The EuroBiotech Journal
REVIEW ARTICLE
Medical Biotechnology

# Digital Twins for Precision Medicine and Drug Discovery

Kosi Gramatikoff[1], Ed Judge[2*], Ted Slater[3], Ljubica Milovic[4], Andrian Minchev[5], Miroslav Karabaliev[6]

## Abstract

Precision medicine keeps promising individualized care and keeps delivering population averages. The gap is structural: single-layer omics analysis cannot capture what a biological system does when its parts interact. Digital Twins (DTs) close that gap, not with more data but with dynamic, patient-specific models that fuse multi-omics, clinical records, and longitudinal monitoring into something that runs forward in time. AI drives the simulation; network biology gives it structure. The result predicts disease trajectories and therapeutic response, rather than merely fitting them.

This review follows the full DT lifecycle, from data ingestion and FAIRification through machine learning personalization to clinical deployment. We propose a five-point operational definition that separates genuine DTs from relabeled statistical models, resting on one requirement: a bidirectional virtual-physical loop, not a static prediction. We also examine the gut microbiome, still underserved despite its complexity, and show how ecological frameworks map host-microbe dynamics for precision nutrition, IBD management, and cancer immunotherapy.

Case studies in oncology and neurodegeneration ground the argument, but the bottlenecks matter more: the drug discovery "Attrition Paradox," misaligned multi-modal data, and uncertainty quantification most pipelines skip. Ethical exposure runs alongside these gaps. Privacy law (GDPR/HIPAA) and algorithmic bias are not peripheral; left unaddressed, they widen disparities rather than close them. We close with stakeholder-specific recommendations, for researchers, industry, funders, regulators, aimed at Digital Twin ecosystems that are safe and equitable by design, not by afterthought.

**Keywords:** Digital Twins; Precision Medicine; Multi-omics Integration; Network Biology; Machine Learning Personalization; Gut Microbiome; Health Disparities; Algorithmic Bias

## Introduction

Precision medicine aims to deliver the right treatment to the right patient at the right time, guided by molecular, clinical, and environmental information. Traditional approaches, often reliant on single-omics or coarse clinical phenotyping, fail to capture the dynamic, multifactorial nature of complex diseases. Digital Twins (DTs) provide a paradigm shift by offering computational models that integrate multi-layered biological, clinical, and environmental data to generate dynamic, patient-specific virtual representations [1, 2, 3]. These virtual entities enable simulation of disease progression, response to interventions, and prediction of adverse events, effectively bridging the gap between data generation and actionable insights.

The concept of DTs in biomedicine builds upon advances in multi-omics profiling, which encompass genomics, transcriptomics, proteomics, metabolomics, and microbiomics, providing comprehensive molecular characterization at unprecedented resolution [7, 8]. Network biology complements these datasets by framing diseases as perturbations within interconnected molecular networks, allowing identification of disease modules and prioritization of therapeutic targets based on systemic effects rather than isolated gene-centric views [22]. Artificial intelligence (AI) functions as the integrative engine of DTs, harmonizing heterogeneous data, learning latent representations, generating novel molecular candidates, and simulating patient-specific trajectories [17, 18, 19, 20, 21].

Concurrently, FAIR principles ensure that data remain findable, accessible, interoperable, and reusable, which is essential for constructing federated DT ecosystems that span multiple institutions, datasets, and national boundaries [16].

[1]Medical University Varna, "Professor Marin Drinov" 55, 9002 Bulgaria.
[2]EPAM Systems Ltd, 9 Devonshire Square, 5th Floor, London, EC2M 4YF, UK.
[3]EPAM Systems Inc, Newtown, PA, US.
[4]EPAM Systems LLC, Blvd. Vojvode Misica 17, 11000 Belgrade, Serbia.
[5]Trakia University, Faculty of Technics and Technologies, 38 Graf Ignatiev str., Yambol 8600, Bulgaria.
[6]Trakia University, Faculty of Medicine, Department of Physics, Biophysics, Roentgenology and Radiology, Stara Zagora, Bulgaria.

***Corresponding author:**
Edison Judge
E-mail: ed.judge@outlook.com

**DOI:** 10.2478/ebtj-2026-0011

The translation of DTs into clinical and drug discovery applications necessitates rigorous pipelines, encompassing data ingestion, harmonization, network-aware feature engineering, model training and personalization, virtual trial simulations, and iterative wet-lab validation [4, 5]. Case studies in oncology, metabolic disease, neurodegeneration, and, as newly documented in this review, the gut microbiome exemplify the potential of DTs to generate patient-specific predictions, optimize therapeutic strategies, and accelerate drug development [6, 12, 14, 63, 64, 65]. At the same time, regulatory, ethical, and privacy challenges, including patient consent, data governance, algorithmic bias, and equitable access, must be proactively addressed to ensure trust and societal benefit [30, 31].

This critical review synthesizes current methodologies for constructing and deploying DTs, encompassing multi-omics integration, machine learning and hybrid modeling approaches, FAIR-compliant data engineering, practical pipelines, regulatory landscapes, and translational case studies. By highlighting open challenges and proposing actionable recommendations for researchers, industry, funders, and regulators, we provide a comprehensive roadmap for harnessing DTs to advance precision medicine and drug discovery while ensuring ethical, equitable, and sustainable application [9, 1, 2].

## 2. Precision Medicine and Multi-Omics Overview

Precision medicine seeks to tailor prevention, diagnosis, and treatment strategies to the individual characteristics of each patient, encompassing genetic, molecular, environmental, and lifestyle factors [2, 3]. Multi-omics technologies provide the molecular granularity required to operationalize this vision, capturing complex biological information across multiple layers of cellular function. Genomics reveals inherited variants and somatic mutations that predispose or contribute to disease; transcriptomics quantifies gene expression dynamics; proteomics assesses protein abundance and post-translational modifications; metabolomics maps metabolic states; and microbiomics characterizes the composition and functional potential of microbial communities influencing host physiology [7, 8, 9].

The integration of these omics layers enables the identification of patient subtypes, molecular signatures of disease progression, and potential therapeutic targets with higher precision than any single layer alone [10, 11, 15]. Single-omics approaches, while informative, fail to capture emergent properties arising from cross-layer interactions, such as signaling cascades, metabolic flux, and epigenetic regulation, which often drive complex disease phenotypes [7, 11]. Multi-omics analyses also facilitate longitudinal monitoring, allowing the detection of early molecular changes before clinical manifestation, the assessment of treatment responses, and the identification of adaptive resistance mechanisms [9, 14].

Beyond static profiling, temporal and spatial multi-omics approaches are emerging as critical tools for understanding disease heterogeneity at the cellular and tissue levels. Single-cell transcriptomics and proteomics resolve cell-type-specific states and interactions, revealing cellular hierarchies and rare subpopulations that may drive pathology [13, 23, 26]. Spatial transcriptomics and imaging-based proteomics contextualize molecular data within tissue architecture, bridging molecular observations with histopathological and functional insights [23, 24, 25]. Together, these approaches provide a rich substrate for constructing patient-specific Digital Twins, capable of representing not only molecular states but also their spatial and temporal dynamics [1, 2, 4].

The high dimensionality and heterogeneity of multi-omics datasets necessitate advanced computational strategies for integration and interpretation. Machine learning and network-based approaches offer powerful solutions, identifying latent structures, capturing cross-omics interactions, and enabling predictive modeling [7, 8, 10, 11]. These computational frameworks form the foundation for Digital Twins, which synthesize multi-omics data into coherent, actionable models for personalized medicine [1, 2, 3]. The architecture through which these multi-omics layers are harmonized into a coherent, adaptive Digital Twin, with patient data flowing into an AI integration engine and predictions fed back through a closed learning cycle — is illustrated in Figure 1.

It is important to acknowledge the limitations intrinsic to each omics layer. Genomics captures inherited risk but cannot predict regulatory plasticity without transcriptomic or epigenomic context. Transcriptomics reflects snapshot gene expression states but misses post-translational control that governs protein activity. Metabolomics integrates upstream biology into biochemical readouts but is sensitive to pre-analytical variability and compositional data challenges. Microbiomics introduces additional complexity through inter-individual variability, compositional structure requiring specialized statistical methods (e.g., ALDEx2, ANCOM), and the absence of universal reference standards across profiling platforms [8]. Recognizing these limitations is essential for setting appropriate expectations when building Digital Twins grounded in multi-omics data.

## 3. Digital Twin Concept in Biomedicine

The concept of a Digital Twin (DT) originates from engineering and industrial applications, where virtual replicas of physical assets enable simulation, optimization, and predictive maintenance. In biomedicine, this concept has been adapted to create patient-specific virtual representations that capture the molecular, cellular, organ-level, and even systemic dynamics of human biology [1, 2, 3]. Digital Twins integrate diverse multi-omics data, clinical measurements, imaging, and longitudinal monitoring to construct dynamic models that mirror individual physiological and pathological states [2, 4].

The utility of Digital Twins lies in their ability to facilitate in silico experimentation. By simulating disease progression, drug responses, and environmental perturbations within a computational model, DTs enable hypothesis testing that would be costly, time-consuming, or ethically challenging in

real patients [1, 3, 5]. For instance, a tumor-specific DT can simulate the effects of combination therapies on different clonal populations, thereby prioritizing treatment strategies and predicting resistance pathways before clinical administration [6, 12, 14]. Similarly, metabolic DTs can integrate continuous glucose monitoring, metabolomics, and genetic risk factors to predict individual glycemic responses to dietary interventions [3, 4].

