Pharmacophore

ISSN-2229-5402

Journal home page: http://www.pharmacophorejournal.com

EXPLAINABLE DRUG REPURPOSING MODELS USING TRANSCRIPTOMICS, DRUG SIGNATURES, PATHWAYS, AND TARGET NETWORKS

Carlos Ramirez[1]*, Elena Torres[2], Pablo Ortega[1], Sofia Mendes[3]

1. ***Department of Pharmaceutical AI and Drug Analytics, Faculty of Pharmacy, University of Barcelona, Barcelona, Spain.***
2. ***Department of Computational Drug Sciences, Faculty of Pharmacy, University of Lisbon, Lisbon, Portugal.***
3. ***Department of Intelligent Pharmaceutical Systems, Faculty of Pharmacy, University of Porto, Porto, Portugal.***

ARTICLE INFO

Received:
10 March 2025
Received in revised form:
02 June 2025
Accepted:
03 June 2025
Available online:
28 June 2025

Keywords: Explainable AI, Drug repurposing, Transcriptomics, Drug signatures, Pathway enrichment, Target networks

ABSTRACT

Drug repurposing can accelerate the transition from biological hypothesis to therapeutic evaluation by leveraging compounds with existing pharmacological knowledge; however, many computational predictions remain challenging to act upon because their underlying biological rationale is not explicit. Current repurposing models often treat drug–disease associations as black-box predictions, limiting their utility for biologists and clinicians who need to understand the pathways, targets, or transcriptomic relationships supporting a proposed indication. To address this, an explainable machine learning model can be designed to predict drug repurposing opportunities while providing transparent biological explanations for each prediction, highlighting influential transcriptomic signatures, pathway signals, and target proteins. Such a multi-modal model could integrate disease expression profiles, drug perturbation signatures, pathway enrichment features, and protein–protein interaction network proximity, with SHAP-based attribution and pathway-level attention decomposing predictions into interpretable biological components. Conceptually, the system would output a ranked set of drug–disease pairs alongside evidence narratives that specify the relevant pathways, target proteins, and directions of transcriptomic reversal, rendering each prediction biologically plausible. By providing interpretable insights, an explainable repurposing model would transform computational repositioning from a mere screening exercise into a hypothesis-generation framework, enabling scientists to prioritize predictions for experimental or translational follow-up.

To Cite This Article: Ramirez C, Torres E, Ortega P, Mendes S. Explainable Drug Repurposing Models Using Transcriptomics, Drug Signatures, Pathways, and Target Networks. Pharmacophore. 2025;16(3):32-41. https://doi.org/10.51847/0ih6F5zcFs

Introduction

Drug repurposing is strategically attractive because it can generate therapeutic hypotheses from existing pharmacological knowledge rather than beginning with entirely new chemical entities. Curated repurposing resources and biomedical knowledge graphs have shown how known drug properties, disease associations, and molecular evidence can be systematically organized for computational inference [1, 2]. Yet a predicted drug–disease association alone is rarely sufficient to motivate experimental follow-up, because translational researchers need to know why a candidate is plausible before investing resources. This creates a strong need for repurposing models that return mechanistic evidence rather than only ranked candidates.

The biological data available for repurposing are unusually rich, spanning disease transcriptomic profiles, drug-induced gene expression signatures, pathway databases, target annotations, and protein interaction networks. The Connectivity Map and L1000 resources established a large-scale foundation for comparing disease states with drug perturbation profiles [3], while LINCS-oriented systems resources have expanded the use of cellular response signatures as reusable computational features [4]. Drug signature platforms such as L1000FWD further support the idea that perturbational transcriptomics can be queried as a mechanistic space for candidate discovery [5]. However, without interpretability, these signals may remain detached from the pathway and target biology that would make a prediction actionable.

Corresponding Author: Carlos Ramirez; Department of Pharmaceutical AI and Drug Analytics, Faculty of Pharmacy, University of Barcelona, Barcelona, Spain. E-mail: carlos.ramirez@gmail.com

Recent progress in explainable artificial intelligence makes it possible to attribute predictions to genes, targets, pathways, graph paths, or latent biological features. Network-based deep learning models have been proposed for in silico repositioning [6], while interpretable knowledge-graph and reasoning-path approaches demonstrate that model outputs can be connected to biological entities rather than treated as opaque scores [7, 8]. Explainability is especially important in systems pharmacology because drug action is distributed across molecular networks and cellular programs, not confined to a single isolated feature. A useful XAI model should therefore translate statistical evidence into biological reasoning that medicinal chemists, pharmacologists, and disease experts can examine.

The central thesis of this article is that an explainable repurposing model should predict drug–disease relationships and simultaneously provide a pathway- and target-level rationale for each prediction. Such a model would combine transcriptomic reversal, pathway enrichment, target-network proximity, and knowledge-graph context into a unified decision process [9, 10]. Its explanation layer would identify which disease genes, perturbation signatures, targets, and biological pathways contributed most to the predicted opportunity. This design would help bridge the gap between computational screening and experimental validation by making predictions easier to critique, prioritize, and refine.

