (for instance, towards a healthy state). Simulating the dynamic state of a patient, including the tumor itself, the immune system, different organs and how the patient environment is impacting all of them will probably require a combination of models of different scales (e.g., molecular, metabolic, cellular, etc.) as well as a better understanding of regulation and correspondence between nodes in different layers. Further down the line, with a better understanding of the underlying multi-level biological processes, we might be able to construct executable models comprising the entire variety of datasets, constituting DTs that will be extremely useful for *in silico* clinical trials that consider the patient as a whole (Clarke and Fisher, 2020; Niarakis et al., 2024).

Several approaches to modeling the TME have been proposed (Metzcar et al., 2019; Johnson et al., 2025), and these might enable a dynamic perspective, which is essential to understand the

adaptation of tumors to therapeutic molecules. More effective treatment strategies might require combinations of different approaches at different times, requiring a good understanding of the time scales of the different biological processes involved in cancer (Wang and Deisboeck, 2019).

To move closer to virtual clinical trials, a more complete representation of patients would be required. The flexibility of the multilayer approach should enable the integration of diverse data types from the tumor and from the patient as a whole. For instance, integrating prior comorbidities could be of great value in predicting patient outcomes, as was the case for predicting pancreatic cancer onset (Placido et al., 2023). Indeed, patient comorbidity profiles might work as proxies for identifying patient endotypes or characteristics that are relevant for patient stratification (Sánchez-Valle et al., 2020), while integrating exposures or side effects with tumor characteristics via multilayer networks and multi-level models could produce more useful DTs for drug discovery and development of *in silico* clinical trials.

Conclusion

This review has examined the methodologies, tools, and challenges in the implementation of multilayer networks as cornerstones of DTs for cancer, building on their potential to provide a comprehensive and integrated view of the patient, a central element required to predict disease progression, response to treatment, and side effects. Despite the hopes in DTs, some challenges remain, specifically in data integration, computational efficiency, and data privacy. Future research should focus on developing computational models and enhancing scalability, without neglecting ethical frameworks for patient data use. Collaboration between the private sector, clinicians and researchers could help overcome these challenges. DTs offer opportunities to transform cancer care with more effective and individualized treatments by providing dynamic, multiscale models that simulate tumor biology and patient trajectories over time. Further advancements in this field, with the integration of more patient-level data, are crucial, enabling better *in silico* clinical trials and accelerating cancer treatment development.

Author contributions

HC: Conceptualization, Writing – original draft, Writing – review and editing. MM: Writing – review and editing. TJ: Writing – review and editing. AZ: Supervision, Writing – review and editing. VP: Supervision, Writing – review and editing.

Funding

The author(s) declared that financial support was received for this work and/or its publication. This work was funded by the Agence Nationale de la Recherche (ANR-23-CE45-0014), granted to VP. HC was funded by a CIFRE fellowship (No. 231967A10).

established between VP, AZ, and HC, in collaboration with Inserm Transfert, CNRS, Université de Toulouse, the CHU de Toulouse, and EVOTEC (France) SAS. This study has been partially supported through the grant EUR CARE No. ANR-18-EURE-0003 in the framework of the Programme des Investissements d'Avenir, granted to HC. The authors declared that this work received funding from EVOTEC (France) SAS through the CIFRE fellowship.

Acknowledgements

The authors would like to thank Joel Herrera Martinez and four reviewers for their constructive feedback on previous versions of this review.

Conflict of interest

Authors HC, TJ, and AZ are employed by Evotec. AZ reports receiving consulting fees from Scientia Lab outside the submitted work.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

