INTERNATIONAL JOURNAL OF PHARMACEUTICAL SCIENCES
[ISSN: 0975-4725; CODEN(USA): IJPS00]
Journal Homepage: https://www.ijpsjournal.com

OPEN ACCESS

Review Article

AI-Driven Drug Discovery: Integrating Machine Learning and Multi-Omics for Pharmacological Target Identification and Drug Repurposing

Omkar Gosavi*, Rutuja Argade, Samiksha Cholke, Nikita Tambe, Pratiksha Shewale

Pravara Education Society's Institute of Pharmacy, Loni, Maharashtra, India.

ARTICLE INFO

Published: 4 Nov 2026

Keywords:
Artificial intelligence, machine learning, multi-omics integration, drug discovery, target identification, drug repurposing

DOI:
10.5281/zenodo.23940069

ABSTRACT

Artificial intelligence (AI) combined with multi-omics data is creating new ways to identify therapeutic targets and find new uses for existing medicines. Traditional drug discovery takes many years, costs a great deal, and often fails before a medicine reaches patients. AI methods can examine large genomic, transcriptomic, proteomic, and metabolomic datasets and help rank disease-related molecules that may be suitable drug –targets. This review describes common machine-learning (ML) and deep-learning (DL) methods, explains the value of each omics layer, and discusses approaches for combining these data to identify targets. It also reviews drug-repurposing methods such as transcriptomic signature reversal, network analysis, and knowledge-graph link prediction, with examples from cancer, infectious disease, and cardiometabolic conditions. Key issues include data quality, model interpretation, regulatory expectations, and the need for experimental confirmation. Future work may benefit from models that combine biological mechanisms with ML, repeated measurements over time, and precision-medicine approaches. These tools can guide research, but predictions must be tested in laboratory studies and clinical trials.

INTRODUCTION

1.1 Evolution of Drug Discovery

Drug discovery has traditionally moved through a series of stages: selecting a target, finding an active compound, improving the lead compound, conducting preclinical studies, and testing it in clinical trials. This approach has produced many useful medicines, but it remains slow and expensive. Development commonly takes more than a decade, and the cost of bringing one drug to market can exceed US$2 billion. Many candidates fail, especially in areas such as cancer and neurodegenerative disease. Common reasons include weak target validation, limited relevance of animal models to

***Corresponding Author:** Omkar Gosavi
Address: Pravara Education Society's Institute of Pharmacy, Loni, Maharashtra, India
Email : gosaviomkar94@gmail.com

Relevant conflicts of interest/financial disclosures: The 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.

human disease, unexpected toxicity, and differences between patients.[2][4]

To address these problems, researchers increasingly use high-throughput experiments, omics technologies, and computational analysis to study disease at a systems level. Instead of examining one gene or protein in isolation, these methods consider how molecular networks change during disease and help identify targets that may be important in a particular biological setting. [3][1]

1.2 Rise of Artificial Intelligence in Pharmacology

AI, especially ML and DL, is now used at several stages of drug development. These methods can learn complex relationships from different kinds of information, including chemical structures, protein sequences, gene-expression patterns, and clinical outcomes. In pharmacology, common applications include:[4][2]

- Predict drug–target interactions and binding affinities
- Prioritize disease-relevant genes and pathways
- Design novel molecules with desired properties
- Forecast ADMET (absorption, distribution, metabolism, excretion, toxicity) profiles
- Stratify patients and optimize clinical trial designs[5][1]

Advances in protein-structure prediction, including AlphaFold, have also supported structure-based drug design. These tools can provide useful three-dimensional models for many proteins and help researchers examine possible binding sites.[6][2]

1.3 Multi-Omics Revolution

At the same time, multi-omics methods have made it possible to study biological systems at several levels:

- **Genomics**: DNA variation, copy number alterations, structural variants
- **Transcriptomic**: bulk and single-cell RNA-seq, spatial transcriptomic
- **Proteomics**: mass spectrometry, protein arrays, phosphoproteomics
- **Metabolomics**: LC-MS, NMR-based metabolite profiling

• epigenomics: DNA methylation, histone modifications, chromatin accessibility

Each omics layer describes a different part of disease biology, from inherited risk and gene activity to protein function and changes in metabolism. However, these datasets are often large, noisy, and affected by tissue type or experimental conditions. Careful computational methods are therefore needed to combine them reliably. [7][8][1]

1.4 Rationale for Integration: AI and Multi-Omics

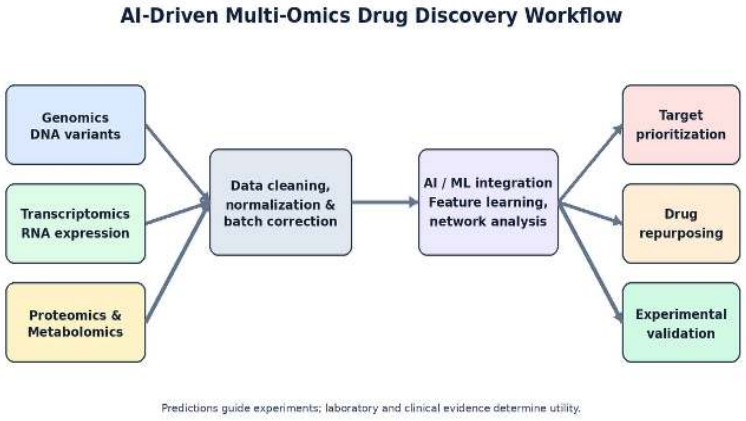


Figure 1: ***Overview of an AI-driven multi-omics workflow for target identification and drug repurposing.***

Combining AI with multi-omics can help address several limitations in drug discovery:

1. **Finding patterns in complex data:** ML and DL can detect hidden relationships within and between

omics layers that may be missed by analyses of single variables or individual pathways.[8][1]

2. **Identifying the mechanism behind an effect:** AI can connect screening results, such as changes in cell survival or appearance, with likely molecular causes by combining CRISPR experiments with transcriptomic and proteomic data.[3][8]
3. **Repurposing based on disease mechanisms:** Comparing disease-related molecular patterns with changes caused by drugs can highlight compounds that may counter disease-associated changes.[9][10][11]
4. **Grouping patients by biology:** Models that combine omics data can identify patient subgroups that may respond differently to treatment, which can support precision medicine and better-designed trials.[12][1]

This review examines how AI and multi-omics can be used together to identify pharmacological targets and suggest new uses for existing drugs. It covers key methods, examples of use, current limitations, and likely future developments.

2. Foundations of Machine Learning in Drug Discovery

2.1 Core ML/DL Architectures

AI-based drug discovery uses methods ranging from traditional statistical learning to complex neural networks.

- **Supervised learning**: Support vector machines (SVMs), random forests, gradient boosting (e.g., XGBoost), and logistic regression are widely used for classification (active vs. inactive) and regression (binding affinity, IC_{50}) tasks. These models perform well when feature engineering is robust and datasets are moderate in size.[13][14]
- **Unsupervised learning**: Clustering (k-means, hierarchical), dimensionality reduction (PCA, t-SNE, UMAP), and autoencoders help discover latent structure in omics data, identify molecular subtypes, and compress high-dimensional inputs for downstream modeling.[7][8]
- **Deep learning:** Convolutional neural networks (CNNs) process grid-like data (e.g., images, some molecular representations); recurrent neural networks (RNNs) and transformers handle sequential data (e.g., SMILES strings, protein sequences); graph neural networks (GNNs) operate directly on molecular graphs and biological networks; diffusion models and variational autoencoders enable generative design of novel compounds.[15][2][6]For structure-based tasks, three-dimensional models such as equivariant graph neural networks (GNNs) and diffusion models account for the spatial arrangement of atoms. They can help predict how a compound may bind to a protein or generate molecules that fit a binding pocket.[14][6][15]

2.2 Feature Representation

ML models depend on how chemical and biological information is represented as input features.

- **Molecular representations**: Fingerprints (e.g., Morgan/ECFP), physicochemical descriptors, SMILES strings, and graph representations (atoms as nodes, bonds as edges) are common. For generative tasks, sequence-based (SMILES) and graph-based encodings dominate.[2][15]
- Biological features: Gene expression vectors, mutation profiles, protein abundance matrices, pathway activity scores, and network embeddings serve as inputs for target identification and repurposing models. Multi-omics integration often relies on modality-specific encoders that project each data type into a shared latent space.[9][8][7]
- **Multimodal fusion**: Hybrid architectures combine chemical, structural, and biological features—e.g., GNNs for molecules paired with transformers for protein sequences, or multi-branch networks for genomics + transcriptomic + proteomics.[1][8]

A reliable model needs a training and testing plan that reflects how it will be used in practice.

Robust model development requires:

- **Data splitting:** Data may be divided randomly, by chemical scaffold, or by time. Time-based testing is particularly useful in drug-repurposing studies because it helps prevent information from later

discoveries leaking into model development.[17] [9]

- **Cross-validation and independent testing:** K-fold cross-validation can help select model settings. Performance should then be checked on an independent dataset, ideally collected by another group or during a different period.
- **Metrics**: Classification (AUC-ROC, AUC-PR, F1), regression (RMSE, MAE, R^2), ranking (early enrichment, BEDROC), and calibration metrics. In repurposing, early retrieval of known drug–disease associations is critical.[10][9]
- **Avoiding overfitting and leakage**: Careful handling of duplicate compounds, shared scaffolds, and temporally overlapping data; use of hard negatives from failed assays or clinical trials.[17]

Table 1: Machine Learning and Deep Learning Architectures Commonly Used in AI-Driven Drug Discovery

Model / method	Typical input	Drug-discovery use	Key consideration
Random forest / gradient boosting	Molecular descriptors, assay and omics features	Activity prediction, hit ranking and toxicity screening	Feature quality and data leakage can affect results.
SVM / logistic regression	Structured numerical features	Compound classification and response prediction	Scaling and suitable validation are important.
CNN	Images or grid-like molecular features	Phenotypic screening and image-pattern recognition	Requires representative training data.
RNN / transformer	SMILES, protein or biological sequences	Molecular generation and sequence modelling	Check data bias and generalisation.
Graph neural network	Molecular or biological graphs	Property prediction and drug –target interaction	Graph construction and external validation matter.
Clustering / autoencoders	High-dimensional omics and chemical profiles	Patient subgrouping and pattern discovery	Clusters need biological interpretation.

3. Multi-Omics Data Types and Their Pharmacological Relevance

3.1 Genomics

Genomic information provides evidence about inherited disease risk and possible differences in drug response.

- **GWAS and sequencing**: Genome-wide association studies (GWAS) and whole-exome/genome sequencing identify variants associated with disease risk, progression, and drug metabolism. However, most signals lie in non-coding regions, complicating causal gene assignment.[18]
- **Pharmacogenomics**: Variants in drug-metabolizing enzymes (e.g., CYP450s), transporters, and targets influence efficacy and toxicity, informing dose adjustments and patient stratification.
- **Druggability assessment**: Genomic context (e.g., essentiality, copy number, mutation frequency) helps prioritize targets likely to yield therapeutic benefit with acceptable safety.