Background

Principles of Drug Repurposing

Computational drug repurposing includes signature matching, network-based prioritization, literature mining, knowledge-graph completion, and supervised learning over known drug–disease relationships. Signature-based approaches compare drug-induced molecular changes with disease-associated changes, while network approaches examine how drug targets relate to disease modules in the interactome [11, 12]. Literature-derived and knowledge-graph methods add another layer by connecting drugs and diseases through genes, pathways, phenotypes, and biomedical concepts [13, 14]. Across these strategies, biological plausibility remains essential because a candidate that is statistically ranked but mechanistically unexplained is unlikely to inspire confidence.

Transcriptomic Signatures and Drug Perturbation Data

Transcriptomic signatures provide a systematic way to represent both disease states and drug-induced cellular responses. The Connectivity Map, L1000 platform, and related perturbational resources make it possible to ask whether a compound could reverse a disease-associated expression program [3, 5]. Deep representation learning has also been used to encode gene expression profiles for drug repurposing, suggesting that disease and compound signatures can be compared in a learned molecular space [15]. Still, interpretation is complicated by context dependence, including cell type, perturbation conditions, platform differences, and batch effects.

Biological Pathways and Gene Set Enrichment

Pathway and gene set enrichment analyses help convert high-dimensional gene-level measurements into interpretable biological themes. Instead of asking whether thousands of genes individually support a repurposing prediction, a pathway-aware model can ask whether immune signaling, metabolic regulation, cell-cycle control, or stress-response programs are directionally altered. Pathway-guided neural models and enrichment-oriented drug response frameworks show how structured biological knowledge can constrain or explain machine learning outputs [16, 17]. This makes pathway-level features valuable both for prediction and for communicating the biological logic behind a candidate.

Target Networks and Proximity in the Interactome

Target-network methods rest on the idea that effective drugs often act near disease-associated proteins within the human interactome. Network medicine studies of repurposing have used proximity between drug targets and disease modules to reason about candidate mechanisms [11, 12]. Integrative frameworks can combine target networks with expression, pathway, and clinical knowledge to generate mechanistically grounded hypotheses [18, 19]. In an explainable model, the same network features should not only influence prediction but also reveal which target–disease relationships make the candidate plausible.

Explainable AI for Biological Discovery

Explainable AI methods can help transform model outputs into biological interpretations by attributing predictions to features, graph edges, pathways, or molecular entities. Knowledge-guided graph neural networks, transformer-based architectures, and visible neural models have been used to make drug response or drug activity prediction more interpretable [20, 21]. Studies comparing interpretable and non-interpretable architectures also highlight that transparency must be evaluated carefully rather than assumed from model design alone [22]. For drug repurposing, the key challenge is not merely to produce explanations, but to ensure that the explanations are biologically coherent and useful for downstream decision-making.

Model Development Overview

High-Level Framework

The proposed framework would ingest disease transcriptomic signatures, drug perturbation signatures, pathway enrichment scores, target-network features, and known drug–target relationships. A supervised learning module could estimate the likelihood that a drug–disease pair represents a plausible repurposing opportunity, drawing inspiration from network-based

deep learning and knowledge-graph approaches [6, 10]. A separate explanation module would then decompose the prediction into transcriptomic, pathway, target, and graph-based contributions. The goal is to make the model's reasoning visible without claiming that computational evidence alone establishes therapeutic efficacy.

Figure 1 illustrates the proposed explainable drug repurposing architecture, showing how transcriptomic signatures, pathway enrichment, and target-network evidence are transformed into ranked drug–disease hypotheses with biologically interpretable explanations.

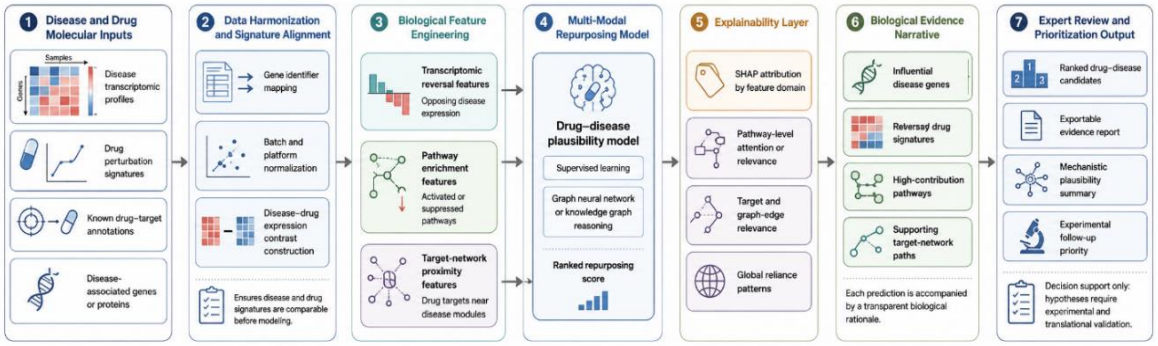

Figure 1. Explainable Drug Repurposing Architecture Using Transcriptomics, Drug Signatures, Pathways, and Target Networks.

