Pharmacophore

ISSN-2229-5402

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

KNOWLEDGE-GRAPH FRAMEWORK FOR RARE DISEASE DRUG REPURPOSING USING GENES, PATHWAYS, AND ORPHAN DRUG LABELS

Jose Martinez[1]*, Carmen Lopez[1]

1. *Department of Intelligent Pharmaceutical Analytics, Faculty of Pharmacy, University of Salamanca, Salamanca, Spain.*

ARTICLE INFO

Received:
01 February 2026
Received in revised form:
20 April 2026
Accepted:
23 April 2026
Available online:
28 April 2026

Keywords: Rare diseases, Knowledge graph, Drug repurposing, Orphan drugs, Gene–disease associations, Biological pathways

ABSTRACT

Rare disease drug development faces significant challenges due to small patient populations, fragmented evidence, and limited commercial incentives, making drug repurposing an attractive strategy when genetic, pathway, pharmacological, and regulatory evidence are considered together. Existing computational repurposing approaches often focus on biological similarity or drug–target relationships while treating regulatory knowledge as external context, which restricts the ability to prioritize candidates with orphan drug precedent, approved indications, or interpretable translational relevance. To address this, we propose a knowledge-graph framework that integrates gene–disease associations, biological pathway relationships, drug–target evidence, and orphan drug regulatory labels, enabling the generation and ranking of repurposing hypotheses for rare diseases without claiming experimental validation or quantitative performance. The system incorporates entity-resolution workflows, a graph database, a reasoning layer, and a hypothesis-ranking module, linking disease, gene, pathway, drug, orphan designation, and label-indication entities while preserving provenance for each relationship. This framework allows researchers to query for drugs that target pathways disrupted by rare disease genes and that have orphan designation or approval evidence in related conditions, making mechanistic and regulatory evidence visible within the same reasoning environment. By embedding orphan drug regulatory knowledge into a biological knowledge graph, the approach helps bridge the translational gap between mechanistic discovery and clinical development, supporting transparent, evidence-aware prioritization of repurposing opportunities for rare diseases.

To Cite This Article: Martinez J, Lopez C. Knowledge-Graph Framework for Rare Disease Drug Repurposing Using Genes, Pathways, and Orphan Drug Labels. Pharmacophore. 2026;17(2):123-31. https://doi.org/10.51847/4hPgLZX1cA

Introduction

Rare diseases collectively impose a substantial clinical burden while remaining difficult targets for conventional drug development because evidence is dispersed, patient cohorts are small, and commercial incentives are uneven. Analyses of orphan drug designations and approvals show that regulatory incentives have stimulated rare disease drug development, yet many conditions still lack approved treatment options [1]. Longitudinal reviews of FDA orphan drug activity further indicate that drug development attention is distributed unevenly across disease areas, reinforcing the need for systematic approaches that can surface underexplored opportunities [2]. Drug repurposing is therefore attractive because it can begin from existing pharmacological and regulatory knowledge rather than from a wholly new discovery program.

Current repurposing efforts depend on disconnected sources that describe different fragments of the translational picture. Disease genetics may be represented in DisGeNET [3], pathway biology in Reactome [4], KEGG, and WikiPathways, while drug structures, targets, and bioactivity evidence are distributed across DrugBank [5] and ChEMBL [6]. These resources are individually valuable, but their separation makes it difficult to inspect whether a drug target, disrupted pathway, rare disease gene, and orphan regulatory precedent all converge on the same hypothesis. A rare disease researcher may therefore need to manually assemble evidence that should be available as an integrated computational graph.