Explainable AI methods can help connect genome-wide association study signals with regulatory regions, likely causal genes, and biological pathways. These links provide hypotheses that can be tested experimentally.[18]

3.2 Transcriptomic

Transcriptomics measures gene activity and helps describe how cells change during disease or after treatment.

- **Bulk RNA-seq**: Identifies differentially expressed genes and pathways between conditions (e.g., tumor vs. normal, treated vs. untreated).
- **Single-cell and spatial transcriptomic**: Resolve cell-type-specific and spatially resolved expression patterns, revealing heterogeneity in tumors, inflamed tissues, and neurodegenerative lesions.
- **Signature reversal**: Drug repurposing often relies on comparing disease expression signatures with drug-induced signatures; compounds that reverse

disease-associated transcriptional changes are prioritized.[11][9][10]

AI can learn useful representations of gene-expression data to estimate pathway activity, predict drug–target relationships, and rank compounds for further repurposing studies.

3.3 Proteomics

Proteomics examines proteins, which carry out many cellular functions and are common targets of medicines.

- **Mass spectrometry and arrays**: Quantify protein abundance, post-translational modifications (e.g., phosphorylation), and protein–protein interactions (PPIs).
- **PPI networks**: Map signaling and metabolic pathways; AI can identify hub nodes, bottlenecks, and context-specific modules as candidate targets.
- **Target engagement and off-target prediction**: Proteomics data inform models that predict which proteins a compound binds in cells, aiding mechanism deconvolution and safety assessment.[1][3]

3.4 Metabolomics

Metabolomics measures small molecules produced or used by cells and can show the combined effects of changes in genes, RNA, and proteins.

- **LC-MS and NMR**: Profile small molecules involved in energy metabolism, signaling, and detoxification.
- **Pathway mapping**: Altered metabolite levels indicate pathway dysregulation (e.g., glycolysis, TCA cycle, lipid metabolism) and can suggest metabolic vulnerabilities.
- **Biomarker and toxicity prediction**: Metabolite signatures correlate with drug response and adverse events, supporting patient stratification and safety profiling.[12][1]

3.5 epigenomics and Other Layers

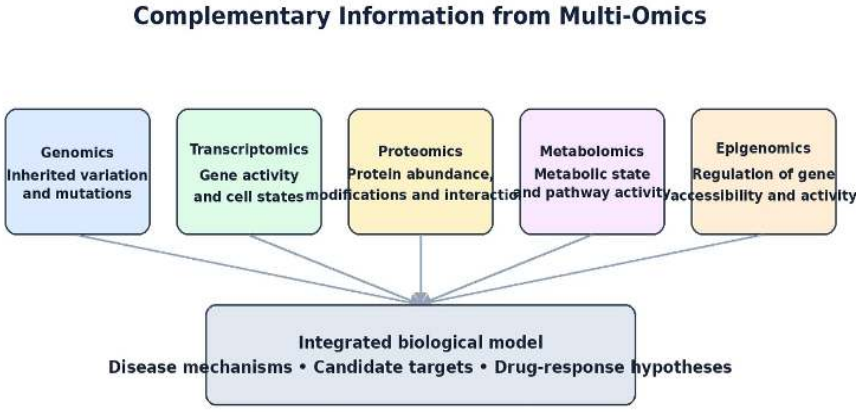


Figure 2: ***Complementary biological information captured by major omics layers.***

epigenomics examines changes that regulate gene activity without changing the DNA sequence itself.

- **DNA methylation, histone marks, chromatin accessibility**: Define cell-type-specific regulatory landscapes; dysregulation contributes to cancer, neurodegeneration, and metabolic diseases.
- **Microbiome, lipidomics, glycomics**: Emerging layers that influence drug metabolism, immune responses, and disease phenotypes.

When AI combines epigenomics and other molecular layers with genomic and transcriptomic data, it may reveal regulatory processes and highlight enzymes or other regulators as possible treatment targets.[19][20][1]

Table 2 lists the main omics layers, common measurement methods, their value in pharmacology, and examples of public databases.

Table 2: Multi-Omics Data Types, Platforms, and Pharmacological Applications

Omics layer	Common methods / data	Pharmacological relevance	Example public resources
Genomics	GWAS, exome/genome sequencing, variant calls	Disease-risk genes, target prioritisation and pharmacogenomics	GWAS Catalog; Ensembl
Transcriptomics	Bulk/single-cell RNA-seq, expression signatures	Disease profiling and drug-signature reversal	GEO; Expression Atlas
Proteomics	Mass spectrometry, protein arrays, phosphoproteomics	Target engagement, pathway activity and off-target assessment	PRIDE; UniProt
Metabolomics	LC–MS, GC–MS, NMR profiles	Metabolic pathways, biomarkers and response signals	MetaboLights; HMDB
Epigenomics	DNA methylation, histone marks, chromatin accessibility	Regulatory mechanisms and cell-state characterisation	ENCODE; GEO
Integrated multi-omics	Matched or harmonised data across molecular layers	Network-level target prioritisation and patient stratification	Choose resources by disease, assay and organism.

4. AI-Driven Target Identification

4.1 Target Deconvolution

Target deconvolution means working out which molecular changes are responsible for an observed effect, such as cell death, cell differentiation, or pathway activation.