References

- Acharya, D., and Mukhopadhyay, A. (2024). A comprehensive review of machine learning techniques for multi-omics data integration: challenges and applications in precision oncology. *Briefings Funct. Genomics* 23 (5), 549–560. doi:10.1093/bfgp/ela013
- Angermueller, C., Pärnamaa, T., Parts, L., and Stegle, O. (2016). Deep learning for computational biology. *Mol. Syst. Biol.* 12 (7), 878. doi:10.15252/msb.20156651
- Argelaguet, R., Velten, B., Arnol, D., Dietrich, S., Zenz, T., Marioni, J. C., et al. (2018). Multi-omics factor Analysis—a framework for unsupervised integration of multi-omics data sets. *Mol. Syst. Biol.* 14 (6), MSB178124. doi:10.15252/msb.20178124
- Argelaguet, R., Arnol, D., Bredikhin, D., Deloro, Y., Velten, B., Marioni, J. C., et al. (2020). MOFA+: a statistical framework for comprehensive integration of multi-modal single-cell data. *Genome Biol.* 21 (1), 111. doi:10.1186/s13059-020-02015-1
- Asghar, U. S., and Chung, C. (2025). Application of digital twins for personalized oncology. *Nat. Rev. Cancer* 25 (11), 823–825. doi:10.1038/s41568-025-00850-7
- Bahrami, M., Tejada-Lapuerta, A., Becker, S., Hashemi G., F. S., and Theis, F. J. (2025). scConcept: contrastive pretraining for technology-agnostic single-cell representations beyond reconstruction. *bioRxiv*. doi:10.1101/2025.10.14.682419
- Ballard, J. L., Wang, Z., Li, W., Shen, L., and Long, Q. (2024). Deep learning-based approaches for multi-omics data integration and analysis. *BioData Min.* 17, 38. doi:10.1186/s13040-024-00391-z
- Baptista, A., Gonzalez, A., and Baudot, A. (2022). Universal multilayer network exploration by random walk with restart. *Commun. Phys.* 5 (1), 170. doi:10.1038/s42005-022-00937-9
- Baptista, A., Brière, G., and Baudot, A. (2024). Random walk with restart on multilayer networks: from node prioritisation to supervised link prediction and beyond. *BMC Bioinforma.* 25 (1), 70. doi:10.1186/s12859-024-05683-z
- Bastian, M., Heymann, S., and Jacomy, M. (2009). Gephi: an open source software for exploring and manipulating networks. *Proc. Int. AAAI Conf. Web Soc. Media* 3 (1), 361–362. doi:10.1609/icwsm.v3i1.13937
- Ben Guebila, M., Wang, T., Lopes-Ramos, C. M., Fanfani, V., Weighill, D., Burkholz, R., et al. (2023). The network zoo: a multilingual package for the inference and analysis of gene regulatory networks. *Genome Biol.* 24 (1), 45. doi:10.1186/s13059-023-02877-1
- Benzekry, S., Lamont, C., Beheshti, A., Tracz, A., Ebos, J. M. L., Hlatky, L., et al. (2014). Classical mathematical models for description and prediction of experimental tumor growth. *PLoS Comput. Biol.* 10 (8), e1003800. doi:10.1371/journal.pcbi.1003800
- Bersanelli, M., Mosca, E., Remondini, D., Giampieri, E., Sala, C., Castellani, G., et al. (2016). Methods for the integration of multi-omics data: mathematical aspects. *BMC Bioinforma.* 17 (2), S15. doi:10.1186/s12859-015-0857-9
- Boccaletti, S., Bianconi, G., Criado, R., Del Genio, C. I., Gómez-Gardeñes, J., Romance, M., et al. (2014). The structure and dynamics of multilayer networks. *Phys. Rep.* 544 (1), 1–122. doi:10.1016/j.physrep.2014.07.001
- Boyd, L. N. C., Ali, M., Comandatore, A., Garajova, I., Kam, L., Puik, J. R., et al. (2023). Prediction model for early-stage pancreatic cancer using routinely measured blood biomarkers. *JAMA Netw. Open* 6 (8), e2331197. doi:10.1001/jamanetworkopen.2023.31197
- Bravo González-Blas, C., De Winter, S., Hulselmans, G., Hecker, N., Matetovici, I., Christiaens, V., et al. (2023). SCENIC+: single-cell multiomic inference of enhancers and gene regulatory networks. *Nat. Methods* 20 (9), 1355–1367. doi:10.1038/s41592-023-01938-4
- Bruynseels, K., Santoni de Sio, F., and van den Hoven, J. (2018). Digital twins in health care: ethical implications of an emerging engineering paradigm. *Front. Genet.* 9, 31. doi:10.3389/fgene.2018.00031
- Cai, Z., Poulos, R. C., Liu, J., and Zhong, Q. (2022). Machine learning for multi-omics data integration in cancer. *iScience* 25 (2), 103798. doi:10.1016/j.isci.2022.103798
- Cantini, L., Zakeri, P., Hernandez, C., Naldi, A., Thieffry, D., Remy, E., et al. (2021). Benchmarking joint multi-omics dimensionality reduction approaches for the study of cancer. *Nat. Commun.* 12 (1), 124. doi:10.1038/s41467-020-20430-7
- Cao, K., Hong, Y., and Wan, L. (2021). Manifold alignment for heterogeneous single-cell multi-omics data integration using pamon. *Bioinformatics* 38 (1), 211–219. doi:10.1093/bioinformatics/btab594
- Chatzianastasis, M., Vazirgiannis, M., and Zhang, Z. (2023). Explainable multilayer graph neural network for cancer gene prediction. *Bioinformatics* 39 (11), btad643. doi:10.1093/bioinformatics/btad643
- Chen, H., Ryu, J., Vinyard, M. E., Lerer, A., and Pinello, L. (2024). SIMBA: single-cell embedding along with features. *Nat. Methods* 21 (6), 1003–1013. doi:10.1038/s41592-023-01899-8
- Choobdar, S., Ahsen, M. E., Crawford, J., Tomasoni, M., Fang, T., Lamparter, D., et al. (2019). Assessment of network module identification across complex diseases. *Nat. Methods* 16 (9), 843–852. doi:10.1038/s41592-019-0509-5
- Chow, R., Drkulec, H., Im, J. H. B., Tsai, J., Nafees, A., Kumar, S., et al. (2024). The use of wearable devices in oncology patients: a systematic review. *Oncol.* 29 (4), e419–e430. doi:10.1093/oncol/oyad305
- Cirillo, D., Núñez-Carpintero, I., and Valencia, A. (2021). Artificial intelligence in cancer research: learning at different levels of data granularity. *Mol. Oncol.* 15 (4), 817–829. doi:10.1002/1878-0261.12920

Generative AI statement

The author(s) declared that generative AI was not used in the creation of this manuscript.