Constructing a reliable biomedical Digital Twin requires three critical components: comprehensive data acquisition, mechanistic modeling, and advanced computational inference. Comprehensive data acquisition encompasses multi-omics profiling, longitudinal clinical measurements, wearable device outputs, and imaging modalities, all annotated with rich metadata [4, 16]. Mechanistic modeling leverages knowledge of biochemical networks, signaling pathways, and physiological processes to structure predictive models [2, 22]. Computational inference employs machine learning and AI to integrate heterogeneous datasets, extract latent features, predict outcomes, and generate counterfactual scenarios for patient-specific intervention strategies [7, 8, 10, 11].

Digital Twins are inherently dynamic; they evolve over time as new data become available, allowing continuous refinement of predictions. This adaptability is essential for precision medicine, where patient conditions change in response to treatment, environmental factors, and disease progression [3, 4, 5]. Unlike static models, DTs can continuously incorporate feedback from clinical outcomes and experimental validation, thereby improving both predictive accuracy and mechanistic insight [29, 5].

The adoption of Digital Twins in drug discovery and clinical decision-making also hinges on standardization and interoperability. Data must comply with FAIR principles (Findable, Accessible, Interoperable, Reusable) to ensure that multi-institutional collaboration is feasible and that models can be reliably updated as new knowledge emerges [16]. Moreover, regulatory and ethical frameworks must evolve to accommodate adaptive, patient-specific computational models while safeguarding privacy, equity, and data governance [30, 31].

Overall, the Digital Twin concept represents a paradigm shift from population-based medicine to individualized, mechanistically informed, and data-driven healthcare. The subsequent section delves into multi-omics integration strategies that provide the molecular foundation for constructing these personalized virtual models.

## 4. Multi-Omics for Digital Twin Construction

Building an accurate and predictive Digital Twin requires comprehensive molecular characterization of the patient. Multi-omics approaches provide a holistic view of biological systems by capturing information across multiple molecular layers, including genomics, transcriptomics, proteomics, metabolomics, and microbiomics [7, 8, 9]. Each omics layer contributes complementary insights: genomics defines inherited variants and somatic mutations; transcriptomics reveals dynamic gene expression patterns; proteomics elucidates functional protein networks and post-translational modifications; metabolomics reflects biochemical activity and pathway fluxes; and microbiomics captures host-microbiota interactions impacting systemic physiology [8, 9, 11].

The integration of multi-omics data is crucial because single-layer analyses often fail to capture the complexity of disease mechanisms. Longitudinal multi-omics further allows tracking of disease progression, therapy response, and adaptive physiological changes over time, enabling Digital Twins to evolve dynamically in parallel with the patient [3, 4, 9]. Integrating multi-omics into a coherent Digital Twin also benefits from network biology frameworks, which map molecular entities onto interaction networks, pathways, or regulatory circuits [22].

From a technical perspective, each omics modality presents unique challenges. High-throughput sequencing generates large-scale genomic and transcriptomic datasets requiring careful preprocessing, alignment, and normalization. Proteomics and metabolomics data often suffer from missing values, batch effects, and instrument-specific noise, necessitating robust quality control and harmonization methods [7, 8]. Microbiome profiling introduces additional complexity due to compositional data structures, requiring log-ratio transformations, and substantial inter-individual variability that confounds cross-study comparisons [8]. Addressing these challenges ensures that integrated datasets are of sufficient quality and consistency to inform mechanistic modeling and predictive analytics.

The combination of multi-omics and longitudinal data provides the substrate for AI-driven modeling, enabling patient-specific simulations, outcome prediction, and virtual experimentation [1, 2, 3, 4]. Machine learning methods leverage the richness of integrated omics data to extract latent features, detect nonlinear interactions, and personalize predictive models.

## 5. Multi-Omics Integration Methods for Digital Twins

Transforming multi-omics datasets into actionable insights for Digital Twins requires robust integration strategies that capture cross-layer interactions while accommodating the heterogeneity inherent to biological data [7, 11].

### 5.1 Integration Architectures: Early, Late, and Intermediate Fusion

Early fusion (feature-level integration) combines molecular features from different omics layers into a unified matrix prior to modeling, directly capturing correlations across layers [7, 10, 11]. However, it faces challenges related to high dimensionality, scale differences, and missing data. Late fusion constructs separate models for each omics layer and combines predictions through ensemble strategies, accommodating heterogeneity naturally but potentially overlooking subtle cross-layer interactions [15]. Intermediate fusion, implemented

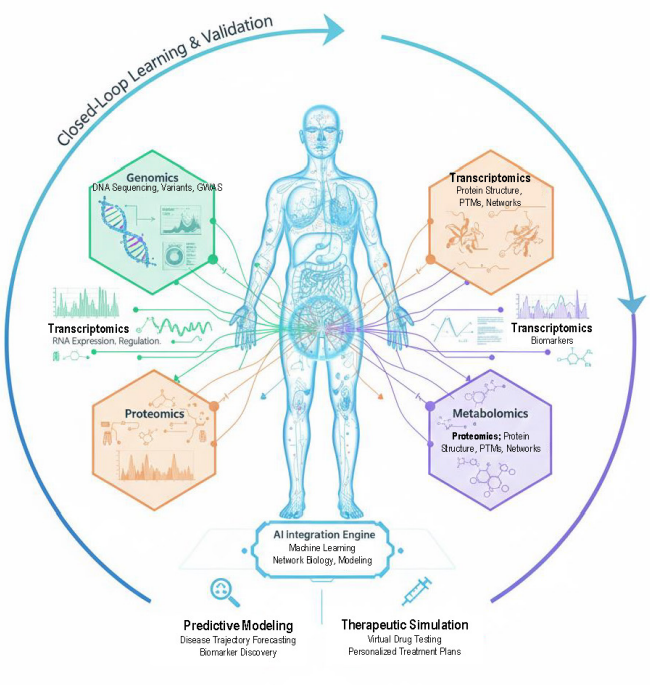


**Figure 1.** Multi-omics integration architecture for Digital Twin construction and the closed-loop learning cycle.
Omics data ecosystem is multi-layered, and underpins biomedical Digital Twin construction, organized around a central human body schematic representing the patient. Five principal omics modalities converge on the patient model: Genomics (DNA sequences, variants, GWAS studies; upper left, green hexagon), providing the inherited mutational landscape; Transcriptomics (RNA expression dynamics and regulatory networks; left, connecting arrows), capturing dynamic gene activity; Proteomics (protein structure, post-translational modifications, and interaction networks; lower left, orange hexagon), reflecting functional protein states; Metabolomics (metabolic biomarkers and pathway fluxes; lower right, purple hexagon), integrating biochemical activity; and a fifth layer combining transcriptomic and pharmacodynamic readouts (upper right, orange hexagon). All five streams feed into a central AI Integration Engine, employing machine learning, network modeling, Bayesian inference, and causal modeling, which translates heterogeneous molecular signals into two translational outputs shown at the base of the figure: Predictive Modeling (disease trajectory forecasting, biomarker discovery) and Therapeutic Simulation (virtual drug testing, personalized treatment planning). The outer circular arrow labeled "Closed-Loop Learning & Validation" represents the bidirectional feedback architecture that distinguishes a Digital Twin from a static model: predictions generated in silico are validated against clinical outcomes, and new patient data continuously refine the twin's parameters, enabling real-time personalization. This closed-loop architecture is the operationally defining feature of a Digital Twin as distinguished from conventional computational models or systems biology simulations.

through multi-view learning, tensor decomposition, or graph-based methods, attempts to balance these strengths by jointly modeling multi-omics while preserving layer-specific structures [7, 11, 22].

### 5.2 Modeling Approaches: Deterministic, Stochastic, and Hybrid

A critical, and often underappreciated, distinction in DT modeling is between deterministic and stochastic frameworks. Deterministic models (ordinary differential equations, constraint-based metabolic models) produce single-trajectory predictions given fixed parameters, offering mechanistic interpretability but potentially underestimating biological uncertainty. Stochastic models (stochastic differential equations, agent-based models, probabilistic graphical models) explicitly represent randomness in molecular processes such as transcription bursting and stochastic cell fate decisions, which is essential when modeling rare events or small cell populations [22].

Hybrid approaches combine mechanistic knowledge with data-driven models, addressing the limitations of purely data-driven or purely mechanistic paradigms [2, 4, 22]. Physics-informed neural networks embed conservation laws and reaction kinetics into neural network training. Multi-scale models link mechanistic simulations at molecular or cellular levels with surrogate machine learning approximations at higher scales. Causal discovery approaches leverage interventional data from CRISPR screens and perturbation experiments to refine network structures and support mechanistic interpretation [10, 22].