- **Phenotypic screening + omics**: High-content imaging or viability screens are coupled with transcriptomic, proteomics, or CRISPR knockout/activation screens to link phenotypes to genes and pathways.
- **CRISPR screens + ML**: Genome-wide CRISPR loss-/gain-of-function screens identify genes whose perturbation affects drug sensitivity or disease phenotypes. ML models integrate these screens with multi-omics to prioritize causal drivers.[8][3]
- **Knowledge graphs: Biomedical knowledge graphs integrate literature, pathways, and omics to infer plausible target–disease links via graph embeddings and reasoning.[21][9][3]**

4.2 Prioritization Frameworks

AI supports systematic target prioritization by integrating multiple criteria:

- **Druggability scoring: Predictions based on protein class, structural features, ligandability, and historical success rates.**
- **Tissue specificity and safety**: Expression profiles (e.g., GTEx) and proteomics data inform whether a target is enriched in disease-relevant tissues and low in critical organs, reducing toxicity risk.
- **Multi-omics consensus**: Targets supported by convergent evidence across genomics (mutations), transcriptomic (differential expression), proteomics (abundance/phosphorylation), and metabolomics (pathway alterations) are more likely to be robust.[19][8][1]

Network methods also rank targets according to their position in a biological network, their role within a disease-related module, and the effects of changing their activity. This can provide a broader view than ranking genes one at a time.

4.3 Case Studies

4.3.1 Cancer

In cancer research, combining AI with multi-omics data has helped researchers identify several types of potential targets:

- **Oncogene addiction nodes**: Dependencies on specific kinases or transcription factors revealed by integrating mutation, expression, and phosphoproteomics data.
- **Synthetic lethality partners**: Gene pairs where co-perturbation is lethal, discovered via CRISPR screens and ML models trained on multi-omics features.
- **Immune modulation targets**: Checkpoint regulators and antigen presentation machinery components prioritized by combining tumor genomics, transcriptomics, and immune infiltrate profiling. [12][1]

4.3.2 Neurodegeneration

For Alzheimer's and Parkinson's diseases, these approaches have been used to:

- Link GWAS loci to specific cell-types (e.g., microglia, dopaminergic neurons) using single-cell transcriptomic and epigenomics.
- Identify protein aggregation pathways and proteostasis regulators as targets by integrating proteomics, interactomics, and morphological data. [18][1]

• 4.3.3 Cardiovascular and Metabolic Diseases

For cardiovascular and metabolic conditions, AI models that combine multiple omics layers have been used to:

- Prioritize lipid-metabolism enzymes and nuclear receptors based on genomics, metabolomics, and clinical biomarkers.
- Identify inflammation-related targets by integrating cytokine proteomics, transcriptomic profiles of immune cells, and genetic risk scores.[1][12]

5. AI for Drug Repurposing

5.1 Computational Strategies

- Drug repurposing looks for new medical uses for medicines that already have some safety, pharmacokinetic, or pharmacodynamics information. This may reduce the time and cost needed to investigate a new indication. AI can support this process in several ways:
- **Signature matching (transcriptomic reversal)**: Disease-associated gene-expression signatures are compared to drug-induced signatures (e.g., from CMap, LINCS). Compounds that induce opposite changes are hypothesized to counteract the disease state. Enrichment methods (e.g., GSEA) quantify reversal strength.[9][10][11]
- **Network proximity and graph mining**: Drug–target–disease networks are constructed from curated databases; graph algorithms (random walks, diffusion, meta-paths) identify drugs topologically close to disease modules.[10][21][9]
- **Knowledge graph embeddings**: Models such as node2vec, GraphSAGE, and transformer-based graph models learn low-dimensional representations of entities (drugs, proteins, diseases) and predict missing links (drug–disease associations). Temporal validation ensures generalization to future discoveries.[21][9][17]
- Combined models use molecular signatures, network relationships, and chemical similarity together. Using more than one type of evidence may improve candidate ranking and make the results easier to interpret.[11][10]

5.2 Data Sources

Public databases provide much of the information used by AI systems to identify possible new indications for existing medicines.

- **Drug and target databases**: Drug Bank, ChEMBL, Binding DB, PubChem, Therapeutic Target Database (TTD).[22][23]
- **Disease and pathway databases**: OMIM, KEGG, Reactome, DisGeNET, CORUM, GO.

- **Omics repositories**: GEO, TCGA, GTEx, ProteomeXchange, Metabolites.
- **Signature databases**: CMap, LINCS L1000, which provide gene-expression profiles for thousands of compounds across cell lines.[10][11]
- **Clinical and real-world data**: ClinicalTrials.gov, electronic health records (EHRs), pharmacovigilance databases (e.g., FAERS), and claims data for hypothesis generation and validation.[21][10]

Knowledge graphs, including resources such as PrimeKG-Plus, bring together relationships from several biomedical databases. They can be used to suggest links between drugs and diseases, including conditions that are less well studied.[21]

5.3 Success Stories

Examples of drug repurposing show how a medicine developed for one purpose may later be investigated for another:

- **Remdesivir**: Originally developed for Ebola, it was repurposed for SARS-CoV-2 based on antiviral activity and structural considerations; AI models later supported its prioritization among candidate antivirals.[24]
- **Colchicine**: An anti-gout agent repurposed for cardiovascular inflammation, with trials showing reduced events in coronary disease patients; network and pathway analyses supported its anti-inflammatory mechanism.[24]
- **Ketamine/esketamine**: From anesthetic to rapid-acting antidepressant; transcriptomic and signaling models helped elucidate NMDA receptor–dependent and independent mechanisms relevant to depression.[24]
- **Oncology repurposing**: Antibiotics, antihypertensive, and psychotropics have shown anticancer activity in preclinical and early clinical studies, often prioritized via signature reversal and network proximity approaches.[10][24]

AI-based systems have proposed new uses for existing compounds, and some candidates have moved into clinical studies. However, a prediction from an AI model is not proof that a treatment works, and regulatory approval depends on suitable evidence of quality, safety, and efficacy.[25]