Core Input Features

The core feature set would include disease-associated differentially expressed genes, drug-induced expression changes, pathway enrichment summaries, and protein-network proximity between drug targets and disease proteins. Transcriptomic features would encode whether a drug perturbation is expected to oppose or reinforce disease biology, following the broader logic of perturbational signature matching [3, 15]. Pathway features would summarize coordinated biological programs, while target features would describe whether a compound's known targets sit close to disease-relevant proteins in the interactome [11]. Known drug–target annotations and repurposing resources would provide structured supervision and biological context for training [1].

Design Principles

The model should output more than a repurposing score; it should also produce a ranked explanation identifying influential pathways, targets, genes, and transcriptomic contrasts. Knowledge-graph completion and reasoning-path models show that drug–disease predictions can be accompanied by interpretable routes through biological entities [2, 8]. This principle can be extended by presenting directional contributions, such as whether reversal of an inflammatory pathway increases support for the prediction or whether target-network proximity drives the score. The explanation should be understandable to a biologist while remaining faithful to the model's actual decision process.

Data Sources and Feature Engineering

Assembling Disease and Drug Transcriptomic Profiles

Disease expression profiles could be curated from public repositories, disease-focused studies, or cancer transcriptomic resources, while drug perturbation profiles could be drawn from Connectivity Map, LINCS L1000, or related drug-signature collections. Harmonization would require consistent gene identifiers, comparable expression contrasts, and careful normalization before disease and drug signatures are compared [3, 4]. Query tools such as L1000FWD illustrate how perturbational profiles can be organized for interpretable exploration of drug-induced molecular states [5]. Feature engineering should preserve both gene-level directionality and higher-level biological context so that later explanations remain traceable.

Table 1 defines the main evidence domains that should be integrated into the proposed explainable repurposing model and clarifies how each domain contributes separately to prediction and biological interpretation.

Table 1. Multi-Modal Evidence Domains for Explainable Drug Repurposing Prediction

Evidence domain	Primary biological question addressed	Representative input features	How the domain contributes to prediction	How the domain supports explanation	Main interpretive risk
Disease transcriptomic signature	What molecular program characterizes the disease state?	Differentially expressed genes, disease-associated expression contrasts, tissue- or disease-	Defines the disease molecular state against which candidate drugs are compared	Identifies genes and biological programs that make the disease representation interpretable	Disease signatures may vary by tissue, cohort, platform, disease stage, or preprocessing method

		specific expression profiles			
Drug perturbation signature	What molecular changes does a compound induce?	Drug-induced expression changes, L1000 or Connectivity Map perturbation profiles, dose- and time-dependent cellular responses	Allows comparison between disease biology and compound-induced molecular reversal or reinforcement	Shows whether the drug plausibly opposes disease-associated expression patterns	Cell-line context, dose, exposure time, and assay conditions may not match the disease setting
Transcriptomic reversal features	Does the drug oppose the disease-associated expression pattern?	Directional reversal scores, gene-level concordance or discordance, disease–drug contrast vectors	Provides a core repurposing signal by estimating whether a drug may counteract disease biology	Explains predictions through specific reversed genes or gene programs	Apparent reversal may reflect technical artifacts or non-disease-specific stress responses
Pathway enrichment features	Which biological processes organize the gene-level evidence?	Enriched pathways, gene-set activity scores, disease-activated and drug-suppressed pathway contrasts	Reduces high-dimensional gene data into biologically meaningful predictive units	Enables pathway-level explanations such as inflammatory, metabolic, cell-cycle, or stress-response mechanisms	Pathway databases are incomplete, overlapping, and biased toward well-studied mechanisms
Drug–target annotations	Which proteins are directly or indirectly modulated by the candidate drug?	Known targets, target classes, mechanism-of-action annotations, curated pharmacological labels	Connects candidate drugs to molecular mechanisms and known pharmacology	Supports target-centric rationales for why a drug may influence disease biology	Target annotations may be incomplete, context-dependent, or biased toward approved and well-studied drugs
Protein–protein interaction network proximity	Are drug targets located near disease-relevant proteins in the interactome?	Drug target–disease protein distance, shared network neighborhoods, disease-module proximity	Adds systems-level plausibility beyond isolated target matching	Highlights network neighborhoods or target–disease paths supporting the prediction	Interactome incompleteness and false-positive edges may distort proximity-based explanations
Knowledge-graph context	What structured biomedical relationships connect drug, target, disease, pathway, and phenotype?	Drug–target–pathway–disease relations, graph paths, semantic links, known drug–disease associations	Supports relational inference over heterogeneous biomedical evidence	Produces interpretable reasoning paths that can be inspected by experts	Reasoning paths are limited by the quality, coverage, and bias of the underlying graph
Prior repurposing knowledge	Has related drug–disease evidence been observed before?	Known indications, historical repurposing pairs, curated drug resources, disease–drug associations	Provides supervision and contextual grounding for candidate ranking	Helps distinguish novel hypotheses from candidates supported by prior biomedical evidence	Leakage and circular evidence can inflate apparent model performance if not controlled