Any alternative text (alt text) provided alongside figures in this article has been generated by Frontiers with the support of artificial intelligence and reasonable efforts have been made to ensure accuracy, including review by the authors wherever possible. If you identify any issues, please contact us.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

- Clarke, M. A., and Fisher, J. (2020). Executable cancer models: successes and challenges. *Nat. Rev. Cancer* 20 (6), 343–354. doi:10.1038/s41568-020-0258-x
- Coorey, G., Figtree, G. A., Fletcher, D. F., Snelson, V. J., Vernon, S. T., Winlaw, D., et al. (2022). The health digital twin to tackle cardiovascular disease—a review of an emerging interdisciplinary field. *Npj Digit. Med.* 5 (1), 1–12. doi:10.1038/s41746-022-00640-7
- Corso, G., Stark, H., Jegelka, S., Jaakkola, T., and Barzilay, R. (2024). Graph neural networks. *Nat. Rev. Methods Prim.* 4 (1), 1–13. doi:10.1038/s43586-024-00294-7
- Cowen, L., Ideker, T., Raphael, B. J., and Sharan, R. (2017). Network propagation: a universal amplifier of genetic associations. *Nat. Rev. Genet.* 18 (9), 551–562. doi:10.1038/nrg.2017.38
- Csárdi, G., Nepusz, T., Traag, V., Horvát, S., Zanini, F., Noom, D., et al. (2006). *Igraph: network analysis and visualization*. doi:10.32614/CRAN.package.igraph
- Cui, H., Wang, C., Maan, H., Pang, K., Luo, F., Duan, N., et al. (2024). scGPT: toward building a foundation model for single-cell multi-omics using generative AI. *Nat. Methods* 21 (8), 1470–1480. doi:10.1038/s41592-024-02201-0
- De Domenico, M. (2022). “Multilayer networks: overview,” in *Multilayer networks: analysis and visualization*. Editor M. De Domenico (Cham: Springer International Publishing), 23–29. doi:10.1007/978-3-030-75718-2_2
- De Domenico, M. (2023). More is different in real-world multilayer networks. *Nat. Phys.* 19 (9), 1247–1262. doi:10.1038/s41567-023-02132-1
- De Domenico, M., Lancichinetti, A., Arenas, A., and Rosvall, M. (2015). Identifying modular flows on multilayer networks reveals highly overlapping organization in interconnected systems. *Phys. Rev. X* 5 (1), 011027. doi:10.1103/PhysRevX.5.011027
- De Domenico, M., Nicosia, V., Arenas, A., and Latora, V. (2015a). Structural reducibility of multilayer networks. *Nat. Commun.* 6 (1), 6864. doi:10.1038/ncomms7864
- De Domenico, M., Porter, M. A., and Arenas, A. (2015b). MuxViz: a tool for multilayer analysis and visualization of networks. *J. Complex Netw.* 3 (2), 159–176. doi:10.1093/comnet/cnu038
- De Domenico, M., Granell, C., Porter, M. A., and Arenas, A. (2016). The physics of spreading processes in multilayer networks. *Nat. Phys.* 12 (10), 901–906. doi:10.1038/nphys3865
- De Domenico, M., Allegri, L., Caldarelli, G., d’Andrea, V., Di Camillo, B., Rocha, L. M., et al. (2025). Challenges and opportunities for digital twins in precision medicine from a complex systems perspective. *npj Digital Medicine* 8, 37. doi:10.1038/s41746-024-01402-3
- Demirel, H. C., Arici, M. K., and Tuncbag, N. (2022). Computational approaches leveraging integrated connections of multi-omic data toward clinical applications. *Mol. Omics* 18 (1), 7–18. doi:10.1039/D1MO00158B
- Didier, G., Brun, C., and Baudot, A. (2015). Identifying communities from multiplex biological networks. *PeerJ* 3, e1525. doi:10.7717/peerj.1525
- Dimitrov, D., Schäfer, P. S. L., Farr, E., Rodriguez-Mier, P., Lobentanzer, S., Badia-I-Mompel, P., et al. (2024). LIANA+ provides an all-in-one framework for cell–cell communication inference. *Nat. Cell Biol.* 26 (9), 1613–1622. doi:10.1038/s41556-024-01469-w
- Dwork, C., and Roth, A. (2014). The algorithmic foundations of differential privacy. *Found. Trends® Theor. Comput. Sci.* 9 (3–4), 211–407. doi:10.1561/04000000042