5.4 Validation Pipelines

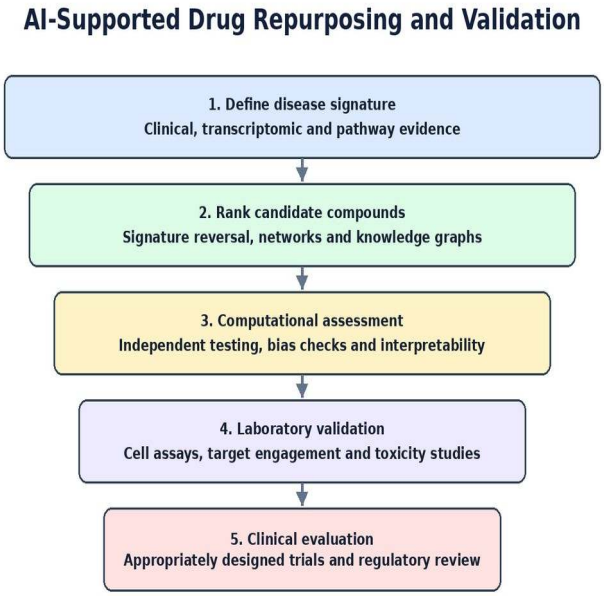


Figure 3: ***A proposed sequence for computational prioritization and validation of repurposing candidates.***

Predictions from computer models need to be checked through laboratory and clinical research.

- **In silico → in vitro**: Top candidates are tested in cell-based assays measuring viability, pathway modulation, and phenotypic readouts.
- **In vivo models**: Rodent or alternative models assess efficacy, PK/PD, and toxicity in disease-relevant contexts.
- **Clinical trials**: Repurposed drugs may enter phase II/III trials, often with biomarker-guided patient selection. Adaptive designs and real-world evidence can accelerate evaluation.[11][24]

Testing a model on newer data than those used for training, as well as on independent datasets, is important for estimating whether it will perform well in real settings.[9][17]

Table 3 outlines major computational approaches to drug repurposing, including their main concepts, data inputs, methods, benefits, and limitations.

Strategy	Data / method	Potential benefit	Limitation / validation
Transcriptomic signature reversal	Compare disease and drug-induced gene-expression signatures	Prioritises compounds that may counter disease-associated changes	Cell type, dose and assay context affect signatures; confirm experimentally.
Network proximity	Drug–target–disease networks and graph algorithms	Finds compounds connected to disease-related modules	Network coverage and edge reliability can bias ranking.
Knowledge-graph link prediction	Graph embeddings or graph neural networks over biomedical entities	Suggests missing drug–disease or drug–target relationships	Temporal and independent validation are needed.
Chemical similarity	Fingerprints, molecular descriptors and similarity metrics	Identifies compounds related to known active molecules	Similar structures may have different efficacy or toxicity.
Integrated evidence ranking	Combine signatures, networks and chemical features	Uses complementary evidence to rank candidates	Avoid data leakage and document weighting and uncertainty.

6. Integration Frameworks and Tools

6.1 Multi-Omics Integration Methods

Methods for combining multi-omics data are often grouped according to the stage at which the data are joined.

- **Early fusion**: Raw or preprocessed features from multiple omics layers are concatenated into a single input matrix. Simple but sensitive to modality imbalance and batch effects.[26][7][8]
- **Intermediate fusion**: Modality-specific encoders (e.g., autoencoders, GNNs) learn latent representations that are then combined, allowing the model to capture non-linear cross-talk while preserving layer-specific structure. Methods like MOGONET and OmiEmbed exemplify this approach.[16][8]
- **Late fusion**: Separate models are trained per omics layer, and their predictions or decision scores are aggregated (e.g., via voting, stacking). This preserves modality-specific patterns but may miss deeper interactions.[27][7][8]

Recent reviews describe growing use of graph-based and hybrid methods to combine multi-omics data, including in cancer research. Researchers must balance model performance with interpretability and decide how to handle samples that are missing one or more data types.[20][7][8]

6.2 AI Platforms and Software

Both open-source software and commercial platforms are available for AI-based drug discovery.

- **Open-source**:

DeepChem, PyTorch Geometric, DGL-LifeSci for molecular DL and GNNs

scikit-learn, XGBoost for classical ML

MOGONET, OmiEmbed, and custom PyTorch/TensorFlow pipelines for multi-omics

- **Commercial**:

Platforms from Insilco Medicine, BenevolentAI, Atomwise, and others offer end-to-end AI workflows for target identification, generative design, and ADMET prediction.[5][2]

Container tools such as Docker and Singularity, workflow systems such as Nextflow and Snakemake, and cloud services such as AWS, GCP, and Azure can make analyses easier to scale and repeat.[28]

6.3 FAIR Data and Interoperability

The FAIR principles—Findable, Accessible, Interoperable, and Reusable—provide a framework for managing data so that other researchers can locate and work with them.

- **Findable/Accessible**: Depositing data in public archives (GEO, ProteomeXchange, MetaboLights) with rich metadata and persistent identifiers.[29][28]
- **Interoperable**: Using standard ontologies (Gene Ontology, ChEBI, Mesh, HPO) and controlled vocabularies to annotate samples, assays, and variables.
- **Reusable**: Providing detailed pipeline descriptions, version-controlled code, and container images to enable reanalysis and extension.[28][29]

Regulators increasingly emphasize well-documented data and transparent analytical workflows when AI and omics evidence are used to support scientific conclusions.[30][28]

7. Challenges and Limitations

7.1 Data-Related Issues

Despite progress in this field, data quality and consistency remain important concerns.