### 5.3 Network Types and Their Roles

Network-based representations facilitate identification of disease modules, hub regulators, and intervention points

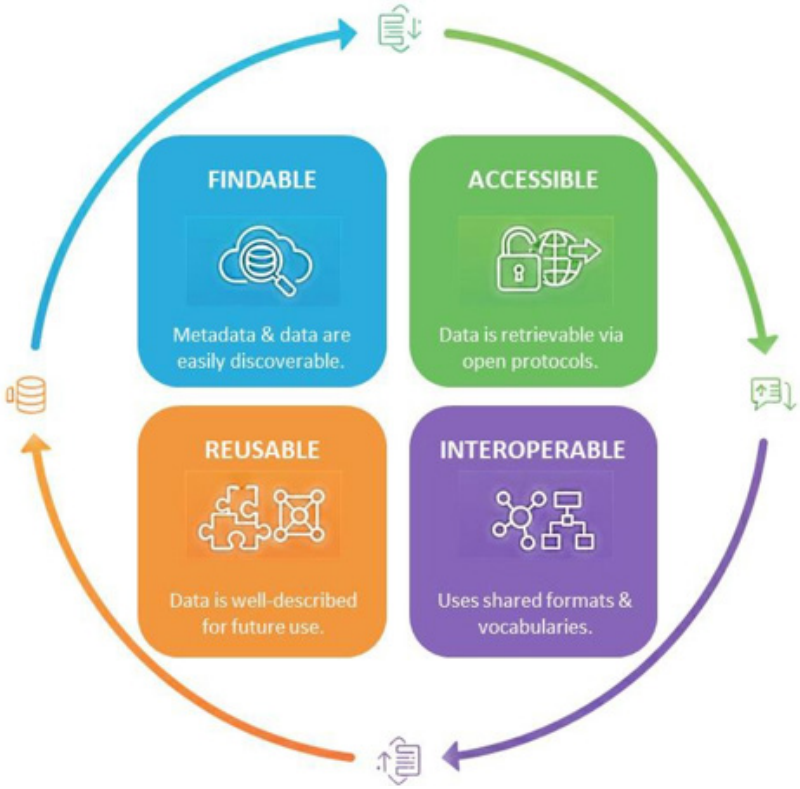


**Figure 2. The FAIR principles as a continuous data stewardship cycle for Digital Twin ecosystems.**
The four FAIR principles are presented as an interconnected cycle rather than an independent checklist, reflecting their interdependent and cumulative nature in practice. The cycle proceeds clockwise from the upper left through four quadrants, each represented by a distinct color and icon: FINDABLE (blue, upper left), metadata and data are assigned globally unique persistent identifiers and deposited in searchable registries so that both humans and automated systems can discover them; ACCESSIBLE (green, upper right), data are retrievable through standardized, open protocols (APIs, SPARQL endpoints), with metadata permanently retained even when original data are removed; INTEROPERABLE (purple, lower right), data employ community-accepted formats, controlled vocabularies, and ontologies (Gene Ontology, Human Phenotype Ontology, Disease Ontology) so they can be integrated seamlessly across platforms and workflows; REUSABLE (orange, lower left), data are richly described with provenance, experimental context, quality metrics, and clear licensing terms (Creative Commons, Data Use Agreements) so that secondary analyses and federated Digital Twin networks can build upon them without ambiguity. The connecting arrows and central cycle symbol emphasize that FAIR compliance is not a one-time deposition event but an ongoing stewardship process: findability enables access, access enables interoperability, interoperability enables reuse, and demonstrated reuse motivates further data sharing. For Digital Twins specifically, FAIR-compliant data infrastructure is the prerequisite for federated DT ecosystems, cross-institutional model training, and reproducible longitudinal analyses.

[10, 11, 22]. Relevant network types include: protein-protein interaction (PPI) networks capturing physical binding relationships; gene regulatory networks (GRNs) representing transcriptional control logic; metabolic networks encoding enzymatic reactions and metabolic flux; and ecological interaction networks (generalized Lotka-Volterra models) relevant to microbiome Digital Twins. Each network type carries distinct modeling assumptions, PPI networks are static snapshots; GRNs are context-dependent; metabolic networks require stoichiometric constraints; ecological networks assume specific competition-cooperation dynamics. Selecting the appropriate network type for the biological question at hand is therefore a design decision that directly affects DT validity.

### 5.4 Machine Learning, Deep Learning, and Generative Models

Machine learning and deep learning approaches increasingly dominate multi-omics integration for Digital Twins. Unsupervised methods (iCluster, MOFA, t-SNE, UMAP, autoencoders) reveal latent structures without labeled outcomes [7, 11, 13]. Supervised methods (random forests, elastic net, multi-task learning) predict clinical outcomes from integrated omics features [9, 14, 15]. Graph neural networks integrate molecular network topology with omics features, while transformer architectures dynamically weight omics contributions [10, 11, 13]. Generative models (variational autoencoders, GANs, diffusion models) enable in silico perturbation experiments by generating synthetic omics

profiles and exploring molecular design spaces [18, 19, 21].

**5.5 Rigorous Benchmarking, Validation, and Explainability**

Rigorous benchmarking and validation are essential for clinical translation. Standardized evaluation frameworks assess predictive accuracy, robustness, scalability, and interpretability across datasets [15]. Advanced cross-validation, external validation, and biological validation strategies confirm that models generalize and reflect underlying biology [5, 14, 15]. Explainable AI techniques, such as SHAP values, integrated gradients, attention weights, and pathway enrichment, provide transparency and facilitate regulatory compliance [30, 31]. Uncertainty quantification, via Bayesian approaches, ensembles, and conformal prediction, ensures that clinical decisions are informed by confidence estimates rather than point predictions [5].

## 6. FAIR, Interoperability, and Data Engineering for Digital Twins

Effective multi-omics Digital Twins rely not only on robust integration methods but also on data that is structured, interoperable, and reusable. Ensuring these properties requires adherence to FAIR principles, Findable, Accessible, Interoperable, and Reusable, which have become a cornerstone of modern biomedical data stewardship [16].

Findability requires globally unique persistent identifiers (DOIs, ORCIDs, RRIDs) coupled with rich metadata. Accessibility ensures data can be obtained through clear protocols while balancing privacy requirements. Interoperability is achieved through community-accepted formats, controlled vocabularies, and ontologies (Gene Ontology, Human Phenotype Ontology, Disease Ontology). Reusability emphasizes comprehensive documentation, clear licensing, and quality metrics so that data support future analyses beyond the original study [16]. The interdependence of FAIR properties, and their role as a continuous stewardship cycle rather than a set of independent criteria, is illustrated in Figure 2, which forms the conceptual backbone of the data infrastructure requirements for federated Digital Twin ecosystems.

Despite widespread adoption of FAIR principles, practical barriers persist: metadata incompleteness, persistent identifier instability, legal constraints under GDPR and HIPAA, resource-intensive retroactive FAIRification of legacy datasets, and cultural misalignments driven by publication pressures [30, 31, 16]. Technological infrastructures, data lakes, common data models, federated systems like Personal Health Train and DataSHIELD, can alleviate these barriers by embedding FAIR compliance into routine workflows [16].

Lessons from large-scale initiatives including the European Bioinformatics Institute (EBI), NIH Data Commons Pilot, TCGA, and UK Biobank highlight that technical interoperability alone is insufficient without concomitant cultural and incentive alignment [29, 16]. Key lessons include the importance of upfront investment in metadata, automated validation tools, sustainable funding models, and technical solutions that balance openness with privacy through synthetic data generation and secure computation environments [16].

## 7. AI as the Enabling Engine for DT Creation and Drug Discovery

Building on FAIR data stewardship and multi-omics integration, artificial intelligence (AI) serves as the central engine that transforms raw biological data into actionable Digital Twins. By harmonizing heterogeneous datasets, extracting meaningful features, generating novel molecules, and enabling predictive simulations, AI underpins the creation, personalization, and translational application of DTs in drug discovery [17, 18, 19, 20, 21].

AI performs data harmonization via neural network methods (scVI, Harmony) that remove batch effects while preserving biological variation [7, 8, 13], feature extraction via autoencoders and graph embeddings [10, 22], and generative molecular design via variational autoencoders, GANs, and reinforcement learning [18, 19]. Model personalization leverages transfer learning, Bayesian updating, and meta-learning to adapt population-trained models to individual patients with limited data [3, 4, 10]. AI-driven DTs also enable simulation and counterfactual exploration through neural ODEs, causal inference, and generative sampling [1, 2, 3, 18].

Although long-term clinical outcomes remain unproven, positive impacts of generative AI in molecular design are increasingly evident. Platforms such as Insilico Medicine's Pharma.AI have produced novel fibrosis candidates advancing to Phase II trials in under 18 months [17, 20]. Recursion Pharmaceuticals integrates high-content cellular imaging with generative models to identify drug-target interactions [20], while AlphaFold revolutionized structure prediction and tools such as RFdiffusion and ProteinMPNN enable therapeutic protein design [21, 28, 29].

Despite these advances, limitations remain. Challenges include chemical synthesizability of AI-generated molecules [19, 20], off-target promiscuity, incomplete ADME predictions, and the necessity for experimental validation [5, 17, 31]. Furthermore, reproducibility concerns persist: many AI drug discovery studies lack external validation cohorts, pre-registered analysis plans, and code availability, echoing broader replication crisis concerns in biomedical research. Explainability (SHAP, attention maps, pathway-level interpretation), uncertainty quantification (Bayesian neural networks, conformal prediction), and regulatory compliance (FDA SaMD guidelines) are therefore essential for clinical deployment [30, 31].