- Fuller, A., Fan, Z., Day, C., and Barlow, C. (2020). Digital twin: enabling technologies, challenges and open research. *IEEE Access* 8, 108952–108971. doi:10.1109/ACCESS.2020.2998358
- Gayoso, A., Steier, Z., Lopez, R., Regier, J., Nazor, K. L., Streets, A., et al. (2021). Joint probabilistic modeling of single-cell multi-omic data with totalVI. *Nat. Methods* 18 (3), 272–282. doi:10.1038/s41592-020-01050-x
- Gegner, H. M., Naake, T., Aljakouch, K., Dugourd, A., Kliever, G., Müller, T., et al. (2023). A single-sample workflow for joint metabolomic and proteomic analysis of clinical specimens. *bioRxiv*. doi:10.1101/2023.11.07.561857
- Gkouskou, K., Vlastos, I., Karkaloulos, P., Chaniotis, D., Sanoudou, D., and Eliopoulos, A. G. (2020). The ‘Virtual Digital Twins’ concept in precision nutrition. *Adv. Nutr.* 11 (6), 1405–1413. doi:10.1093/advances/nmaa089
- Gómez, D., Figueira, J. R., and Eusébio, A. (2013). Modeling centrality measures in social network analysis using bi-criteria network flow optimization problems. *Eur. J. Operational Res.* 226 (2), 354–365. doi:10.1016/j.ejor.2012.11.027
- Gómez-López, G., Dopazo, J., Cigudosa, J. C., Valencia, A., and Al-Shahrour, F. (2019). Precision medicine needs pioneering clinical bioinformaticians. *Briefings Bioinforma.* 20 (3), 752–766. doi:10.1093/bib/bbx144
- Granja, J. M., Corces, M. R., Pierce, S. E., Bagdatli, S. T., Choudhry, H., Chang, H. Y., et al. (2021). ArchR is a scalable software package for integrative single-cell chromatin accessibility analysis. *Nat. Genet.* 53 (3), 403–411. doi:10.1038/s41588-021-00790-6
- Grassia, M., De Domenico, M., and Mangioni, G. (2021). mGNN: generalizing the graph neural networks to the multilayer case. *arXiv*. doi:10.48550/arXiv.2109.10119
- Hagberg, A., Swart, P. J., and Schult, D. A. (2008). *Exploring network structure, dynamics, and function using NetworkX*. LA-UR-08-05495; LA-UR-08-5495. Los Alamos, NM (United States): Los Alamos National Laboratory LANL. Available online at: <https://www.osti.gov/biblio/960616> (Accessed September 26, 2024).
- Halu, A., De Domenico, M., Arenas, A., and Sharma, A. (2019). The multiplex network of human diseases. *Npj Syst. Biol. Appl.* 5 (1), 1–12. doi:10.1038/s41540-019-0092-5
- Hasin, Y., Seldin, M., and Lusis, A. (2017). Multi-omics approaches to disease. *Genome Biol.* 18 (1), 83. doi:10.1186/s13059-017-1215-1
- Hauptmann, T., and Kramer, S. (2023). A fair experimental comparison of neural network architectures for latent representations of multi-omics for drug response prediction. *BMC Bioinforma.* 24 (1), 45. doi:10.1186/s12859-023-05166-7
- Hernandez-Boussard, T., Macklin, P., Greenspan, E. J., Gryshuk, A. L., Stahlberg, E., Syeda-Mahmood, T., et al. (2021). Digital twins for predictive oncology will be a paradigm shift for precision cancer care. *Nat. Medicine* 27 (12), 2065–2066. doi:10.1038/s41591-021-01558-5
- Hodgkinson, A., Trucu, D., Lacroix, M., Le Cam, L., and Radulescu, O. (2022). Computational model of heterogeneity in melanoma: designing therapies and predicting outcomes. *Front. Oncol.* 12, 857572. doi:10.3389/fonc.2022.857572
- Hsu, C.-Y., Askar, S., Alshkarchy, S. S., Nayak, P. P., Attabi, K. A. L., Khan, M. A., et al. (2025). AI-driven multi-omics integration in precision oncology: bridging the data deluge to clinical decisions. *Clin. Exp. Med.* 26 (1), 29. doi:10.1007/s10238-025-01965-9
- Huang, S., Chaudhary, K., and Garmire, L. X. (2017). More is better: recent progress in multi-omics data integration methods. *Front. Genet.* 8, 84. doi:10.3389/fgene.2017.00084
- Huizing, G.-J., Deutschmann, I. M., Peyré, G., and Cantini, L. (2023). Paired single-cell multi-omics data integration with mowgli. *Nat. Commun.* 14 (1), 7711. doi:10.1038/s41467-023-43019-2