- **Heterogeneity and batch effects**: Omics data from different platforms, labs, or cohorts exhibit technical variability that can confound AI models. Normalization, batch correction, and domain adaptation techniques are essential but not foolproof.[7][8]
- **Missing data and modality imbalance**: Not all samples have all omics layers measured; models must handle missing modalities without bias.
- **Bias in training sets**: Over-representation of certain populations, diseases, or chemical spaces

limits generalizability and may exacerbate health disparities.[2][18]

- **Privacy and consent**: Clinical and real-world data raise ethical and legal concerns regarding patient privacy, especially when linked to omics profiles. [28][2]

Model design and interpretation also raise several concerns.

AI models face scrutiny regarding:

- **Black-box behavior**: Deep networks often lack transparency, hindering trust and mechanistic insight. Explainable AI (XAI) methods (e.g., SHAP, attention maps, gradient-based attributions) aim to highlight important features but can be inconsistent across methods.[31][32][33]
- **Correlation vs. causation**: Many models learn associative patterns rather than causal mechanisms, risking spurious target hypotheses. Causal inference frameworks and perturbation-based validation are needed to strengthen claims.[33][18]
- **Reproducibility and benchmarking**: Lack of standardized benchmarks, data splits, and evaluation metrics complicates comparison across studies. Community efforts are underway to establish best practices.[14][2]

Moving from a promising computational result to a useful treatment in clinical practice remains difficult.

Bridging computational predictions to clinical impact remains challenging:

Challenge	Impact on discovery	Potential mitigation
Heterogeneous data and batch effects	Technical variation can obscure disease signals	Harmonies metadata, normalise data and assess batch effects.
Missing modalities and dataset bias	Models may generalise poorly across cohorts or populations	Report missingness and applicability; use diverse external validation sets.
Limited interpretability	Predictions may be difficult to connect to biological mechanisms	Use appropriate explainability methods and expert review.
Correlation mistaken for causation	Spurious associations may be prioritised as targets	Use perturbation studies and causal designs where appropriate.
Data leakage and weak reproducibility	Performance may appear better than real-world performance	Use scaffold/time-aware splits, independent tests and versioned workflows.
Clinical translation and regulation	Predictions alone cannot establish efficacy or safety	Require experimental evidence and fit-for-purpose regulatory review.

- **Experimental validation**: In silico hits must be confirmed in biologically relevant models; discrepancies between cell lines, organoids, and animal models can obscure true efficacy.
- **Species differences**: Human disease mechanisms may not be fully recapitulated in model organisms, limiting translation.
- **Clinical trial design**: Repurposed or AI-identified therapies require carefully designed trials, often with biomarker-guided enrollment, to demonstrate benefit in specific subpopulations.[11][24]
- **Regulatory adaptation**: While FDA and EMA have issued guiding principles for AI in drug development, formal frameworks for AI-generated evidence in early discovery are still evolving.

8. Future Directions

8.1 Hybrid and Causal Modeling

Future approaches are likely to bring together biological models and data-driven methods.

- **Mechanistic models**: Ordinary differential equations (ODEs), Boolean networks, and agent-based models that encode known biology.
- **ML/DL components**: Flexible learners that capture residual patterns and high-dimensional interactions.
- **Causal graphs and counterfactual reasoning**: Frameworks that explicitly model interventions (e.g., gene knockouts, drug treatments) to infer causal effects and prioritize targets with higher confidence.[33][18]

These combined methods fit within systems pharmacology, which studies and predicts drug effects across connected biological processes.

8.2 Multimodal and Longitudinal Data

New data sources may help AI models represent disease and treatment response in greater detail.

- **Wearables and digital phenotyping**: Continuous monitoring of physiological parameters (heart rate, activity, sleep) provides real-time, patient-specific data.
- **Imaging and radiomics**: MRI, CT, and PET scans, analyzed via CNNs, reveal structural and functional biomarkers.
- **Time-series omics**: Repeated sampling (e.g., during treatment) enables dynamic modeling of response and resistance mechanisms.
- **Real-world evidence**: EHRs, claims, and patient-reported outcomes integrated with omics can refine repurposing hypotheses and support post-marketing surveillance.[28][1]

8.3 Personalized and Precision Medicine

AI combined with multi-omics may help match treatments to groups of patients with similar biological features.

- **Molecular subtyping**: Clustering patients by multi-omics profiles to identify subgroups with distinct target dependencies and drug sensitivities. [19][12]
- **N-of-1 and adaptive trials**: Using individual patient data to tailor treatment sequences and doses, guided by AI predictions of response and toxicity.
- **Equity and access**: Ensuring diverse representation in training data and equitable deployment of AI tools to avoid widening health disparities.[29] [18]

8.4 Regulatory and Industry Adoption

Regulatory agencies are developing approaches for the responsible use of AI in drug development.

- **FDA and EMA guiding principles**: Ten principles emphasize human-centric design, risk-based approaches, data governance, model lifecycle management, and transparency. While not binding, they set expectations for AI use across the product lifecycle, including nonclinical stages.[30]
- **Context of use and credibility**: AI models must define their intended application and undergo risk-proportionate validation; high-impact uses (e.g., supporting efficacy claims) require more rigorous evidence.
- **Public–private partnerships**: Collaborative initiatives (e.g., ICH, consortia) aim to harmonize standards, share benchmarks, and accelerate responsible AI adoption in pharma R&D.[30]

As AI-supported candidates progress through clinical development, regulatory guidance will need to address more clearly how model-based evidence should be assessed.[25]

CONCLUSION

Combining AI with multi-omics is moving drug-target discovery and drug repurposing beyond single-gene or single-data-layer studies toward a broader view of disease biology. ML and DL methods can identify patterns across genomic, transcriptomic, proteomic, and metabolomic data and help rank targets or existing medicines for further study. Important limitations remain, including inconsistent data, difficulty explaining model predictions, and the challenge of translating computational findings into clinical benefit. Progress in hybrid models, multimodal analysis, explainable AI, and regulatory science may improve the usefulness of these methods. Close cooperation between computational researchers, pharmacologists, clinicians, and regulators will be needed to determine whether AI-supported discoveries can become safe and effective treatments.