Biomedical knowledge graphs provide a practical way to integrate heterogeneous biomedical data and reason across multiple relationship types. Hetionet demonstrated that systematic integration of biomedical knowledge can prioritize drugs for repurposing by connecting diseases, genes, drugs, pathways, and phenotypes within a shared network [7]. More recent precision-medicine knowledge graphs have shown how diverse biomedical entities can be assembled into reusable graph

Corresponding Author: Jose Martinez; Department of Intelligent Pharmaceutical Analytics, Faculty of Pharmacy, University of Salamanca, Salamanca, Spain. E-mail: jose.martinez@gmail.com

resources for computational discovery [8], and OREGANO illustrates how drug-repurposing evidence can be organized as a dedicated knowledge graph [9]. Biomedical knowledge graph learning also supports multi-layer inference over drug, disease, and mechanistic relationships, suggesting that rare disease repurposing can benefit from graph-native representations [10].
The thesis of this AIF article is that rare disease drug repurposing should treat regulatory precedent as a first-class graph component rather than as post hoc narrative context. Literature-based knowledge graph embedding has already been applied to rare disease repurposing [11], while rare disease knowledge graphs derived from curated information sources show how disease-specific entities can be integrated for discovery [12]. Rare disease research and clinical-trial knowledge graphs further demonstrate that research funding, trial design, and rare disease concepts can be represented in structured graph form [13, 14]. The proposed framework builds on this direction by fusing genes, pathways, drug targets, and orphan drug labels so that regulatory evidence contributes directly to hypothesis generation.

Background

Rare Diseases and Drug Repurposing

Rare disease repurposing seeks to identify new therapeutic uses for existing or previously characterized drugs, especially when conventional development is constrained by limited patient populations. Orphan drug policy has created incentives for development, but analyses of designations and approvals suggest that therapeutic progress remains uneven across disease categories [1, 2]. Well-known rare disease repositioning examples, such as sildenafil for pulmonary arterial hypertension and everolimus for tuberous sclerosis complex, illustrate the broader principle that established drugs can become relevant in rare indications when mechanism and clinical context align. A graph-based framework would not assume that such examples generalize automatically, but it could use known repurposing patterns as conceptual anchors for retrospective evaluation and hypothesis design [15].

Biomedical Knowledge Graphs for Drug Discovery

Biomedical knowledge graphs organize entities such as genes, diseases, drugs, phenotypes, targets, and pathways as nodes connected by typed relationships. Hetionet showed that graph integration can support drug-repurposing prioritization [7], while precision-medicine graph resources demonstrate how clinically relevant biomedical concepts can be connected for broader translational reasoning [8]. Embedding and graph-learning methods can transform these relationships into machine-readable representations, enabling candidate drug–disease links to be inferred from graph topology and semantic context [10, 16]. In an AIF framework for rare diseases, such methods should be used conceptually as reasoning modules whose outputs require biological review rather than as definitive predictions [17].

Gene–Disease and Pathway Databases for Rare Diseases

Gene–disease and pathway databases provide the mechanistic substrate for rare disease repurposing, but they differ in scope, identifiers, curation policies, and update practices. DisGeNET captures disease genomics relationships that can support gene–disease edges [3], while Reactome [4], KEGG, and WikiPathways provide complementary pathway representations that can connect rare disease genes to biological processes. Rare disease knowledge graphs derived from curated resources show that disease-specific entity integration is feasible, but they also highlight challenges of semantic alignment and incomplete coverage [12]. Ontology-enriched clustering of rare diseases further suggests that graph context can help interpret sparse disease information by relating poorly characterized disorders to better annotated neighbors [18].

Orphan Drug Regulatory Labels as a Knowledge Source

Orphan drug designations and regulatory labels contain structured and semi-structured evidence about the drug, intended disease, approval status, indication language, and sometimes mechanistic rationale. Studies of FDA orphan drug designations show that these records can be used to describe rare disease development trends and therapeutic focus areas [1, 2]. Clinical-trial knowledge graph work in rare diseases suggests that regulatory and trial-related concepts can be encoded as graph entities rather than being left as disconnected documents [14]. In the proposed framework, orphan drug labels would become a regulatory layer that links drugs to rare disease entities, related indications, development status, and evidence provenance.