- Jin, S., Plikus, M. V., and Nie, Q. (2025). CellChat for systematic analysis of cell–cell communication from single-cell transcriptomics. *Nat. Protoc.* 20 (1), 180–219. doi:10.1038/s41596-024-01045-4
- Johnson, J. A. I., Bergman, D. R., Rocha, H. L., Zhou, D. L., Cramer, E., Mclean, I. C., et al. (2025). Human interpretable grammar encodes multicellular systems biology models to democratize virtual cell laboratories. *Cell* 188 (17), 4711–4733. doi:10.1016/j.cell.2025.06.048
- Junet, V., Matos-Filipe, P., García-Illarramendi, J. M., Ramírez, E., Oliva, B., Farrés, J., et al. (2023). A decision support system based on artificial intelligence and systems biology for the simulation of pancreatic cancer patient status. *CPT Pharmacometrics & Syst. Pharmacol.* 12 (7), 916–928. doi:10.1002/psp4.12961
- Kalfon, J., Samaran, J., Peyré, G., and Cantini, L. (2025a). scPRINT: pre-training on 50 million cells allows robust gene network predictions. *Nat. Commun.* 16 (1), 3607. doi:10.1038/s41467-025-58699-1
- Kalfon, J., Cantini, L., and Peyre, G. (2025b). Towards foundation models that learn across biological scales. *bioRxiv*. doi:10.1101/2025.05.16.653447
- Kamel Boulos, M. N., and Zhang, P. (2021). Digital twins: from personalised medicine to precision public health. *J. Personalized Med.* 11 (8), 745. doi:10.3390/jpm11080745
- Kim, H., Golub, G. H., and Park, H. (2005). Missing value estimation for DNA microarray gene expression data: local least squares imputation. *Bioinformatics* 21 (2), 187–198. doi:10.1093/bioinformatics/bth499
- Kivela, M., Arenas, A., Barthelemy, M., Gleeson, J. P., Moreno, Y., and Porter, M. A. (2014). Multilayer networks. *J. Complex Netw.* 2 (3), 203–271. doi:10.1093/comnet/cnu016
- Lasko, T. A., Stead, W. W., Still, J. M., Li, T. Z., Kammer, M., Barbero-Mota, M., et al. (2025). Unsupervised discovery of clinical disease signatures using probabilistic independence. *J. Biomed. Inf.* 166, 104837. doi:10.1016/j.jbi.2025.104837
- Laubenbacher, R., Niarakis, A., Helikar, T., An, G., Shapiro, B., Malik-Sheriff, R. S., et al. (2022). Building digital twins of the human immune system: toward a roadmap. *Npj Digit. Med.* 5 (1), 1–5. doi:10.1038/s41746-022-00610-z
- Lazar, C., Meganck, S., Taminiau, J., Steenhoff, D., Coletta, A., Molter, C., et al. (2013). Batch effect removal methods for microarray gene expression data integration: a survey. *Briefings Bioinforma.* 14 (4), 469–490. doi:10.1093/bib/bbs037
- LeCun, Y. (2022). *A path towards autonomous machine intelligence*.
- Lee, B., Zhang, S., Poleksic, A., and Xie, L. (2020). Heterogeneous multi-layered network model for omics data integration and analysis. *Front. Genet.* 10, 1381. doi:10.3389/fgene.2019.01381
- Leek, J. T., Scharpf, R. B., Bravo, H. C., Simcha, D., Langmead, B., Johnson, W. E., et al. (2010). Tackling the widespread and critical impact of batch effects in high-throughput data. *Nat. Rev. Genet.* 11 (10), 733–739. doi:10.1038/nrg2825
- Libbrecht, M. W., and Noble, W. S. (2015). Machine learning applications in genetics and genomics. *Nat. Rev. Genet.* 16 (6), 321–332. doi:10.1038/nrg3920
- Liu, Y., Shi, N., Regev, A., He, S., and Hemann, M. T. (2020). Integrated regulatory models for inference of subtype-specific susceptibilities in glioblastoma. *Mol. Syst. Biol.* 16 (9), e9506. doi:10.15252/msb.20209506
- Lococo, F., Boldrini, L., Dieprie, C. D., Evangelista, J., Nero, C., Flamini, S., et al. (2023). Lung cancer multi-omics digital human avatars for integrating precision medicine into clinical practice: the LANTERN study. *BMC Cancer* 23 (1), 540. doi:10.1186/s12885-023-10997-x
- Lopez, R., Regier, J., Cole, M. B., Jordan, M. I., and Yosef, N. (2018). Deep generative modeling for single-cell transcriptomics. *Nat. Methods* 15 (12), 1053–1058. doi:10.1038/s41592-018-0229-2