Pathway Enrichment as Structured Features

For each disease and drug signature, enrichment analysis could be performed against pathway and gene-set resources to create structured activity profiles. These pathway-level summaries would reduce the dimensionality of transcriptomic data while retaining interpretable biological meaning, as shown conceptually by pathway-guided and enrichment-based predictive models [16, 17]. The resulting features could represent whether a pathway is activated in disease, suppressed by a drug, or directionally reversed between the two. This design allows the explanation layer to refer to recognizable biological processes rather than only isolated genes.

Network Proximity and Target Features

Target-network features would be derived by mapping drug targets and disease-associated proteins onto a human protein–protein interaction network. The model could represent how close a drug's targets are to disease proteins, whether targets and disease proteins share pathways, and whether relevant graph neighborhoods contain known mechanistic links [11, 12]. COVID-19 repurposing studies using network medicine and integrative frameworks illustrate how target proximity and systems-level context can support candidate prioritization [18, 19]. In the proposed model, these features would also support target-centric explanations that identify which proteins or network neighborhoods are responsible for a prediction.

Explainable Model Architecture

Predictive Model Core

The predictive core could be implemented as a gradient-boosted tree ensemble over engineered multi-modal features or as a graph neural network over a drug–target–gene–pathway–disease graph. DeepDR demonstrates the feasibility of network-based deep learning for drug repositioning [6], while graph neural approaches to repurposing show how multiple sources of biological evidence can be harmonized in a single predictive structure [23]. Knowledge-graph frameworks provide an alternative architecture when the goal is to reason over typed biomedical relationships rather than fixed feature vectors [10, 24]. In all cases, the model should be treated as a hypothesis generator rather than a source of definitive therapeutic claims.

Global and Local Explainability with SHAP

Post-hoc SHAP values could be calculated for each feature to estimate how much transcriptomic reversal, pathway enrichment, target proximity, or graph-derived evidence contributes to a specific prediction. These local explanations would then be grouped into biological categories, such as cytokine signaling, kinase inhibition, metabolic regulation, or immune-cell activation, so that the output is concise enough for expert review. Interpretable drug-response and pathway-based models show that attribution becomes more useful when model features are aligned with biological structure [17, 25]. Global summaries could also reveal which categories of evidence the model generally relies on across diseases or drug classes.

Additional Interpretability Layers

If the model uses a graph neural network, attention or edge-level relevance mechanisms could highlight specific drug–target, target–pathway, or pathway–disease connections that support a candidate. Interpretable graph and transformer models in drug-response prediction illustrate how attention-like mechanisms can connect predictions to molecular entities and biological knowledge [20]. Visible neural architectures similarly suggest that model structure can be aligned with known biology so that intermediate representations correspond to interpretable functional modules [21]. These layers should complement SHAP-style attribution by showing not only which features matter, but also how evidence flows through the biological graph.

Linking Predictions to Biological Pathways and Targets

From SHAP Values to Pathway Narratives

The explanation module would aggregate gene-level and pathway-level feature attributions into a pathway narrative for each predicted drug–disease pair. For example, instead of reporting only that a compound could be relevant to an inflammatory disorder, the model would describe whether the prediction is supported by reversal of cytokine signaling, stress-response pathways, or immune-cell activation programs. Pathway-guided prediction studies show that biological structure can make model outputs easier to interpret when pathways are used as explicit explanatory units [16, 17]. The narrative should therefore translate attribution patterns into cautious mechanistic language, such as stating that a candidate is supported by coordinated reversal of a disease-associated pathway and by target interactions consistent with that pathway.

Target-Centric Explanation

A target-centric explanation would identify whether the drug's known targets are close to disease-associated proteins, embedded in relevant pathways, or connected through interpretable graph paths. Network medicine frameworks support this logic by treating drug targets and disease modules as interacting regions of the human interactome [11, 12]. Knowledge-graph repurposing models add a complementary view by tracing interpretable relationships between drugs, targets, diseases, and biological concepts [8, 26]. For a drug developer, this explanation would help distinguish a candidate supported by coherent target biology from one supported mainly by indirect statistical similarity.