## 8. Practical Pipeline: From Raw Omics to a Deployed Digital Twin

The implementation of Digital Twins in drug discovery requires a robust, end-to-end pipeline that transforms raw omics data into actionable insights [1-3, 5]. This pipeline bridges theoretical concepts, multi-omics integration, and AI-

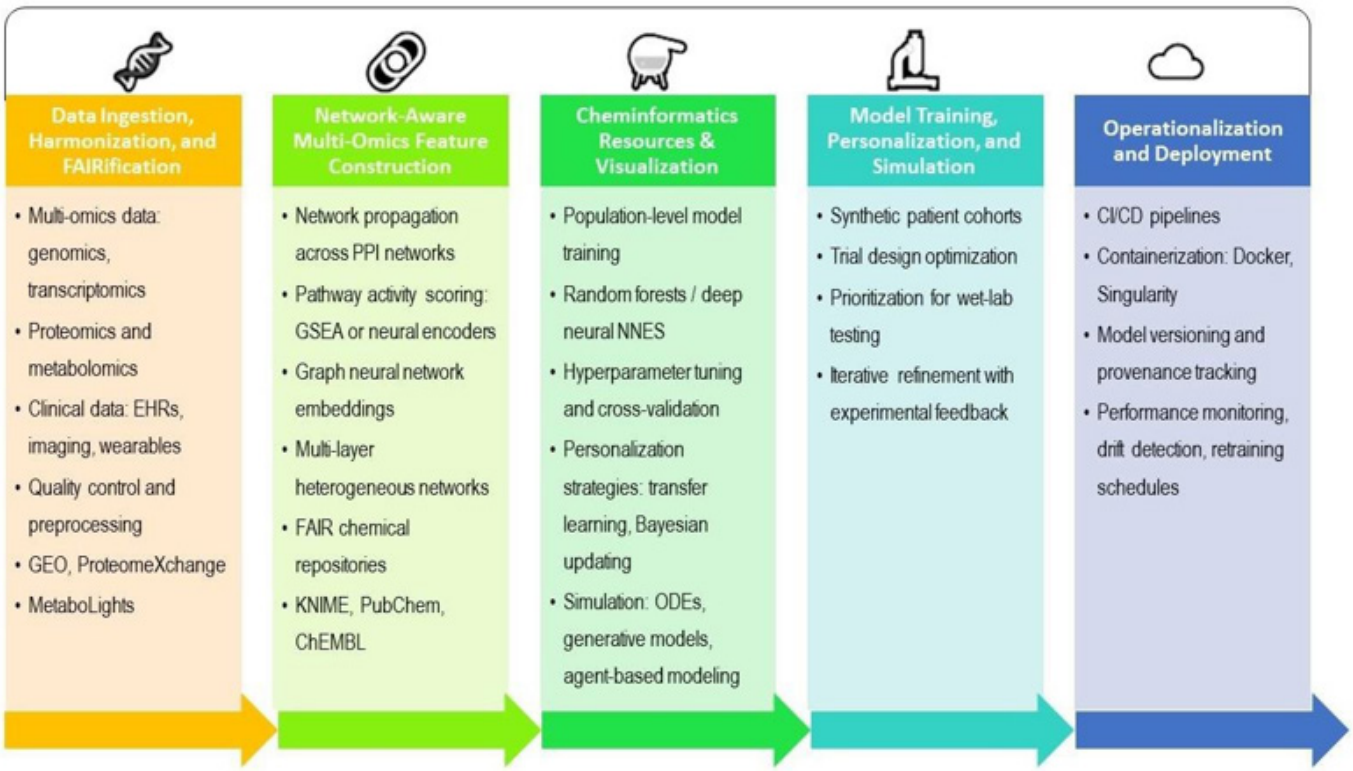


**Figure 3. End-to-end pipeline for Digital Twin development in drug discovery: from raw omics data to deployed clinical system**.

Five-stage operational pipeline are presented, through which raw multi-omics data are transformed into a deployed, patient-specific Digital Twin for precision medicine and drug discovery. The stages are arranged as sequential horizontal arrows with color-coding to indicate functional domain: Stage 1: Data Ingestion, Harmonization, and FAIRification (yellow/orange): Multi-omics datasets (genomics, transcriptomics, proteomics, metabolomics), clinical information (EHRs, imaging, wearable device streams), and experimental metadata are imported, subjected to quality control and preprocessing, and deposited in FAIR-compliant repositories (GEO, ProteomeXchange, MetaboLights) with standardized metadata and persistent identifiers. Stage 2: Network-Aware Multi-Omics Feature Construction (green): Harmonized data are integrated with biological network knowledge through network propagation across protein-protein interaction (PPI) networks, pathway activity scoring (GSEA or neural encoders), graph neural network embeddings, and multi-layer heterogeneous network construction. FAIR chemical repositories (PubChem, ChEMBL) and workflow platforms (KNIME) support molecular feature integration. Stage 3: Cheminformatics Resources and Visualization (green, continued): Population-level model training (random forests, deep neural networks) proceeds in parallel with chemical structure analysis and lead optimization, incorporating hyperparameter tuning, cross-validation, personalization strategies (transfer learning, Bayesian updating), and simulation via ODEs, generative models, and agent-based modeling. Stage 4: Model Training, Personalization, and Simulation (teal/blue): Synthetic patient cohorts are generated for virtual trial design; trial parameters (sample size, inclusion criteria, endpoints) are optimized; high-confidence predictions are prioritized for wet-lab validation (cell assays, organoids, animal models); and experimental feedback drives iterative model refinement. Stage 5: Operationalization and Deployment (grey-blue): Validated models are deployed through CI/CD pipelines for automated updates, containerization (Docker, Singularity) for reproducible environments, model versioning and provenance tracking, and real-time performance monitoring with drift detection and retraining schedules. The sequential arrow structure encodes the directionality of the pipeline, while the color progression from warm (data input) to cool (deployment) reflects the transformation from raw biological material to operational clinical tool. Each stage is designed to be modular, enabling independent updating as new omics technologies, AI methods, or regulatory requirements evolve.

powered simulation with practical deployment for translational applications. The pipeline consists of five integrated stages with feedback loops enabling iterative refinement.

**8.1 Data Ingestion, Harmonization, and FAIRification**

The first stage involves importing multi-omics datasets (genomic sequences, transcriptomic counts, proteomics mass spectrometry outputs, metabolomic measurements) alongside clinical information such as EHRs, imaging data, and wearable device streams. Quality control procedures assess data integrity using platform-specific tools (FastQC, MaxQuant), followed by preprocessing, batch effect correction (ComBat, Harmony, limma), and FAIRification: assigning persistent identifiers, annotating metadata following community standards (ISA framework, DATS), mapping entities to ontologies (GO, HPO, ChEBI), and depositing data in appropriate repositories (GEO, ProteomeXchange, MetaboLights) [4, 16]. The complete five-stage pipeline, spanning data ingestion and FAIRification through to operationalization and deployment, is illustrated in Figure 3; each stage is described in detail in the subsections that follow.

**8.2 Network-Aware Multi-Omics Feature Construction**

Once data are harmonized and FAIR-compliant, the pipeline advances to network-aware feature construction [7-11]. Network propagation techniques diffuse molecular signals across protein interaction networks. Pathway activity scoring (GSEA, neural network encoders) summarizes expression into biologically meaningful features. Graph neural networks leverage molecular networks annotated with omics data to generate embeddings that capture both local molecular measurements and global network context [10, 11] (see Figure 3, Stage 2 embeddings: PPI network propagation, GSEA, and GNN). Multi-layer network integration connects genes, proteins, metabolites, and phenotypes into heterogeneous networks [7, 8].

**8.3 Cheminformatics Resources and Visualization Tools**

Effective implementation of DT pipelines in drug discovery depends on robust cheminformatics infrastructure [32, 33]. Open-source platforms such as Ketcher provide web-based chemical structure editors; Miew enables high-performance 3D rendering of protein structures; tools like AiZynthFinder, ASKCOS, IBM RXN, and ChemAIRS enable AI-driven retrosynthetic analysis [42-45]. Structure-activity relationship analysis through StarDrop and Vortex guides lead optimization, while workflow integration platforms (KNIME, RDKit) enable reproducible cheminformatics pipelines [47-51]. FAIR-compliant repositories (PubChem, ChEMBL, DrugBank, ZINC) form the foundation for training AI models and enabling community-wide collaboration [46, 51].

**8.4 Model Training, Personalization, and Simulation**

Population-level models are trained using stratified data splits, with hyperparameters optimized through cross-validation or Bayesian optimization. Personalization strategies leverage transfer learning, Bayesian parameter updating, and meta-learning (MAML) to adapt models to individual patients. Simulation capabilities powered by neural ordinary differential equations, generative latent-space modeling, and agent-based modeling enable counterfactual generation, allowing exploration of hypothetical interventions [1, 2, 22].

A critical design choice is between deterministic simulation (e.g., ODE-based pharmacodynamic models) and stochastic simulation (e.g., agent-based cellular models). Deterministic models are computationally efficient and interpretable but cannot capture inherent stochasticity. Stochastic agent-based models are more realistic for heterogeneous cell populations but require orders of magnitude more compute and are harder to validate. The appropriate choice depends on the biological question and the scale of the DT being constructed.

**8.5 In-Silico Trials, Virtual Cohorts, and Wet-Lab Validation**

Synthetic patient cohorts generated from DT sampling simulate diverse populations and evaluate drug efficacy, safety, and treatment response heterogeneity. Trial design is optimized using DT predictions to determine sample size, inclusion criteria, and relevant endpoints prior to actual enrollment [5]. Predictions with high confidence and translational potential are prioritized for wet-lab testing using cell-based assays, organoids, or animal models, with iterative refinement of DT models using experimental feedback [5, 31].

**8.6 Operationalization and Deployment**

Validated models are operationalized through CI/CD pipelines for automated model updates, containerization (Docker, Singularity) for reproducible environments, model versioning and provenance tracking, and real-time performance monitoring with drift detection [2, 31]. The complete pipeline produces AI-integrated, patient-specific Digital Twins capable of predicting responses to therapeutic interventions and supporting drug discovery through virtual screening and compound optimization.

**9. Regulatory, Ethical, and Privacy Challenges**

As Digital Twins transition from computational frameworks to applied tools, navigating regulatory, ethical, and privacy landscapes becomes imperative [30, 31].