- Mardinoglu, A., Turkez, H., Shong, M., Srinivasulu, V. P., Nielsen, J., Palsson, B. O., et al. (2025). Longitudinal big biological data in the AI era. *Mol. Syst. Biol.* 21 (9), 1147–1165. doi:10.1038/s44320-025-00134-0
- Merotto, L., Vairo, F., and Macklin, P. (2024). “Chapter two - next-generation deconvolution of transcriptomic data to investigate the tumor microenvironment,” in *International review of cell and molecular biology*. Editors A. D. Garg, and L. Galluzzi (Academic Press), 103–143. doi:10.1016/bs.ircmb.2023.05.002
- Messina, F., Giombini, E., Agrati, C., Vairo, F., Ascoli Bartoli, T., Al Moghazi, S., et al. (2020). COVID-19: viral-host interactome analyzed by network based-approach model to study pathogenesis of SARS-CoV-2 infection. *J. Transl. Med.* 18 (1), 233. doi:10.1186/s12967-020-02405-w
- Metzcar, J., Wang, Y., Heiland, R., and Macklin, P. (2019). A review of cell-based computational modeling in cancer biology. *JCO Clin. Cancer Inf.* 3 (3), 1–13. doi:10.1200/CCI.18.00069
- Mucha, P. J., Richardson, T., Macon, K., Porter, M. A., and Onnela, J. P. (2010). Community structure in time-dependent, multiscale, and multiplex networks. *Science* 328 (5980), 876–878. doi:10.1126/science.1184819
- Munjal, K., and Bhatia, R. (2023). A systematic review of homomorphic encryption and its contributions in healthcare industry. *Complex & Intelligent Syst.* 9 (4), 3759–3786. doi:10.1007/s40747-022-00756-z
- Niarakis, A., Laubenbacher, R., An, G., Ilan, Y., Fisher, J., Flobak, Å., et al. (2024). Immune digital twins for complex human pathologies: applications, limitations, and challenges. *Syst. Biol. Appl.* 10 (1), 141. doi:10.1038/s41540-024-00450-5
- Nikolski, M., Hovig, E., Al-Shahrour, F., Blomberg, N., Scollen, S., Valencia, A., et al. (2024). Roadmap for a European cancer data management and precision medicine infrastructure. *Nat. Cancer* 5 (3), 367–372. doi:10.1038/s43018-023-00717-6
- Noronha, A., Modamio, J., Jarosz, Y., Guerard, E., Sompairac, N., Preciat, G., et al. (2019). The virtual metabolic human database: integrating human and gut microbiome metabolism with nutrition and disease. *Nucleic Acids Res.* 47 (D1), D614–D624. doi:10.1093/nar/gky992
- Palla, G., Spitzer, H., Klein, M., Fischer, D., Schaar, A. C., Kuemmerle, L. B., et al. (2022). Squidpy: a scalable framework for spatial omics analysis. *Nat. Methods* 19 (2), 171–178. doi:10.1038/s41592-021-01358-2
- Pancaldi, V. (2026). Network analysis to understand gene regulation: across spatial scales and toward multi-level models. *Curr. Opin. Genet. Dev.* 98, 102473. doi:10.1016/j.cdev.2026.102473
- Parvizi, P., Azuaje, F., Theodoratou, E., and Luz, S. (2022). BioNE: integration of network embeddings for supervised learning. *bioRxiv*. doi:10.1101/2022.04.26.489560
- Pellegrino, G., Gervasi, M., Angelelli, M., and Corallo, A. (2025). A conceptual framework for digital twin in healthcare: evidence from a systematic meta-review. *Inf. Syst. Front.* 27 (1), 7–32. doi:10.1007/s10796-024-10536-4
- Perlstein, D., Shlagman, O., Kogan, Y., Halevi-Tobias, K., Jakobson, A., Lazarev, I., et al. (2019). Personal response to immune checkpoint inhibitors of patients with advanced melanoma explained by a computational model of cellular immunity, tumor growth, and drug. *PLOS ONE* 14 (12), e0226869. doi:10.1371/journal.pone.0226869
- Picard, M., Scott-Boyer, M. P., Bodein, A., Périn, O., and Droit, A. (2021). Integration strategies of multi-omics data for machine learning analysis. *Comput. Struct. Biotechnol. J.* 19, 3735–3746. doi:10.1016/j.csbj.2021.06.030
- Pio-Lopez, L., Valdeolivas, A., Tichit, L., Remy, É., and Baudot, A. (2021). MultiVERSE: a multiplex and multiplex-heterogeneous network embedding approach. *Sci. Rep.* 11 (1), 8794. doi:10.1038/s41598-021-87987-1
- Placido, D., Yuan, B., Hjaltelin, J. X., Zheng, C., Haue, A. D., Chmura, P. J., et al. (2023). A deep learning algorithm to predict risk of pancreatic cancer from disease trajectories. *Nat. Med.* 29 (5), 1113–1122. doi:10.1038/s41591-023-02332-5
- Pontén, F., Jirstrom, K., and Uhlen, M. (2008). The human protein Atlas—A tool for pathology. *J. Pathology* 216 (4), 387–393. doi:10.1002/path.2440
- Pournaghi, S. M., Bayat, M., and Farjami, Y. (2020). MedSBA: a novel and secure scheme to share medical data based on blockchain technology and attribute-based encryption. *J. Ambient Intell. Humaniz. Comput.* 11 (11), 4613–4641. doi:10.1007/s12652-020-01710-y
- Qi, Q., Tao, F., Hu, T., Anwer, N., Liu, A., Wei, Y., et al. (2021). Enabling technologies and tools for digital twin. *J. Manuf. Syst.* 58, 3–21. doi:10.1016/j.jmsy.2019.10.001
- Quail, D. F., and Joyce, J. A. (2013). Microenvironmental regulation of tumor progression and metastasis. *Nat. Med.* 19 (11), 1423–1437. doi:10.1038/nm.3394
- Ren, L., Wang, J., Li, Z., Li, Q., and Yu, G. (2023). scMCs: a framework for single-cell multi-omics data integration and multiple clusterings. *Bioinformatics* 39 (4), 133. doi:10.1093/bioinformatics/btad133
- Rockne, R. C., Andersen, M., Anderson, A. R. A., Basanta, D., Bentinegna, A., Benzekry, S., et al. (2026). The future of mathematical oncology in the age of AI. *Syst. Biol. Appl.* 12 (1), 22. doi:10.1038/s41540-026-00656-9
- Rohart, F., Gautier, B., Singh, A., and Lê Cao, K. A. (2017). mixOmics: an R package for ‘omics feature selection and multiple data integration. *PLOS Comput. Biol.* 13 (11), e1005752. doi:10.1371/journal.pcbi.1005752