Visualising Transcriptomic Reversal

Transcriptomic reversal could be visualised by showing genes whose disease-associated expression changes are directionally opposed by a drug perturbation signature. Connectivity Map and L1000-based resources provide the conceptual basis for this type of drug–disease comparison [3, 5]. Deep embedding approaches further suggest that expression profiles can be organized into a learned molecular space while still requiring interpretable views that expose the genes and pathways driving similarity or reversal [15]. A useful visualization would therefore connect gene-level reversal to pathway-level summaries rather than presenting a dense gene list without biological structure. **Figure 2** illustrates how disease-associated gene-expression changes can be compared with drug perturbation signatures to identify directional reversal, summarize affected pathways, and judge whether the model explanation is biologically coherent enough to support candidate prioritization.

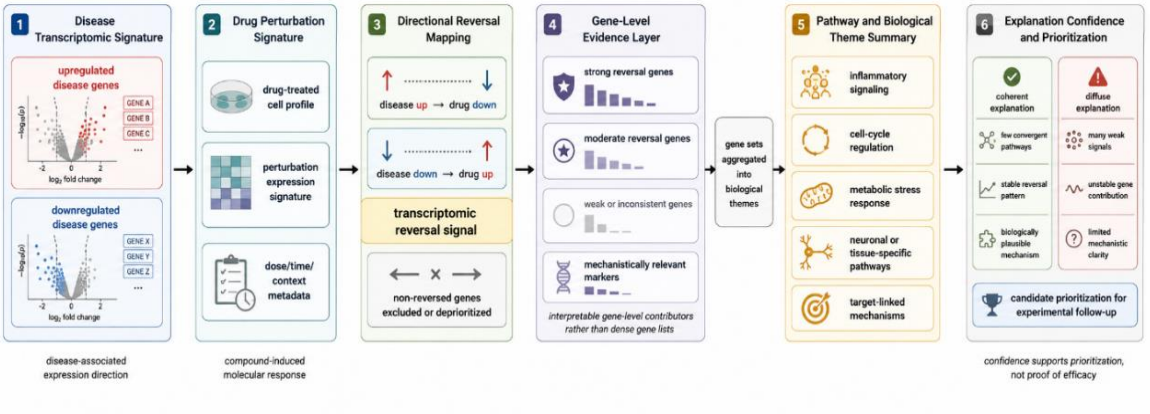

Figure 2. Gene-to-pathway visualization of transcriptomic reversal and explanation confidence

Confidence and Limitations of the Explanation

The model should distinguish predictions supported by a small number of coherent biological themes from predictions supported by many weak and diffuse signals. This distinction matters because an apparently high-ranking candidate may be less actionable if its explanation cannot be connected to plausible disease biology or target pharmacology. Interpretable architecture studies caution that transparency alone does not guarantee biological correctness, so explanations should be assessed for coherence, stability, and consistency with known mechanisms [22]. Explanation confidence should therefore be framed as a guide for prioritization rather than as proof that the candidate will succeed experimentally.

Explainability Methods for Drug Developers

Interactive Explanation Dashboard

An interactive dashboard could allow scientists to browse predicted drug–disease pairs, inspect the top pathway themes, and drill down into the genes, targets, and network paths that support each candidate. Query-oriented perturbation resources demonstrate how drug-induced expression signatures can be made searchable for biological interpretation [5], while knowledge-graph frameworks show how drug–disease evidence can be represented through connected biomedical entities [2, 10]. The dashboard should present transcriptomic, pathway, and target evidence in parallel so that users can evaluate whether the model's rationale matches their domain knowledge. This interface would make explainability part of routine hypothesis review rather than a separate technical appendix.

Exportable Evidence Reports

For each prediction, the system could generate a concise evidence report summarizing the disease signature, the drug perturbation signature, the most influential pathways, and the target-network rationale. The Drug Repurposing Hub illustrates the value of curated drug annotations for connecting compounds to mechanisms, targets, and therapeutic context [1]. Literature-based and knowledge-graph approaches further support evidence reports by linking candidates to biomedical concepts that can be reviewed by experts [14, 24]. Such reports should be written as decision-support documents, not as claims of efficacy, and should make clear which evidence comes from transcriptomics, networks, pathways, or prior knowledge.

Validation of Biological Coherence

Explanation quality should be evaluated by asking whether the model recovers mechanisms that experts would consider plausible for known or historically supported repurposing examples. Mechanism-driven screening and drug-repurposing frameworks show that computational prioritization becomes more useful when model outputs can be interpreted through disease biology and drug action [27]. Integrative COVID-19 repurposing studies also illustrate how target mechanisms, pathway context, and transcriptional evidence can be combined into a biologically grounded hypothesis [18, 19]. The aim of this evaluation would not be to claim prospective success, but to assess whether the explanation layer produces mechanistic stories that domain experts can scrutinize.

Feedback Loop from Experimental Follow-Up

When researchers experimentally validate or invalidate a repurposing hypothesis, the outcome could be added to the model's knowledge base and used to refine both prediction and explanation. Established disease–drug pair knowledge has been used to support computational repurposing, showing how prior evidence can structure learning over candidate associations [28]. After feedback is incorporated, SHAP-style attribution or graph relevance analysis could be repeated to identify which transcriptomic, pathway, or target features distinguished supported hypotheses from unsupported ones. This feedback loop would help the explanation engine evolve with biological evidence instead of remaining fixed at the time of model development.