**9.1 Patient Privacy, Consent, and Data Governance**

Patient privacy underpins trust in DT applications [2]. GDPR in Europe mandates rights including erasure, data minimization, and purpose limitation; HIPAA in the US protects personal health information; PIPEDA (Canada) and LGPD (Brazil) add complexity to international collaborations. Informed consent models range from specific (restrictive but precise) to broad and dynamic (flexible but raising autonomy concerns). De-identification strategies include irreversible anonymization, pseudonymization, and differential privacy providing mathematical leakage guarantees [2]. Federated architectures and data access committees ensure that analysis is permissible

**Table 1.** Representative Digital Twin Studies Across Therapeutic Areas

| Therapeutic Area | Study / Platform | DT Type & Data | Outcome / Impact |
|---|---|---|---|
| Oncology | NCI-MATCH / TAPUR trials [12,14] | Tumor multi-omics DT (WES, RNA-seq, proteomics); network module analysis | Genomics-guided therapy improved response; identified PI3K/MAPK synthetic lethal targets |
| Oncology | Insilico Medicine Pharma.AI [17,20] | Generative AI + DT; phenotypic screening + target identification | Novel fibrosis drug candidate reached Phase II in <18 months; validated DT pipeline |
| Metabolic Disease | Zeevi et al. Cell 2015; Mendes-Soares Nat Med 2019 | Metabolic DT: genomics + gut microbiome + CGM + wearables | Personalized dietary advice outperformed generic; microbiome key predictor of glycemia |
| Microbiome / Nutrition | Quinn-Bohmann et al. Nat Microbiol 2024 [63] | Mechanistic MCMM gut microbiome DT; dietary fiber to SCFA mapping | Individualized SCFA production predictions; transparent, mechanistic, clinically translatable |
| Microbiome / IBD | Caenepeel et al. Gastroenterology 2024 [65] | Multi-omics ML DT; dysbiosis scoring for FMT response prediction | AUC 0.96 for FMT response in UC; first prospective microbiome DT for biologic selection |
| Microbiome / Neonatal | Sizemore et al. Sci Adv 2024 [64] | Generative Q-net DT of infant gut ecosystem; longitudinal 16S + clinical | Predicted neurodevelopmental risk before symptoms; identified key microbial cross-talk taxa |
| Neurodegeneration | Alzheimer multi-scale DT [23-27] | Single-cell transcriptomics + PET/MRI imaging + network perturbation | Identified APOE lipid / TREM2 microglial modules; virtual anti-amyloid trial design |
| Cardiology | Corral-Acero et al. 2020 [2] | Heart DT integrating MRI, hemodynamics, clinical parameters | 25% improvement in outcomes vs. standard care; real-time cardiac DT demonstrated |
| Drug Discovery | AlphaFold / RFdiffusion [28,29] | Protein structure DT for target validation and de novo enzyme design | Revolutionized structure prediction; accelerated target validation from months to days |

DT, Digital Twin; NCI-MATCH, National Cancer Institute Molecular Analysis for Therapy Choice trial; TAPUR, Targeted Agent and Profiling Utilization Registry trial; WES, whole-exome sequencing; RNA-seq, RNA sequencing; PI3K/MAPK, phosphoinositide 3-kinase / mitogen-activated protein kinase signaling pathways; CGM, continuous glucose monitoring; MCMM, Microbial Community-Scale Metabolic Model; SCFA, short-chain fatty acid; ML, machine learning; FMT, fecal microbiota transplantation; UC, ulcerative colitis; AUC, area under the receiver operating characteristic curve; Q-net, quantum-inspired generative network; 16S, 16S ribosomal RNA gene sequencing; PET, positron emission tomography; MRI, magnetic resonance imaging; APOE, apolipoprotein E; TREM2, triggering receptor expressed on myeloid cells 2.

and data remain at source institutions [2].

### 9.2 Validation and Evidence Requirements for Regulatory Acceptance

Regulatory approval for DT-based diagnostics demands robust evidence: analytical validity (accuracy, precision, robustness), clinical validity (meaningful associations with outcomes), and clinical utility (proven improvement in patient management) [5, 30, 31]. Study design must include prospective validation cohorts independent of training data, pre-specified statistical plans, blinded evaluations, and sufficient power. DTs also facilitate adaptive trial designs, basket trials, umbrella trials, master protocols, enabling efficient multi-drug testing within unified infrastructure [31].

### 9.3 Equity, Bias, and Access Issues

AI-driven DTs risk perpetuating health disparities. Algorithmic bias can emerge from training data imbalances, underrepresentation of minority populations, and historical healthcare disparities embedded in labels. Mitigation strategies include curating diverse training cohorts, applying fairness metrics (equalized odds, demographic parity), performing stratified validation across subgroups, and formal bias audits [2, 30]. Access barriers, high cost of multi-omics profiling, infrastructure limitations, digital literacy gaps, require proactive equity strategies, including initiatives modeled on

H3Africa for equitable DT deployment in low- and middle-income countries.

## 10. Case Studies: Digital Twins in Practice

It is instructive to examine real-world applications of Digital Twins across diverse therapeutic areas [2, 6]. These case studies illustrate how multi-omics integration, AI-driven modeling, and network biology principles converge to guide patient stratification, therapeutic optimization, and in silico experimentation while navigating practical and ethical constraints. Table 1 provides a consolidated overview of representative DT studies across therapeutic domains.

### 10.1 Oncology: Tumor Multi-Omics and Drug Response Prediction

Cancer heterogeneity necessitates precision approaches leveraging multi-omics profiling to reveal tumor vulnerabilities and guide therapy selection [12, 14]. Patient-specific tumor DTs integrate whole-exome sequencing (somatic mutations, copy number alterations), RNA sequencing (expression programs), and proteomic quantification (signaling pathway activation) [14]. By constructing tumor-specific molecular networks, activated oncogenic modules such as PI3K/AKT/mTOR and RAS/MAPK can be identified alongside synthetic lethal dependencies [12, 14].

Drug response prediction leverages pharmacogenomic data (GDSC, CCLE) to simulate sensitivity and evaluate combination therapies within patient DTs [12]. Genomics-guided trials such as NCI-MATCH and TAPUR highlight both feasibility and challenges in acquiring timely multi-omics data and translating predictions to clinical decisions. Lessons: single-omics approaches are insufficient; expression and proteomic states are essential; longitudinal monitoring captures resistance mechanisms [12, 14, 5].

### 10.2 Metabolic Disease: Metabolic Networks and Lifestyle-Intervention DTs

Metabolic disorders involve dysregulated energy metabolism understandable through network-based approaches [10]. Personalized nutrition DTs incorporate baseline genomics (diabetes risk variants), gut microbiome metagenomics and metabolomics (microbial and metabolic states), and continuous glucose, activity, and sleep data from wearable devices [2, 3]. These DTs model personalized metabolic networks predicting glycemic responses to dietary intake, accounting for genetic background, microbiome composition, and circadian rhythms [3]. Resulting interventions, optimized for individual metabolic profiles, have been shown to outperform generic dietary advice (Zeevi et al., Cell 2015; Mendes-Soares et al., Nat Med 2019). Lessons include the importance of temporal dynamics and the need for continuously adaptive DTs that update recommendations as physiology changes [2, 3].

### 10.3 Neurodegeneration: Single-Cell and Imaging DTs for Target Identification

Neurodegenerative diseases involve complex interactions among neuronal vulnerability, inflammation, and protein aggregation, illuminated by single-cell and spatial multi-omics [23-27].

Alzheimer's DTs integrate single-cell transcriptomic profiling to identify disease-associated microglia, astrocyte states, and vulnerable neuronal populations, while spatial transcriptomics maps gene expression to tissue architecture (amyloid plaques, tau tangles) [23-25, 27]. Imaging modalities (PET, MRI, fMRI) generate multi-scale DTs capturing disease progression and heterogeneity [2]. Network perturbation analysis identifies disease modules (APOE-driven lipid metabolism, TREM2-mediated microglial activation) that prioritize therapeutic targets [26]. Virtual clinical trials simulate anti-amyloid therapies, predicting responders vs. non-responders based on molecular subtypes [1, 2, 6]. Lessons: single-cell resolution is essential to capture heterogeneity masked by bulk profiling [23-26], and longitudinal biomarkers (CSF proteins, imaging) are needed to validate DT predictions against actual disease trajectories [5].

### 10.4 Microbiome Digital Twins: A New Frontier

The gut microbiome, comprising trillions of microbial cells encoding >100-fold more genes than the human genome, is increasingly recognized as a dynamic "organ" critically influencing metabolism, immunity, neurodevelopment, and drug pharmacokinetics. Modeling this ecosystem as a Digital Twin presents unique opportunities and challenges that distinguish microbiome DTs from other biomedical DT applications (Table 2).

#### 10.4.1 Mechanistic and Ecological Modeling of the Gut Microbiome

Microbial Community-Scale Metabolic Models (MCMMs) represent a major advance in microbiome DT construction. Quinn-Bohmann et al. (Nature Microbiology, 2024) developed an MCMM framework incorporating tens of thousands of metabolites and enzymes across dozens of microbial organisms to predict personalized short-chain fatty acid (SCFA) production profiles in the human gut [63]. Unlike black-box machine learning approaches, MCMMs are transparent and mechanistic, mapping dietary fiber intake quantitatively to SCFA outputs at the individual level, a capacity directly relevant to precision nutrition and metabolic disease management. This work represents a foundational step toward deploying microbiome DTs in the clinic.