- Romero-Tapiador, S., Tolosana, R., Morales, A., Lacruz-Pleguezuelos, B., Bosch Pastor, S., Marcos-Zambrano, L. J., et al. (2026). Personalized weight loss management through wearable devices and artificial intelligence. *Comput. Biol. Med.* 209, 111676. doi:10.1016/j.combiomed.2026.111676
- Rothstein, M. A. (2015). Ethical issues in big data health research. *J. Law, Med. & Ethics* 43 (2), 425–429. doi:10.1111/jlme.12258
- Rozenblatt-Rosen, O., Stubbington, M. J. T., Regev, A., and Teichmann, S. A. (2017). The human cell atlas: from vision to reality. *Nature* 550 (7677), 451–453. doi:10.1038/550451a
- Ruiz, C., Zitnik, M., and Leskovec, J. (2021). Identification of disease treatment mechanisms through the multiscale interactome. *Nat. Commun.* 12 (1), 1796. doi:10.1038/s41467-021-21770-8
- Sánchez-Valle, J., Tejero, H., Fernández, J. M., Juan, D., Urda-García, B., Capella-Gutiérrez, S., et al. (2020). Interpreting molecular similarity between patients as a determinant of disease comorbidity relationships. *Nat. Commun.* 11 (1), 2854. doi:10.1038/s41467-020-16540-x
- Sang-aram, C., Browaeys, R., Seurinck, R., and Saeys, Y. (2025). Unraveling cell-cell communication with NicheNet by inferring active ligands from transcriptomics data. *Nat. Protoc.* 20 (6), 1439–1467. doi:10.1038/s41596-024-01121-9
- Santiago-Rodriguez, T. M., and Hollister, E. B. (2021). Multi ‘omic data integration: a review of concepts, considerations, and approaches. *Seminars Perinatology* 45 (6), 151456. doi:10.1016/j.semperi.2021.151456
- Shamanna, P., Dharmalingam, M., Sahay, R., Mohammed, J., Mohamed, M., Poon, T., et al. (2021). Retrospective study of glycemic variability, BMI, and blood pressure in diabetes patients in the digital twin precision treatment program. *Sci. Rep.* 11 (1), 14892. doi:10.1038/s41598-021-94339-6
- Shannon, P., Markiel, A., Ozier, O., Baliga, N. S., Wang, J. T., Ramage, D., et al. (2003). Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.* 13 (11), 2498–2504. doi:10.1101/gr.1239303
- Sharifi-Noghabi, H., Zolotareva, O., Collins, C. C., and Ester, M. (2019). MOLI: multi-omics late integration with deep neural networks for drug response prediction. *Bioinformatics* 35 (14), i501–i509. doi:10.1093/bioinformatics/btz318
- Sherwani, M. K., Ruuskanen, M. O., Feldner-Busztin, D., Nisantzis Firbas, P., Boza, G., Móréh, Á., et al. (2025). Multi-omics time-series analysis in microbiome research: a systematic review. *Briefings Bioinforma.* 26 (5), bbaf502. doi:10.1093/bib/bbaf502
- Singh, A., Shannon, C. P., Gautier, B., Rohart, F., Vacher, M., Tebbutt, S. J., et al. (2019). DIABLO: an integrative approach for identifying key molecular drivers from multi-omics assays. *Bioinformatics* 35 (17), 3055–3062. doi:10.1093/bioinformatics/bty1054
- Srivastava, U., Aguilar, B., and Asadpoure, A. (2024). “Chapter 2 - types of omics data: Genomics, metagenomics, epigenomics, transcriptomics, proteomics, metabolomics, and phenomics,” in *Integrative omics*. Editor M. K. Gupta (Academic Press), 13–34. doi:10.1016/B978-0-443-16092-9.00002-3
- Stahl, P. L., Salmén, F., Vickovic, S., Lundmark, A., Navarro, J. F., Magnusson, J., et al. (2016). Visualization and analysis of gene expression in tissue sections by spatial transcriptomics. *Science* 353 (6294), 78–82. doi:10.1126/science.aaf2403
- Stahlberg, E. A., Abdel-Rahman, M., Aguilar, B., Asadpoure, A., Beckman, R. A., Borkon, L. L., et al. (2022). Exploring approaches for predictive cancer patient digital twins: opportunities for collaboration and innovation. *Front. Digital Health* 4, 1007784. doi:10.3389/fdgh.2022.1007784
- Stunnenberg, H. G., Hirst, M., and International Human Epigenome Consortium (2016). The international human epigenome consortium: a blueprint for scientific collaboration and discovery. *Cell* 167 (5), 1145–1149. doi:10.1016/j.cell.2016.11.007
- Taieb, R., Kedaigle, A., and Soltis, A. R. (2026). *Mechanistic learning for survival prediction in NSCLC using routine blood biomarkers and tumor kinetics*.
- Thamotharan, P., Srinivasan, S., Kesavadev, J., Krishnan, G., Mohan, V., Seshadri, S., et al. (2023). Human digital twin for personalized elderly type 2 diabetes management. *J. Clin. Med.* 12 (6), 2094. doi:10.3390/jcm12062094
- Thapa, C., and Camtepe, S. (2021). Precision health data: requirements, challenges and existing techniques for data security and privacy. *Comput. Biol. Med.* 129, 104130. doi:10.1016/j.combiomed.2020.104130
- Trimbour, R., Deutschmann, I. M., and Cantini, L. (2024). Molecular mechanisms reconstruction from single-cell multi-omics data with HuMMuS. *Bioinformatics* 40 (5), btac143. doi:10.1093/bioinformatics/btac143
- Tuncbag, N., Gosline, S. J. C., Kedaigle, A., Soltis, A. R., Gitter, A., and Fraenkel, E. (2016). Network-based interpretation of diverse high-throughput datasets through the omics integrator software package. *PLOS Comput. Biol.* 12 (4), e1004879. doi:10.1371/journal.pcbi.1004879
- Valdeolivas, A., Tichit, L., Navarro, C., Perrin, S., Odelin, G., Levy, N., et al. (2019). Random walk with restart on multiplex and heterogeneous biological networks. *Bioinformatics* 35 (3), 497–505. doi:10.1093/bioinformatics/bty637
- Vallée, A. (2024). Envisioning the future of personalized medicine: Role and realities of digital twins. *J. Med. Internet Res.* 26, e50204. doi:10.2196/50204
- Vallée, A. (2026). Multi-scale digital twins for personalized medicine. *Front. Digital Health* 8, 1753906. doi:10.3389/fdgh.2026.1753906