Graph-Based Inference and Explainability

Graph-based inference can identify candidate drug–disease relationships through path-based reasoning, knowledge graph completion, graph neural networks, or hybrid symbolic and statistical approaches. Explainable drug repurposing via path-based knowledge graph completion shows how predicted associations can be accompanied by interpretable graph paths [19]. Case-based explainable graph neural network approaches further indicate that mechanistic drug repositioning can benefit from explanations that resemble prior examples rather than opaque scores [20]. Rare disease repurposing frameworks should therefore prioritize evidence paths, provenance, and human review, especially when using multimodal or graph-neural methods to generate hypotheses [21-24].

Framework Architecture Overview

High-Level Design

The proposed architecture begins with ingestion of biomedical and regulatory sources into a graph database, followed by entity normalization into a shared identifier space. A graph reasoning module then evaluates candidate drug–disease links by traversing gene, pathway, target, and orphan-designation relationships, similar in spirit to biomedical knowledge graph resources built for precision medicine [8]. OREGANO provides a useful precedent for representing drug-repurposing knowledge in graph form [9], while clinician-centered repurposing frameworks indicate how graph-derived hypotheses could be made more actionable for translational users [25]. The user-facing layer would visualize evidence paths, provenance, and the distinction between mechanistic support and regulatory precedent.

Figure 1 illustrates the proposed regulatory-aware knowledge-graph architecture for generating explainable rare disease drug repurposing hypotheses from genes, pathways, drug-target evidence, and orphan drug labels.

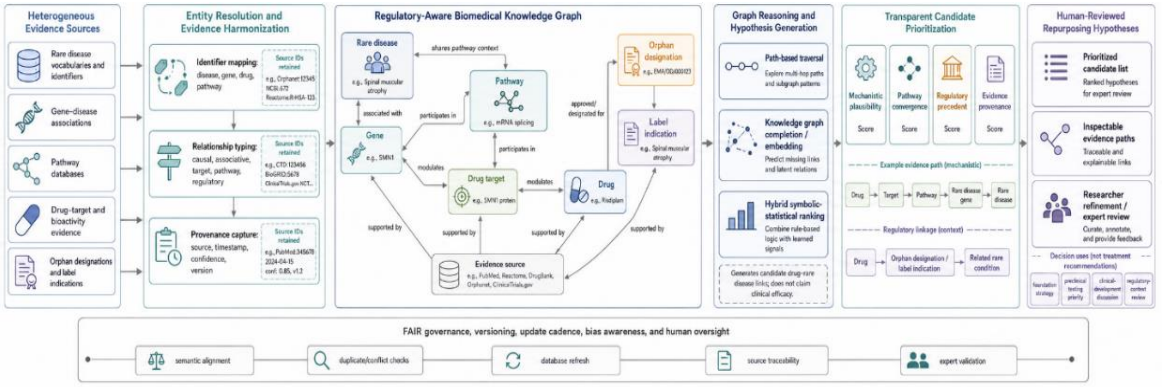

Figure 1. Knowledge-graph framework for rare disease drug repurposing using genes, pathways, drug targets, and orphan regulatory labels

Core Graph Components

The core graph would include disease, gene, pathway, drug, target, orphan designation, and label-indication nodes. Gene–disease edges could be informed by disease genomics resources [3], pathway membership edges by Reactome [4], and drug-target or bioactivity edges by DrugBank [5] and ChEMBL [6]. Orphan designation and label-indication nodes would connect drugs to rare disease entities through regulatory evidence rather than through purely biological similarity. This design allows a candidate hypothesis to be represented as a multi-step route from a drug to its target, from target to pathway, from pathway to disease mechanism, and from drug to orphan regulatory history.

Table 1 defines the graph schema required to convert heterogeneous rare disease, pathway, pharmacological, and orphan regulatory evidence into traceable repurposing hypotheses.