Integration into Repurposing Pipelines

Early-Stage Hypothesis Generation

In an early-stage pipeline, the model could scan approved or clinically characterized drugs against a new disease signature and return a shortlist of candidates with attached biological rationales. Repurposing libraries and curated drug resources provide a practical foundation for such screening because they organize compounds with known mechanisms, annotations, and pharmacological context [1]. Network and graph-based methods can then connect these compounds to disease modules, transcriptomic states, and pathway perturbations [11, 23]. The resulting shortlist would be most useful when reviewed jointly by computational scientists, pharmacologists, clinicians, and disease biologists.

Regulatory and Intellectual Property Considerations

An explainable model could support translational planning by clarifying the biological mechanism proposed for a new drug indication. Knowledge-graph and systems-level repurposing frameworks are useful in this setting because they can connect a candidate to target biology, disease pathways, and prior biomedical evidence [2, 9]. A mechanistic explanation may also help teams formulate a coherent development rationale, although it should not be treated as a substitute for pharmacological, toxicological, or clinical evidence. The model's role would be to organize and justify a testable hypothesis, not to determine regulatory acceptability.

Evaluation Strategy

Predictive Performance

Predictive evaluation should use standard drug–disease association assessment strategies while avoiding leakage between training and evaluation examples. Knowledge-graph completion, supervised repurposing, and graph neural network studies all demonstrate the importance of evaluating predictions against known associations while preserving separation between related drugs, diseases, or mechanisms [10, 13, 23]. Metrics such as ranking quality and precision-oriented retrieval can be discussed conceptually, but the model should not be presented with invented performance values. The evaluation should emphasize whether the framework is suitable for prioritization and explanation rather than claiming validated therapeutic discovery.

Explanation Quality and Actionability

Explanation quality should be judged by whether the model's pathway, target, and transcriptomic rationales are biologically plausible, internally consistent, and useful for expert decision-making. Interpretable drug sensitivity and drug-response models show that biological features can be exposed to users, but they also highlight the need to evaluate whether explanations are meaningful rather than merely available [20, 25, 29]. Expert review could examine whether the explanation identifies credible pathways, relevant targets, and coherent transcriptomic reversal for each candidate. Actionability should be understood as the degree to which the explanation helps scientists decide what to test next.

Table 2 proposes an evaluation framework that separates predictive performance from explanation faithfulness, biological coherence, and practical actionability for translational decision-making.

Table 2. Evaluation Framework for Prediction Quality, Explanation Quality, and Translational Actionability

Evaluation dimension	Core evaluation question	Suggested assessment approach	Evidence of strong performance	Evidence of weak performance	Relevance to drug developers and translational teams
Candidate ranking validity	Does the model prioritize biologically plausible drug–disease pairs?	Retrospective ranking against known drug–disease associations using leakage-aware splits	Known or historically supported candidates appear near the top without relying on duplicated evidence	Rankings are driven by data leakage, popularity bias, or overrepresented disease areas	Determines whether the model is useful for early hypothesis generation
Time-aware generalization	Could the model have identified later-supported hypotheses using only earlier knowledge?	Train on data available before a defined cutoff and evaluate against later evidence	Later-supported hypotheses are prioritized with plausible pre-existing evidence	Performance collapses when future knowledge is removed	Tests whether the system behaves like a realistic discovery aid rather than a retrospective pattern matcher
Transcriptomic explanation coherence	Do gene-level and signature-level explanations support a biologically meaningful disease–drug relationship?	Expert review of reversed genes, expression contrasts, and disease-relevant transcriptional programs	Key reversed signatures align with known or plausible disease biology	Explanations emphasize irrelevant genes, generic stress signatures, or contradictory expression effects	Helps users judge whether transcriptomic evidence deserves experimental follow-up

Pathway explanation coherence	Are high-attribution pathways mechanistically relevant to the proposed indication?	Review top pathway attributions for biological plausibility, directionality, and consistency	Explanations identify coherent disease-relevant pathways and specify whether they are reversed or reinforced	Pathway rationales are diffuse, redundant, overly broad, or disconnected from disease mechanisms	Supports mechanism-based prioritization rather than score-based candidate selection
Target-network interpretability	Do target and network explanations clarify how the drug could influence disease biology?	Inspect target proximity, graph neighborhoods, and target–pathway–disease paths	Drug targets are connected to disease modules through interpretable mechanisms or pathways	Target evidence is indirect, unstable, or dependent on poorly supported network edges	Helps distinguish candidates with actionable target biology from statistical associations
Explanation faithfulness	Do explanations reflect the model's actual decision logic?	Compare SHAP, attention, ablation, and feature-removal analyses	Removing highly attributed features meaningfully changes predictions	Explanations appear plausible but do not affect model outputs when removed	Prevents attractive but misleading biological narratives
Explanation stability	Are explanations consistent under reasonable data or model perturbations?	Repeat attribution after resampling, alternative normalization, or model retraining	Major pathway and target rationales remain stable across perturbations	Explanations change substantially despite similar candidate scores	Increases confidence that the rationale is not an artifact of one model run
Expert actionability	Does the explanation help experts decide what to test next?	Structured review by pharmacologists, disease biologists, medicinal chemists, and clinicians	Reports suggest concrete assays, mechanisms, biomarkers, or validation experiments	Reports are descriptive but do not guide experimental design or prioritization	Converts explainable AI output into practical translational decision support
Limitation transparency	Does the system clearly state uncertainty and evidence gaps?	Review evidence reports for missing data, context mismatch, annotation uncertainty, and unsupported assumptions	Reports identify weak evidence domains and avoid efficacy claims	Reports overstate computational predictions or hide data limitations	Maintains appropriate caution in preclinical and translational planning