- Venkatesh, K. P., Brito, G., and Boulos, M. N. K. (2024). Health digital twins in life science and health care innovation. *Annu. Rev. Pharmacol. Toxicol.* 64, 159–170. doi:10.1146/annurev-pharmtox-022123-022046
- Verstraete, N., Jurman, G., Bertagnolli, G., Ghavasieh, A., Pancaldi, V., and De Domenico, M. (2020). CovMulNet19, integrating proteins, diseases, drugs, and symptoms: a network medicine approach to COVID-19. *Netw. Syst. Med.* 3 (1), 130–141. doi:10.1089/nsm.2020.0011
- Wang, Z., and Deisboeck, T. S. (2019). Dynamic targeting in cancer treatment. *Front. Physiology* 10, 96. doi:10.3389/fphys.2019.00096
- Wang, B., Mezlini, A. M., Demir, F., Fiume, M., Tu, Z., Brudno, M., et al. (2014). Similarity network fusion for aggregating data types on a genomic scale. *Nat. Methods* 11 (3), 333–337. doi:10.1038/nmeth.2810
- Wang, Y., Tang, S., Ma, R., Zamit, I., Wei, Y., and Pan, Y. (2022). Multi-modal intermediate integrative methods in neuropsychiatric disorders: a review. *Comput. Struct. Biotechnol. J.* 20, 6149–6162. doi:10.1016/j.csbj.2022.11.008
- Wang, H., Arulraj, T., Ippolito, A., and Popel, A. S. (2024). From virtual patients to digital twins in immuno-oncology: lessons learned from mechanistic quantitative systems pharmacology modeling. *Npj Digit. Med.* 7 (1), 1–6. doi:10.1038/s41746-024-01188-4
- Weinstein, J. N., Weinstein, J. N., Collisson, E. A., Mills, G. B., Shaw, K. R. M., Ozenberger, B. A., et al. (2013). The cancer genome atlas pan-cancer analysis project. *Nat. Genet.* 45 (10), 1113–1120. doi:10.1038/ng.2764
- Wishart, D. S., Tzur, D., Knox, C., Eisner, R., Guo, A. C., Young, N., et al. (2007). HMDB: the human metabolome database. *Nucleic Acids Res.* 35 (Suppl. 1_1), D521–D526. doi:10.1093/nar/gkl923
- Wu, D., Wang, D., Zhang, M. Q., and Gu, J. (2015). Fast dimension reduction and integrative clustering of multi-omics data using low-rank approximation: application to cancer molecular classification. *BMC Genomics* 16 (1), 1022. doi:10.1186/s12864-015-2223-8
- Wu, C., Zhou, F., Ren, J., Li, X., Jiang, Y., and Ma, S. (2019). A selective review of multi-level omics data integration using variable selection. *High-Throughput* 8 (1), 4. doi:10.3390/ht8010004
- Zhang, M. (2022). “Graph neural networks: link prediction,” in *Graph neural networks: foundations, frontiers, and applications*. Editor L. Wu (Singapore: Springer Nature), 195–223. doi:10.1007/978-981-16-6054-2_10
- Zhang, W., and Plevritis, S. (2018). Abstract 1297: studying tumor metabolic reprogramming through integration of metabolomics and transcriptomics. *Cancer Res.* 78 (13_Suppl. ment), 1297. doi:10.1158/1538-7445.AM2018-1297
- Zhang, J., Baran, J., Cros, A., Guberman, J. M., Haider, S., Hsu, J., et al. (2011). International cancer genome consortium data Portal-a one-stop shop for cancer genomics data. *Database* 2011 (0), bar026. doi:10.1093/database/bar026
- Zheng, G. X. Y., Terry, J. M., Belgrader, P., Ryvkin, P., Bent, Z. W., Wilson, R., et al. (2017). Massively parallel digital transcriptional profiling of single cells. *Nat. Commun.* 8 (1), 14049. doi:10.1038/ncomms14049

Glossary

Multi-omics integration	The process of combining data from multiple omics layers to obtain a more comprehensive understanding of biological systems
Multilayer network	A network structure in which multiple relationship types or data modalities are represented as distinct but connected layers, allowing the modeling of complex biological systems across different scales
Multiplex network	A specific class of multilayer networks in which the same set of nodes is connected across layers through different types of interactions
Inter-layer edge	A connection linking nodes across different layers of a multilayer network, representing relationships between entities at distinct biological levels
Digital Twin (DT)	A dynamic, patient-specific computational model that integrates multi-scale data to simulate biological processes, disease progression, or potential responses to therapeutic interventions
Tumor microenvironment (TME)	The complex cellular and molecular environment surrounding tumor cells, including non-cancerous cells, such as immune cells and stromal cells, signaling molecules, and extracellular matrix components Intra-tumor
Intra-tumoral heterogeneity	The presence of diverse cellular, genetic and molecular profiles within a single tumor