REFERENCES

1. Vamathevan J, Clark D, Czodrowski P, Dunham I, Ferran A, Lee G, et al. Applications of machine learning in drug discovery and development. Nat Rev Drug Discov. 2019;18 (6):463-477. Doi:10.1038/s41573-019-0024-5.

2. Jumper J, Evans R, Pritzel A, Green T, Figurnov M, Ronneberger O, et al. Highly accurate protein structure prediction with AlphaFold. Nature. 2021;596 (7873):583-589. Doi:10.1038/s41586-021-03819-2.
3. Baek M, DiMaio F, Anishchenko I, Dauparas J, Ovchinnikov S, Lee GR, et al. Accurate prediction of protein structures and interactions using a three-track neural network. Science. 2021;373 (6557):871-876. Doi:10.1126/science.abj8754.
4. Zhang O, Lin H, Zhang X, Wang X, Wu Z, Ye Q, et al. Graph neural networks in modern AI-aided drug discovery. Chem Rev. 2025;125 (20):10001-10103. Doi:10.1021/acs.chemrev.5c00461.
5. Gaudelet T, Day B, Jamasb AR, Soman J, Regep C, Liu G, et al. Utilizing graph machine learning within drug discovery and development. Brief Bioinform. 2021;22 (6):bbab159. Doi:10.1093/bib/bbab159.
6. Zhavoronkov A, Ivanenkov YA, Aliper A, Veselov MS, Aladinskiy VA, Aladinskaya AV, et al. Deep learning enables rapid identification of potent DDR1 kinase inhibitors. Nat Biotechnol. 2019;37 (9):1038-1040. Doi:10.1038/s41587-019-0224-x.
7. Méndez-Lucio O, Baillif B, Clevert DA, Rouquié D, Wichard J. De novo generation of hit-like molecules from gene-expression signatures using artificial intelligence. Nat Commun. 2020;11 (1):10. Doi:10.1038/s41467-019-13807-w.
8. Ekins S, Puhl AC, Zorn KM, Lane TR, Russo DP, Klein JJ, et al. Exploiting machine learning for end-to-end drug discovery and development. Nat Mater. 2019;18 (5):435-441. Doi:10.1038/s41563-019-0338-z.
9. Kearnes S, McCloskey K, Berndl M, Pande V, Riley P. Molecular graph convolutions: moving beyond fingerprints. J Comput Aided Mol Des. 2016;30 (8):595-608. Doi:10.1007/s10822-016-9938-8.
10. Wu Z, Ramsundar B, Feinberg EN, Gomes J, Geniesse C, Pappu AS, et al. MoleculeNet: a benchmark for molecular machine learning. Chem Sci. 2018;9 (2):513-530. Doi:10.1039/C7SC02664A.
11. Hasin Y, Seldin M, Lusis A. Multi-omics approaches to disease. Genome Biol. 2017;18 (1):83. Doi:10.1186/s13059-017-1215-1.
12. Subramanian I, Verma S, Kumar S, Jere A, Anamika K. Multi-omics data integration, interpretation, and its application. Bioinform Biol Insights. 2020;14:1177932219899051. Doi:10.1177/1177932219899051.
13. Huang S, Chaudhary K, Garmire LX. More is better: recent progress in multi-omics data integration methods. Front Genet. 2017;8:84. Doi:10.3389/fgene.2017.00084.
14. Olowolayemo A, Souag A, Sirlantzis K, Turner S, Wilson C. A review of multi-omics integration techniques across five machine learning method families. Bioinform Adv. 2026;6 (1):vbag108. Doi:10.1093/bioadv/vbag108.
15. Rappoport N, Shamir R. Multi-omic and multi-view clustering algorithms: review and cancer benchmark. Nucleic Acids Res. 2018;46 (20):10546-10562. Doi:10.1093/nar/gky889.
16. Argelaguet R, Arnol D, Bredikhin D, Deloro Y, Velten B, Marioni JC, Stegle O. MOFA+: a statistical framework for comprehensive integration of multi-modal single-cell data. Genome Biol. 2020;21 (1):111. Doi:10.1186/s13059-020-02015-1.
17. Wishart DS. Emerging applications of metabolomics in drug discovery and precision medicine. Nat Rev Drug Discov. 2016;15 (7):473-484. Doi:10.1038/nrd.2016.32.
18. Uhlén M, Fagerberg L, Hallström BM, Lindskog C, Oksvold P, Mardinoglu A, et al. Proteomics: tissue-based map of the human proteome. Science. 2015;347 (6220):1260419. Doi:10.1126/science.1260419.
19. Weinstein JN, Collisson EA, Mills GB, Shaw KRM, Ozenberger BA, Ellrott K, et al. The Cancer Genome Atlas Pan-Cancer analysis project. Nat Genet. 2013;45 (10):1113-1120. Doi:10.1038/ng.2764.