Prospective Validation

A prospective evaluation strategy would train the model using information available before a defined time point and then examine whether later-supported repurposing hypotheses would have been prioritized with plausible explanations. Time-aware evaluation is important because biomedical knowledge graphs, drug annotations, and transcriptomic resources evolve, which can otherwise make historical prediction tasks unrealistically easy [2, 10]. The explanation layer should be assessed alongside candidate ranking by asking whether the model represented the relevant mechanism in a form that experts could have interpreted before later evidence emerged. This design would test the model as a realistic hypothesis-generation system rather than as a retrospective pattern-matching exercise.

Limitations

Incompleteness of Biological Knowledge

Pathway databases, drug–target annotations, and protein interaction networks are incomplete and biased toward well-studied genes, diseases, and therapeutic areas. Network-based repurposing depends on the quality of the interactome and disease-module definitions, so missing or inaccurate edges can distort both predictions and explanations [11, 12]. Knowledge-graph methods face similar limitations because their reasoning paths are constrained by the entities and relationships already represented in the graph [8, 26]. As a result, an explainable model may produce clear but incomplete rationales, especially for understudied diseases or mechanisms outside current pathway knowledge.

Transcriptomic Noise and Confounding

Drug perturbation signatures are strongly influenced by cell line, dose, exposure duration, assay platform, and preprocessing choices. Large-scale resources such as Connectivity Map and LINCS L1000 make transcriptomic repurposing feasible, but they also require careful harmonization and interpretation across experimental contexts [3, 4]. A drug that reverses a disease signature in one cellular background may not have the same effect in the target tissue, and an explanation may reflect technical artifacts if confounding is not controlled. For this reason, transcriptomic evidence should be presented as one component of a broader mechanistic hypothesis rather than as a standalone justification.

Conclusion

An explainable drug repurposing model can integrate disease transcriptomics, drug perturbation signatures, pathway enrichment, and target networks into a single hypothesis-generation framework. Its distinctive value is not only the ranking of candidate drug–disease pairs, but the ability to explain why each pair appears biologically plausible.

The strongest feature of this approach is biological transparency. By producing pathway-level attribution narratives, target-centric explanations, and transcriptomic reversal summaries, the model can help researchers interpret predictions in the language of disease mechanisms and pharmacology.

Important challenges remain before such systems can be used routinely. Data quality, pathway incompleteness, target annotation bias, and context-dependent drug responses all limit the reliability of both predictions and explanations.

Future progress will depend on open-source implementations, transparent benchmarks, and community-wide evaluation of explanation quality. Explainable AI for drug repurposing should ultimately be judged by whether it helps scientists generate clearer, more testable, and more biologically grounded therapeutic hypotheses.