Generalized Lotka-Volterra (gLV) ecological models provide an alternative framework, inferring microbial interaction networks, susceptibility to external perturbations (antibiotics, dietary changes, FMT), and growth rates from longitudinal abundance data [68]. Parameters inferred from gLV models can be used in community simulations to predict ecosystem dynamics, identify keystone species, and design targeted microbiome interventions. Complementary approaches include ecological network models (ENM) where nodes represent taxa

**Table 2.** Microbiome Digital Twin Studies: Data, Methods, and Translational Outcomes

| Application / Study | Disease / Context | Omics Data Used | DT / Model Type | Key Finding |
|---|---|---|---|---|
| Zeevi et al. (2015); Mendes-Soares et al. (2019) | Glycemic response / Metabolic DT | Metagenomics, metabolomics, CGM, clinical | Multi-omics regression + personalized glycemic simulator | Microbiome profile + diet predicted postprandial glucose better than genetics alone [3,63] |
| Quinn-Bohmann et al., Nat Microbiol 2024 [63] | Personalized nutrition / SCFA production | Metagenomics, dietary intake, metabolomics | Microbial Community-Scale Metabolic Model (MCMM) | Mechanistic DT maps dietary fiber intake to individual SCFA profiles; outperforms black-box ML [63] |
| Sizemore et al., Sci Adv 2024 [64] | Neonatal neurodevelopment risk | Longitudinal 16S, clinical phenotype | Q-net generative DT of infant microbiome ecosystem | Predicted cognitive risk from microbiome trajectories before clinical manifestation [64] |
| Caenepeel et al., Gastroenterology 2024 [65] | IBD / FMT response prediction | Metagenomics, metabolomics, clinical | ML classifier + dysbiosis scoring | Dysbiosis-informed models predicted FMT response in UC with AUC 0.96 [65] |
| Stein-Thoeringer et al., Nat Med 2023 [66] | Cancer immunotherapy (CAR-T) | Shotgun metagenomics, functional pathways | Predictive ecological model | Microbiome composition predicted CAR-T therapy success; Akkermansia/ Bacteroides as key taxa [66] |
| Khan et al 2025 [67]; Rehman et al 2025 [68] | Microbiome dysbiosis modeling | Multi-omics, ecological network data | Generalized Lotka-Volterra + genome-scale metabolic models | Ecological DT frameworks simulate community dynamics and perturbation responses [67,68] |
| Moysidou et al., Small Science 2024 [69] | Host-microbiome crosstalk / IBD | Transcriptomics, microbiome profiling, 3D bioelectronics | Organ-on-chip + computational DT | Integrates microbial signals with epithelial DT; bridges in vitro and in silico [69] |

DT, Digital Twin; CGM, continuous glucose monitoring; SCFA, short-chain fatty acid; MCMM, Microbial Community-Scale Metabolic Model; ML, machine learning; 16S, 16S ribosomal RNA gene sequencing; Q-net, Quasinet; FMT, fecal microbiota transplantation; UC, ulcerative colitis; IBD, inflammatory bowel disease; AUC, area under the receiver operating characteristic curve; CAR-T, chimeric antigen receptor T-cell therapy.

and edges encode interaction types (mutualism, competition, amensalism) enabling simulation of community restructuring under perturbation scenarios [68].

#### 10.4.2 Neonatal Microbiome DT for Neurodevelopmental Risk Prediction

Sizemore et al. (Science Advances, 2024) demonstrated a landmark application of DT methodology to the infant gut microbiome [64]. Using a quantum-inspired Q-net generative framework, they constructed a digital twin of the gut ecosystem in preterm infants trained on longitudinal fecal samples from UChicago's Comer Children's Hospital and validated on an independent Boston cohort. The model learned n-way time-aware dependencies among microbial taxa without imposing a priori interaction assumptions, substantially outperforming dynamic Bayesian network baselines ($R^2$ improvement >2-fold) in forecasting microbiome maturation trajectories.

Critically, the DT enabled three translational applications: (i) forecasting microbiome maturation from sparse early observations, (ii) predicting neurodevelopmental risk (assessed via head circumference growth as a cognitive proxy) before clinical manifestation, and (iii) identifying a directional influence network among key taxa, moving beyond abundance changes to reveal which organisms drive ecosystem

dynamics. This study represents the first prospective validation of a microbiome DT for a clinical outcome, illustrating that microbiome DTs can predict risk and guide early intervention at a time point when intervention is still possible.

#### 10.4.3 IBD and FMT Response Prediction

Inflammatory bowel disease (IBD), encompassing Crohn's disease and ulcerative colitis, is characterized by dysbiosis, alterations in microbial community composition that perpetuate mucosal inflammation [65, 67]. Fecal microbiota transplantation (FMT) has shown efficacy in ulcerative colitis, but response prediction remains poor using clinical features alone. Caenepeel et al. (Gastroenterology, 2024) demonstrated that multi-omics DT modeling incorporating dysbiosis scoring from metagenomic sequencing and metabolomics achieved an area under the receiver operating characteristic curve (AUC) of 0.963 for FMT response prediction in UC patients [65], substantially exceeding clinical-only models.

Importantly, this study also illustrates the limitations of microbiome DT approaches: a parallel prospective cohort study integrating gut microbiome profiles from metagenomic sequencing and fecal metabolomics found that the predictive utility of microbiome and metabolite data was marginal compared to clinical features alone for predicting response to vedolizumab and ustekinumab in IBD [71]. This contradiction between studies underscores that microbiome DTs are disease- and treatment-context-specific, and that rigorous cross-cohort validation is essential before clinical deployment.

#### 10.4.4 Microbiome DTs for Cancer Immunotherapy

Emerging evidence reveals that gut microbiome composition influences responses to cancer immunotherapy, including checkpoint inhibitors and CAR-T cell therapy. Stein-Thoeringer et al. (Nature Medicine, 2023) demonstrated that in a cross-cohort analysis of 172 patients undergoing CD19 CAR-T cell therapy for lymphoma, ML models based on shotgun metagenomic sequencing identified key microbial taxa (Akkermansia, Bacteroides, Eubacterium) and functional pathways (peptidoglycan biosynthesis) strongly associated with immune activation and long-term responses [66]. The ecological topology of gut microbiota, not just taxon abundance, was predictive of outcomes, suggesting that DT frameworks capturing community interaction structures would outperform single-taxon biomarker approaches [66].

Collectively, these microbiome DT case studies highlight several lessons: (i) mechanistic, ecologically-informed models outperform black-box approaches for interpretability and generalizability; (ii) longitudinal sampling is essential, as snapshot microbiome profiles have limited predictive power; (iii) host-microbiome crosstalk, captured through combined metagenomics, metabolomics, and host omics, substantially improves prediction; and (iv) context-specificity demands that models be validated in the specific disease and treatment context where they will be deployed.

## 11. Limitations, Open Challenges, and Critical Analysis

### 11.1 Limitations and Challenges of Digital Twins

Despite compelling proof-of-concept results, substantial limitations constrain current DT technologies. These are not merely technical, they are fundamental to the field's maturity and should be clearly communicated to inform realistic expectations.

Data quality constraints are pervasive. Multi-omics profiling is expensive, biologically invasive, and prone to batch effects that can systematically bias model parameters. Even with careful preprocessing, residual technical variation can be indistinguishable from genuine biological signal. Longitudinal multi-omics datasets, the lifeblood of adaptive DTs, remain rare due to the cost and patient burden of repeated sampling.

Biological complexity vs. model simplification represents an unavoidable tension. DTs must simplify in order to compute. Current models cannot simultaneously capture intracellular signaling kinetics, intercellular communication, organ-level physiology, microbial ecology, and environmental exposures at the resolution required for reliable personalized predictions. Every DT makes assumptions, and the validity of those assumptions limits the validity of predictions. Explicit characterization of model assumptions and their consequences should be a standard component of DT publications.

External validation failures are common in the biomedical AI literature, and DTs are not immune. Models that perform impressively on training data or internal validation cohorts frequently underperform when applied to independent cohorts drawn from different institutions, populations, or time periods. Transfer learning and domain adaptation partially address this, but robust generalization across populations remains an open problem [15].

Cost and scalability present barriers to widespread deployment. Comprehensive multi-omics profiling for a single patient can cost thousands of dollars, limiting DT application to well-resourced research settings or affluent healthcare systems. Computational demands for large-scale simulation and continuous DT updating require cloud infrastructure that may not be accessible in low-resource settings.

### 11.2 Technical Gaps

Multimodal alignment remains significant [13, 23-27]. Different omics layers vary in sampling scales, time points, and anatomical locations, complicating integration. Emerging spatial multi-omics technologies (10x Visium, CODEX, imaging mass cytometry) provide unprecedented resolution but require computational methods capable of aligning these modalities both spatially and temporally. Longitudinal sampling is critical for adaptive DTs, yet most datasets remain cross-sectional [2, 3]. Causal inference from observational omics data is constrained by the predominance of correlational relationships, while Mendelian randomization, instrumental variable analysis, and causal discovery algorithms advance this frontier with restricted applicability [22].

### 11.3 Organizational and Cultural Change

Technical advances alone are insufficient; organizational and cultural factors shape DT development [16]. Incentivizing FAIR data sharing remains challenging, as traditional academic reward systems prioritize publications over meticulous data curation. Interdisciplinary collaboration is essential but structurally difficult, given distinct vocabularies and incentive systems across clinical, biological, data science, and engineering domains [1, 2]. Industry-academia partnerships offer leverage through precompetitive consortia (IMI, FNIH) [17, 21]. Regulatory evolution remains a key factor, with current frameworks often lagging behind technological advances [30, 31].

### 11.4 Vision: 'Big AI + DT' Ecosystems for Translational Impact

The ultimate vision encompasses interconnected ecosystems spanning healthcare systems, research institutions, and industry partners [2, 6]. Federated DT networks can learn from local datasets while sharing insights globally through privacy-preserving techniques such as secure multi-party computation and homomorphic encryption [2]. Foundation models for biology (ESM protein language model, Nucleotide Transformer genomic models) can capture universal biological patterns, fine-tuned for specific diseases [28, 29]. Real-time adaptive DTs, continuously updated from wearables, EHRs, and periodic multi-omics measurements, promise dynamic, personalized predictions integrated into clinical workflows via decision support systems [2, 3]. Global DT commons supported by WHO, the Wellcome Trust, and the Gates Foundation could democratize access to precision medicine, particularly for under-resourced settings.