20. Barretina J, Caponigro G, Stransky N, Venkatesan K, Shaw K, Margolin AA, et al. The Cancer Cell Line Encyclopedia enables predictive modelling of anticancer drug sensitivity. Nature. 2012;483 (7391):603-607. Doi:10.1038/nature11003.
21. Ghandi M, Huang FW, Jané-Valbuena J, Kryukov GV, Lo CC, McDonald ER 3rd, et al. Next-generation characterization of the Cancer Cell Line Encyclopedia. Nature. 2019;569 (7757):503-508. Doi:10.1038/s41586-019-1186-3.
22. Tsherniak A, Vazquez F, Montgomery PG, Weir BA, Kryukov G, Cowley GS, et al. Defining a cancer dependency map. Cell. 2017;170 (3):564-576.e16. doi:10.1016/j.cell.2017.06.010.
23. Behan FM, Iorio F, Picco G, Gonçalves E, Beaver CM, Migliardi G, et al. Prioritization of cancer therapeutic targets using CRISPR-Cas9 screens. Nature. 2019;568 (7753):511-516. Doi:10.1038/s41586-019-1103-9.
24. Pushpakom S, Iorio F, Eyers PA, Escott KJ, Hopper S, Wells A, et al. Drug repurposing: progress, challenges and recommendations. Nat Rev Drug Discov. 2019;18 (1):41-58. Doi:10.1038/nrd.2018.168.
25. Lamb J, Crawford ED, Peck D, Modell JW, Blat IC, Wrobel MJ, et al. The Connectivity Map: using gene-expression signatures to connect small molecules, genes, and disease. Science. 2006;313 (5795):1929-1935. Doi:10.1126/science.1132939.
26. Subramanian A, Narayan R, Corsello SM, Peck DD, Natoli TE, Lu X, et al. A next generation connectivity map: L1000 platform and the first 1,000,000 profiles. Cell. 2017;171 (6):1437-1452.e17. doi:10.1016/j.cell.2017.10.049.
27. Keiser MJ, Setola V, Irwin JJ, Laggner C, Abbas AI, Hufeisen SJ, et al. Predicting new molecular targets for known drugs. Nature. 2009;462 (7270):175-181. Doi:10.1038/nature08506.
28. Luo H, Wang J, Li M, Luo J, Peng X, Wu FX, Pan Y. Drug repositioning based on comprehensive similarity measures and bi-random walk algorithm. Bioinformatics. 2016;32 (17):2664-2671. Doi:10.1093/bioinformatics/btw228.
29. Himmelstein DS, Lizee A, Hessler C, Brueggeman L, Chen SL, Hadley D, et al. Systematic integration of biomedical knowledge prioritizes drugs for repurposing. eLife. 2017;6:e26726. Doi:10.7554/eLife.26726.
30. Zeng X, Zhu S, Lu W, Liu Z, Huang J, Zhou Y, et al. Target identification among known drugs by deep learning from heterogeneous networks. Chem Sci. 2020;11 (7):1775-1797. Doi:10.1039/C9SC04336E.
31. Zitnik M, Agrawal M, Leskovec J. Modeling polypharmacy side effects with graph convolutional networks. Bioinformatics. 2018;34 (13):i457-i466. Doi:10.1093/bioinformatics/bty294.
32. Chandak P, Huang K, Zitnik M. Building a knowledge graph to enable precision medicine. Sci Data. 2023;10:67. Doi:10.1038/s41597-023-01960-3.
33. Chen F, Zhao R, Han X, Li H, Tang Z. Harnessing AI to fuse phenotypic signatures for drug-target identification: progress in computational modeling. Brief Bioinform. 2026;27 (1):bbag045. Doi:10.1093/bib/bbag045.
34. Bender A, Thomas MC, Scannell JW, Shaywitz DA, Ghiandoni GM, Greener JG, Pruteanu LL, Jacobson RD, Handa K, Hirano M, Seal S. Artificial intelligence in drug discovery—what it is, where we stand and the path forward. Nature Reviews Drug Discovery. 2026 Aug 7:1-22.
35. He J, Lai H, Saigiridharan L, Ghiandoni GM, Jenei K, Gokalp U, Nukovic A, Engkvist O, Janet JP, Genheden S. Democratising real-world drug discovery through agentic AI. Drug discovery today. 2026 Jan 16:104605.
36. Kantesaria R, Panda HS. A review on AI-based data-driven models for optimization of nanocarriers as drug delivery systems. ACS Biomaterials Science & Engineering. 2026 Feb 11;12 (3):1397-418
37. Salvati A, Melone V, Giordano A, Lamberti J, Palumbo D, Palo L, Rea D, Memoli D, Simonis V, Alexandrova E, Silvestro F. Multi-omics based and AI-driven drug repositioning for epigenetic

therapy in female malignancies. Journal of Translational Medicine. 2025 Jul 25;23 (1):837.

38. Othman ZK, Ahmed MM, Kasimieh O, Musa SS, Branda F, Cue EG, Ocampo JM, Prisno III DE, Vimolmangkang S. Artificial intelligence for natural product drug discovery and development: current landscape, applications, and future directions. Intelligence-Based Medicine. 2025 Nov 26:100316.
39. Ripa JD, Ali S, Field M, Smithson J, Wangchuk P. From AI-assisted in silico computational design to preclinical in vivo models: a multi-platform approach to small molecule anti-IBD drug discovery. Pharmaceuticals. 2025 Oct 13;18 (10):1536.
40. Branda F, Ahmed MM, Ciccozzi M, Guzzi PH, Scarpa F. The next paradigm in bioinformatics: a review of multi-agent systems and foundational models for end-to-end scientific discovery. Briefings in Bioinformatics. 2026 May;27 (3):bbag245

HOW TO CITE: Omkar Gosavi, Rutuja Argade, Samiksha Cholke, Nikita Tambe, Pratiksha Shewale, AI-Driven Drug Discovery: Integrating Machine Learning and Multi-Omics for Pharmacological Target Identification and Drug Repurposing, Int. J. of Pharm. Sci., 2026, Vol 4, Issue 11, 790-804. https://doi.org/10.5281/zenodo.23940069