Acknowledgments: None

Conflict of interest: None

Financial support: None

Ethics statement: None

References

1. Corsello SM, Bittker JA, Liu Z, Gould J, McCarren P, Hirschman JE, et al. The Drug Repurposing Hub: a next-generation drug library and information resource. Nat Med. 2017;23(4):405-8.
2. 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.
3. 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):1437-52.
4. Keenan AB, Jenkins SL, Jagodnik KM, Koplev S, He E, Torre D, et al. The library of integrated network-based cellular signatures NIH program: system-level cataloging of human cells response to perturbations. Cell Syst. 2018;6(1):13-24.
5. Wang Z, Lachmann A, Keenan AB, Ma'Ayan A. L1000FWD: fireworks visualization of drug-induced transcriptomic signatures. Bioinformatics. 2018;34(12):2150-2.
6. Zeng X, Zhu S, Liu X, Zhou Y, Nussinov R, Cheng F. deepDR: a network-based deep learning approach to in silico drug repositioning. Bioinformatics. 2019;35(24):5191-8.
7. Yang J, Li Z, Wu WK, Yu S, Xu Z, Chu Q, et al. Deep learning identifies explainable reasoning paths of mechanism of action for drug repurposing from multilayer biological network. Brief Bioinform. 2022;23(6):bbac469.
8. Jiménez A, Merino MJ, Parras J, Zazo S. Explainable drug repurposing via path based knowledge graph completion. Sci Rep. 2024;14(1):16587.
9. Islam MK, Amaya-Ramirez D, Maigret B, Devignes MD, Aridhi S, Smaïl-Tabbone M. Molecular-evaluated and explainable drug repurposing for COVID-19 using ensemble knowledge graph embedding. Sci Rep. 2023;13(1):3643.
10. Gao Z, Ding P, Xu R. KG-Predict: A knowledge graph computational framework for drug repurposing. J Biomed Inform. 2022;132:104133.
11. Morselli Gysi D, Do Valle Í, Zitnik M, Ameli A, Gan X, Varol O, et al. Network medicine framework for identifying drug-repurposing opportunities for COVID-19. Proc Natl Acad Sci U S A. 2021;118(19):e2025581118.
12. Zhou Y, Hou Y, Shen J, Huang Y, Martin W, Cheng F. Network-based drug repurposing for novel coronavirus 2019-nCoV/SARS-CoV-2. Cell Discov. 2020;6(1):14.
13. Zhang R, Hristovski D, Schutte D, Kastrin A, Fiszman M, Kilicoglu H. Drug repurposing for COVID-19 via knowledge graph completion. J Biomed Inform. 2021;115:103696.
14. Sosa DN, Derry A, Guo M, Wei E, Brinton C, Altman RB. A literature-based knowledge graph embedding method for identifying drug repurposing opportunities in rare diseases. In: Pac Symp Biocomput. 2020;25:463-74.
15. Donner Y, Kazmierczak S, Fortney K. Drug repurposing using deep embeddings of gene expression profiles. Mol Pharm. 2018;15(10):4314-25.
16. Deng L, Cai Y, Zhang W, Yang W, Gao B, Liu H. Pathway-guided deep neural network toward interpretable and predictive modeling of drug sensitivity. J Chem Inf Model. 2020;60(10):4497-505.
17. Tang YC, Gottlieb A. Explainable drug sensitivity prediction through cancer pathway enrichment. Sci Rep. 2021;11(1):3128.
18. Ge Y, Tian T, Huang S, Wan F, Li J, Li S, et al. An integrative drug repositioning framework discovered a potential therapeutic agent targeting COVID-19. Signal Transduct Target Ther. 2021;6(1):165.

19. Loucera C, Esteban-Medina M, Rian K, Falco MM, Dopazo J, Peña-Chilet M. Drug repurposing for COVID-19 using machine learning and mechanistic models of signal transduction circuits related to SARS-CoV-2 infection. Signal Transduct Target Ther. 2020;5(1):290.
20. Shin J, Piao Y, Bang D, Kim S, Jo K. DRPreter: interpretable anticancer drug response prediction using knowledge-guided graph neural networks and transformer. Int J Mol Sci. 2022;23(22):13919.
21. Ferraro L, Scala G, Cerulo L, Carosati E, Ceccarelli M. MOViDA: multiomics visible drug activity prediction with a biologically informed neural network model. Bioinformatics. 2023;39(7):btad432.
22. Li Y, Hostallero DE, Emad A. Interpretable deep learning architectures for improving drug response prediction performance: myth or reality? Bioinformatics. 2023;39(6):btad390.
23. Hsieh K, Wang Y, Chen L, Zhao Z, Savitz S, Jiang X, et al. Drug repurposing for COVID-19 using graph neural network and harmonizing multiple evidence. Sci Rep. 2021;11(1):23179.
24. Zhu Y, Che C, Jin B, Zhang N, Su C, Wang F. Knowledge-driven drug repurposing using a comprehensive drug knowledge graph. Health Informatics J. 2020;26(4):2737-50.
25. Kuenzi BM, Park J, Fong SH, Sanchez KS, Lee J, Kreisberg JF, et al. Predicting drug response and synergy using a deep learning model of human cancer cells. Cancer Cell. 2020;38(5):672-84.
26. Ghorbanali Z, Zare-Mirakabad F, Akbari M, Salehi N, Masoudi-Nejad A. Drugrep-kg: Toward learning a unified latent space for drug repurposing using knowledge graphs. J Chem Inf Model. 2023;63(8):2532-45.
27. Pham TH, Qiu Y, Zeng J, Xie L, Zhang P. A deep learning framework for high-throughput mechanism-driven phenotype compound screening and its application to COVID-19 drug repurposing. Nat Mach Intell. 2021;3(3):247-57.
28. Saberian N, Peyvandipour A, Donato M, Ansari S, Draghici S. A new computational drug repurposing method using established disease–drug pair knowledge. Bioinformatics. 2019;35(19):3672-8.
29. Pang W, Chen M, Qin Y. Prediction of anticancer drug sensitivity using an interpretable model guided by deep learning. BMC Bioinformatics. 2024;25(1):182.