## Conclusion

The preceding sections collectively illustrate that Digital Twins represent a transformative paradigm for precision medicine and drug discovery, providing computational frameworks that integrate multi-omics data, network biology principles, AI-driven modeling, and FAIR data practices [1-6, 16]. By creating dynamic, personalized virtual representations of biological systems, DTs enable in silico experimentation, patient stratification, virtual clinical trials, and therapeutic optimization that are difficult or impossible with traditional approaches [1-3].

The insights from multi-omics technologies form the foundation for capturing molecular states across genomic, transcriptomic, proteomic, metabolomic, and microbiomic layers [7-15], while network biology reframes diseases as perturbations of interconnected molecular systems [10, 11, 22]. AI serves as the integrative engine, harmonizing heterogeneous data, generating novel molecular candidates, and simulating interventions [7-15, 17-21]. Adherence to FAIR principles ensures that data remain findable, accessible, interoperable, and reusable, critical for constructing federated DT ecosystems [16].

The gut microbiome emerges as a particularly compelling new frontier for DT applications, combining ecological dynamics, host-microbe crosstalk, and multi-omics readouts in a system where DT modeling can directly guide therapeutic interventions (FMT selection, precision nutrition, immunotherapy optimization). Mechanistic, ecologically-informed models are preferred over black-box approaches, and rigorous prospective validation in disease-specific contexts is essential.

Despite these advances, persistent challenges remain [1, 5, 30, 31]. Technical gaps include multimodal data alignment, longitudinal sampling constraints, causal inference limitations, and scalability. Organizational barriers include misaligned incentives, disciplinary silos, and regulatory lags. Ethical concerns surrounding privacy, algorithmic bias, and equitable access demand proactive strategies.

## Ethical Approval

Not applicable.


## Funding

The authors received no financial support for the research.


## Conflict of Interest

The authors declare that they have no conflict of interest.


## Acknowledgements

None.


## References


1. De Domenico M, Arenas A, Bassett DS, et al. Challenges and opportunities for digital twins in precision medicine from a complex systems perspective. NPJ Digit Med. 2025;8:37. doi:10.1038/s41746-025-00337-x
2. Gevaert O, Sadée C, Nagaraj D, et al. Medical digital twins: enabling precision medicine and medical artificial intelligence. Lancet Digit Health. 2025;7(7):e499–e511. doi:10.1016/S2589-7500(25)00028-7
3. Saratkar SY, Shinde SK, Tiwari RK, et al. Digital twin for personalized medicine development. Front Digit Health. 2025;7:1583466. doi:10.3389/fdgth.2025.1583466
4. Vallée A. Digital twins for personalized medicine require epidemiological data and mathematical modeling: viewpoint. J Med Internet Res. 2025;27:e72411. doi:10.2196/72411
5. Sel K, Heltemes L, Jafari R, et al. Survey and perspective on verification, validation, and uncertainty quantification of digital twins for precision medicine. NPJ Digit Med. 2025;8:40. doi:10.1038/s41746-025-00340-4

6. Lammert J, Lauss M, Harbst K, et al. Large language models-enabled digital twins for precision medicine in rare gynecological tumors. NPJ Digit Med. 2025;8:420. doi:10.1038/s41746-025-00420-0
7. Baião AR, Treveil A, Barros RL, et al. A technical review of multi-omics data integration methods: from classical statistical to deep generative approaches. Brief Bioinform. 2025;26(4):bbaf355. doi:10.1093/bib/bbaf355
8. Yetgin E. Revolutionizing multi-omics analysis with artificial intelligence and data processing. Quant Biol. 2025;13:e70002. doi:10.1002/qub2.70002
9. Lin M, Chen Z, Hu J, et al. Machine learning and multi-omics integration: advancing cardiovascular translational research and clinical practice. J Transl Med. 2025;23:388. doi:10.1186/s12967-025-03888-0
10. Wu Y, Xie L. AI-driven multi-omics integration for multi-scale predictive modeling of genotype-environment-phenotype relationships. Comput Struct Biotechnol J. 2025;27:265–277. doi:10.1016/j.csbj.2025.01.024
11. Nam Y, Kim J, Park S, et al. Harnessing artificial intelligence in multimodal omics data integration: paving the path for the next frontier in precision medicine. Annu Rev Biomed Data Sci. 2025;7:225–250. doi:10.1146/annurev-biodatasci-102523-103801
12. Cunha A, Carlisle B, Bhatt DL, et al. Flexynesis: a deep learning toolkit for bulk multi-omics data integration for precision oncology and beyond. Nat Commun. 2025;16:8261. doi:10.1038/s41467-025-08261-0
13. Ge S, Chen J, Wan L, et al. Multi-omics single-cell data alignment and integration with enhanced contrastive learning and differential attention mechanism. Bioinformatics. 2025;41(8):btaf443. doi:10.1093/bioinformatics/btaf443
14. Wang S, Liu X, Zhang Y, et al. Integrative multi-omics and machine learning reveal critical functions of proliferating cells in prognosis and personalized treatment of lung adenocarcinoma. NPJ Precis Oncol. 2025;9:243. doi:10.1038/s41698-025-0243-1
15. Spooner A, Sowmya A, Sachdev PS, et al. Benchmarking ensemble machine learning algorithms for multi-class, multi-omics data integration in clinical outcome prediction. Brief Bioinform. 2025;26(2):bbaf116. doi:10.1093/bib/bbaf116
16. Tai KH, Buczak P, Brun J, et al. Key concepts in clinical epidemiology: FAIRification of biomedical research data. J Clin Epidemiol. 2025;187:111920. doi:10.1016/j.jclinepi.2025.111920
17. Dharmasivam M, Bhatt S, Walczak K, et al. Leading AI-driven drug discovery platforms: 2025 landscape and global outlook. Pharmacol Rev. 2025;77:100102. doi:10.1016/j.pharmrev.2025.100102
18. Chakraborty C, Bhattacharya M, Chopra H, et al. Generative AI in drug discovery and development: the next revolution of drug discovery and development would be directed by generative AI. Ann Med Surg (Lond). 2024;86:6340–6343. doi:10.1016/j.amsu.2024.08.014
19. Filella -Merce I, García-Jimeno S, Majewski M, et al. Optimizing drug design by merging generative AI with physics-based active learning framework. Commun Chem. 2025;8:238. doi:10.1038/s42004-025-01635-7
20. Niazi SK. Artificial intelligence in small-molecule drug discovery: a critical review of methods, applications, and real-world outcomes. Pharmaceuticals (Basel). 2025;18(9):1271. doi:10.3390/ph18091271
21. Ocaña A, Montero JC, Vera-Badillo FE, et al. Integrating artificial intelligence in drug discovery and early drug development: a transformative approach. Biomark Res. 2025;13:45. doi:10.1186/s40364-025-0045-x
22. Qiao N, Chen Y, Bhatt DL, et al. The evolution of systems biology and systems medicine: from mechanistic models to uncertainty quantification. Annu Rev Biomed Eng. 2025;27:425–448. doi:10.1146/annurev-bioeng-030525-110045
23. Yang Y, Chen Z, Liu X, et al. Spatial integration of multi-omics single-cell data with SIMO. Nat Commun. 2025;16:1265. doi:10.1038/s41467-025-56523-4
24. Sui X, Zhang L, Wang Q, et al. Scalable spatial single-cell transcriptomics and translatomics in 3D thick tissue blocks. Nat Methods. 2025. doi:10.1038/s41592-025-02867-0
25. Duan J, Li H, Zhao X, et al. Single-cell multi-omics and spatial transcriptomics reveal the transcriptional regulatory landscape of clear cell renal cell carcinoma. Transl Androl Urol. 2025;14(10):2953–2974. doi:10.21037/tau-25-1053
26. Ge S, Wang Y, Liu J, et al. Deep learning in single-cell and spatial transcriptomics data analysis: advances and challenges from a data science perspective. Brief Bioinform. 2025;26(2):bbaf136. doi:10.1093/bib/bbaf136
27. Loach M, Bacon W, Bray S, et al. Galaxy single-cell & spatial omics community update: navigating new frontiers in 2025. Cell Genomics. 2025;5(10):101005. doi:10.1016/j.xgen.2025.101005
28. Malhotra Y, Singh A, Patel R, et al. Advancements in protein structure prediction: AlphaFold comparative overview. Comput Biol Med. 2025;188:109842. doi:10.1016/j.compbiomed.2025.109842
29. Rennie ML, Oliver MR. Emerging frontiers in protein structure prediction following the AlphaFold revolution. J R Soc Interface. 2025;22(225):20240886. doi:10.1098/rsif.2024.0886
30. Madaminov F. Regulating the use of AI in drug development: legal challenges and compliance strategies. Food Drug Law J. 2025. Available from: https://www.fdli.org/2025/07/regulating-the-use-of-ai-in-drug-development-legal-challenges-and-compliance-strategies/
31. You JG, Park J, Kim S, et al. Clinical trials informed framework for real world clinical implementation of AI. NPJ Digit Med. 2025;8:107. doi:10.1038/s41746-025-00893-2
32. Marbán -González A, Gómez-Tamayo JC, Cordomi A. Ex-

ploiting PubChem and other public databases for virtual screening in 2025: what are the latest trends? Expert Opin Drug Discov. 2025;20(11):1387–1403. doi:10.1080/17460441.2025.2558161
33. Verma VV, Singh S, Sharma P, Kumar A. A comprehensive review on structural insights through molecular visualization: tools, applications, and limitations. J Mol Model. 2025;31(6):173. doi:10.1007/s00894-025-06402-y
34. Neovarsity. How cheminformatics is used in drug discovery in 2025? Published March 17, 2025. Accessed December 8, 2025. Available from: https://neovarsity.org/blogs/how-cheminformatics-is-used-in-drug-discovery
35. Neovarsity. Top 9 applications of cheminformatics in the biopharma industry [2025]. Published March 6, 2025. Accessed December 8, 2025. Available from: https://neovarsity.org/blogs/cheminformatics-applications-biopharma-2025
36. IntuitionLabs. A comparative analysis of cheminformatics platforms. Published September 27, 2025. Accessed December 8, 2025. Available from: https://intuitionlabs.ai/articles/cheminformatics-platforms-comparison
37. Drug Discovery News. The power of structure-based drug design. Published August 20, 2025. Accessed December 8, 2025. Available from: https://www.drugdiscoverynews.com/structure-based-drug-design-15997
38. Almagro D, Martínez-Jiménez F, López-Bigas N. Assessing molecular docking tools: understanding drug discovery and design. Future J Pharm Sci. 2025;11:33. doi:10.1186/s43094-025-00862-y
39. Baaden M, Glowacki DR. Virtual reality in drug design: benefits, applications and perspectives. Curr Opin Struct Biol. 2025;92:103044. doi:10.1016/j.sbi.2025.103044
40. Deepmirror. Top drug discovery software solutions to watch in 2025. Accessed December 8, 2025. Available from: https://www.deepmirror.ai/post/top-drug-discovery-software-solutions-to-watch-in-2025
41. Verma VV, Singh S, Sharma P, Kumar A. A comprehensive review on structural insights through molecular visualization: tools, applications, and limitations. J Mol Model. 2025;31(6):173. doi:10.1007/s00894-025-06402-y
42. Takeda S, Kanai S, Tsuda K, Ohta M, Tokiwa H. AI-driven synthetic route design incorporated with retrosynthesis knowledge. J Chem Inf Model. 2022;62(3):1166–1175. doi:10.1021/acs.jcim.1c01074
43. ChemCopilot. AI retrosynthesis tools: revolutionizing organic chemistry and drug discovery. Published April 23, 2025. Accessed December 8, 2025. Available from: https://www.chemcopilot.com/blog/ai-retrosynthesis-tools-revolutionizing-organic-chemistry-and-drug-discovery
44. Schomburg L, Häse F, Roch LM, Aspuru-Guzik A. Accelerating compound synthesis in drug discovery: the role of digitalisation and automation. Chem Soc Rev. 2025. doi:10.1039/D5CS00678C
45. MedChemExpress. AI as the new drug discovery accelerator: how it's reshaping the future of medicine. Published December 2025. Accessed December 8, 2025. Available from: https://medium.com/@medchemexpress_llc/ai-as-the-new-drug-discovery-accelerator
46. Marbán -González A, Gómez-Tamayo JC, Cordomi A. Exploiting PubChem and other public databases for virtual screening in 2025: what are the latest trends? Expert Opin Drug Discov. 2025;20(11):1387–1403. doi:10.1080/17460441.2025.2558161
47. Türková A, Zdrazil B. A ligand-based computational drug repurposing pipeline using KNIME and programmatic data access. J Cheminform. 2020;12:71. doi:10.1186/s13321-020-00474-z
48. Sydow D, Burggraaff L, Szengel A, et al. Democratizing cheminformatics: interpretable chemical grouping using an automated KNIME workflow. J Cheminform. 2024;16:68. doi:10.1186/s13321-024-00894-1
49. Ramírez -Márquez CD, Medina-Franco JL. Exploring chemical databases with KNIME. Low Code Data Sci. Published 2025. Accessed December 8, 2025. Available from: https://medium.com/low-code-for-advanced-data-science/exploring-chemical-databases- with-knime
50. IntuitionLabs. A comparative analysis of cheminformatics platforms. Published September 27, 2025. Accessed December 8, 2025. Available from: https://intuitionlabs.ai/articles/cheminformatics-platforms-comparison
51. Irwin JJ, Tang KG, Young J, et al. ZINC20—a free ultralarge-scale chemical database for ligand discovery. J Chem Inf Model. 2020;60(12):6065–6073. doi:10.1021/acs.jcim.0c00675
52. Rácz A, Mihalovits LM, Beckers M, et al. The changing landscape of medicinal chemistry optimization. Nat Rev Drug Discov. 2025;24(11):870–887. doi:10.1038/s41573-025-01225-1
53. Beckers M, Stiefl N, Fechner N, Sirockin F. 25 years of small-molecule optimization at Novartis: a retrospective analysis of chemical series evolution. J Chem Inf Model. 2025;65(4):1523–1538. doi:10.1021/acs.jcim.2c00785
54. Niazi SK. Artificial intelligence in small-molecule drug discovery: a critical review of methods, applications, and real-world outcomes. Pharmaceuticals (Basel). 2025;18(9):1271. doi:10.3390/ph18091271
55. Kombo DC, LaMarche MJ. The logic of chemical optimization. J Med Chem. 2025;68(11):11572–11585. doi:10.1021/acs.jmedchem.5c00445
56. Yung-Chi C, Prusoff WH. Structural simplification: an efficient strategy in lead optimization. Acta Pharm Sin B. 2019;9(6):1104–1119. doi:10.1016/j.apsb.2019.06.008
57. Heinzke AL, Pahl A, Zdrazil B, et al. Occurrence of "natural selection" in successful small molecule drug discovery. J Med Chem. 2024;67(13):11226–11241. doi:10.1021/acs.jmedchem.4c00811

58. Kelley NW, Bhatt S, Walczak P, et al. Diffusion models at the drug discovery frontier: a review on generating small molecules versus therapeutic peptides. Biology (Basel). 2025;14(12):1665. doi:10.3390/biology14121665
59. Shurtleff VW, Layton ME, Parish CA, et al. Accelerating the hit-to-lead optimization of a SARS-CoV-2 Mpro inhibitor series by combining high-throughput medicinal chemistry and computational simulations. J Med Chem. 2025;68(8):8269–8294. doi:10.1021/acs.jmedchem.4c02941
60. Roche P, Varnek A, Morelli X. Integrated strategy for lead optimization based on fragment growing: the diversity-oriented-target-focused-synthesis approach. J Med Chem. 2018;61(14):6315–6330. doi:10.1021/acs.jmedchem.8b00653
61. Atz K, Böcker J, Gabernet G, et al. Expediting hit-to-lead progression in drug discovery through reaction prediction and multi-dimensional optimization. Nat Commun. 2025;16:1005. doi:10.1038/s41467-025-66324-4
62. Kuentz M, Holm R, Elder DP. Dose number as a tool to guide lead optimization for orally bioavailable compounds in drug discovery. J Med Chem. 2022;65(18):12138–12151. doi:10.1021/acs.jmedchem.1c01687
63. Quinn -Bohmann N, Wilmanski T, Sarmiento KR, et al. Microbial community-scale metabolic modelling predicts personalized short-chain fatty acid production profiles in the human gut. Nat Microbiol. 2024;9:2300–2313. doi:10.1038/s41564-024-01728-4
64. Sizemore N, Oliphant K, Zheng R, Martin CR, Claud EC, Chattopadhyay I. A digital twin of the infant microbiome to predict neurodevelopmental deficits. Sci Adv. 2024;10(15):eadj0400. doi:10.1126/sciadv.adj0400
65. Caenepeel C, Falony G, Machiels K, et al. Dysbiosis and associated stool features improve prediction of response to biological therapy in inflammatory bowel disease. Gastroenterology. 2024;166(3):483–495. doi:10.1053/j.gastro.2023.11.304
66. Stein -Thoeringer CK, Saini NY, Zamir E, et al. A non-antibiotic-disrupted gut microbiome is associated with clinical responses to CD19-CAR-T cell cancer immunotherapy. Nat Med. 2023;29:906–916. doi:10.1038/s41591-023-02234-6
67. Khan T, de Koning K, Endresen D, et al. TwinEco: a unified framework for dynamic data-driven digital twins in ecology. Ecol Inform. 2025;91:103407. doi:10.1016/j.ecoinf.2025.103407
68. Rehman SSU, Nasar MI, Mesquita CS, et al. Integrative systems biology approaches for analyzing microbiome dysbiosis and species interactions. Brief Bioinform. 2025;26(4):bbaf323. doi:10.1093/bib/bbaf323
69. Moysidou CM, van Niekerk DC, Stoeger V, et al. Modelling human gut-microbiome interactions in a 3D bioelectronic platform. Small Sci. 2024;4:2300349. doi:10.1002/smsc.202300349
70. Ratiner K, Ciocan D, Abdeen SK, et al. Utilization of the microbiome in personalized medicine. Nat Rev Microbiol. 2024;22:291–308. doi:10.1038/s41579-023-00998-9
71. Prins FM, Hidding IJ, Klaassen MAY, et al. Limited predictive value of the gut microbiome and metabolome for response to biological therapy in inflammatory bowel disease. Gut Microbes. 2024;16(1):2391505. doi:10.1080/19490976.2024.2391505
72. Bethlehem L, Estevinho MM, Grinspan A, et al. Microbiota therapeutics for inflammatory bowel disease: the way forward. Lancet Gastroenterol Hepatol. 2024;9(5):476–486. doi:10.1016/S2468-1253(24)00006-8
73. Bucci V, Tzen B, Li N, et al. MDSINE: microbial dynamical systems INference engine for microbiome time-series analyses. Genome Biol. 2016;17(1):121. doi:10.1186/s13059-016-0980-6