Table 1. Graph entity and relationship schema for a regulatory-aware rare disease repurposing knowledge graph

Graph layer	Core node types	Representative relationship types	Evidence role in repurposing logic	Provenance and quality-control requirements	Interpretive caution
Rare disease identity layer	Rare disease entities; disease ontology identifiers; synonym records; related-condition groupings	same_as; synonym_of; subtype_of; phenotypically_related_to; shares_ontology_context_with	Establishes the disease object that will anchor graph queries and candidate ranking; helps connect ultra-rare disorders to better-annotated neighboring conditions	Preserve source vocabulary, original disease name, normalized identifier, mapping rule, and mapping confidence; flag ambiguous disease-name matches	Semantic similarity should support exploration, not imply that therapeutic evidence transfers directly across conditions
Genetic evidence layer	Genes; variants where available; gene–disease association records	associated_with; causal_for; contributory_to; reported_in; has_variant_evidence	Connects rare disease biology to potentially druggable mechanisms and pathway-level interpretation	Separate causal, contributory, and associative evidence; store source database, evidence type, publication or database accession, and update date	All gene–disease edges should not be treated as equivalent; missing gene links do not mean absence of biological relevance
Pathway and biological process layer	Pathways; biological processes; molecular functions;	participates_in; regulates; upstream_of; downstream_of; shares_pathway_with	Provides an intermediate mechanistic bridge between disease	Track pathway database source, pathway version, membership definition, and cross-	Pathway membership is supportive evidence, not proof that modulating one

	pathway membership records		genes and drug targets	database pathway alignment	target will correct disease biology
Drug and target pharmacology layer	Drugs; compounds; protein targets; mechanisms of action; bioactivity records	targets; inhibits; activates; binds; has_bioactivity_against; has_mechanism	Identifies pharmacological routes through which existing drugs may intersect disease-relevant genes or pathways	Preserve drug identifiers, target identifiers, assay type, activity measure, target confidence, and source database	Bioactivity or target annotation alone is insufficient for therapeutic plausibility without context on exposure, safety, and disease mechanism
Orphan regulatory evidence layer	Orphan drug designations; approved labels; indication-language entities; regulatory-status records	has_orphan_designation_for; approved_for; label_mentions; development_status; related_indication	Adds regulatory precedent directly into candidate prioritization instead of treating it as narrative background	Store agency source, designation or approval status, indication text, date, jurisdiction, and drug–disease mapping rule	Orphan precedent can improve translational relevance but does not establish efficacy for a different rare disease
Evidence source and provenance layer	Databases; publications; regulatory documents; curation events; timestamped graph versions	derived_from; curated_from; updated_on; conflicts_with; supports_edge	Makes each relationship inspectable, auditable, and updateable for expert review	Require source traceability for every edge; record versioning, refresh cadence, curator notes, and conflict flags	Graph transparency depends on provenance completeness; untraceable edges should be down-weighted or excluded from high-priority outputs
Hypothesis entity layer	Candidate drug–disease hypotheses; ranked candidate records; explanation-path objects	supported_by_path; ranked_by; requires_review; excluded_due_to_existing_indication	Converts graph connectivity into reviewable repurposing hypotheses with interpretable support paths	Store ranking features, explanation paths, excluded known indications, review status, and analyst annotations	Hypothesis nodes represent research priorities, not treatment recommendations or validated clinical claims

Design Principles

The framework should follow FAIR-oriented integration principles by making graph entities identifiable, evidence relationships traceable, and provenance available for review. Reviews of knowledge graphs for drug repurposing emphasize that database interoperability, semantic harmonization, and method selection shape the usefulness of graph-based discovery systems [15]. Explainable path-based completion methods show why a repurposing recommendation should include the graph route that supports it rather than only a ranked output [19]. For rare diseases, ontology-enriched graph organization can also help users reason across related conditions while acknowledging that similarity does not automatically imply therapeutic transferability [18].

Knowledge Graph Construction and Data Ingestion

Entity Resolution and Data Normalization

Entity resolution would map rare diseases to UMLS, Orphanet, or other disease identifiers, genes to stable gene identifiers, and drugs to ChEMBL and DrugBank records. DisGeNET provides disease genomics relationships that can seed gene–disease edges [3], while DrugBank [5] and ChEMBL [6] can support harmonization of drug names, targets, and bioactivity concepts. Rare disease knowledge graph work shows that curated disease resources can be integrated, but that naming variation and ontology mapping require explicit normalization rules [12]. The framework would preserve original source identifiers alongside normalized graph identifiers so that users can inspect where each assertion originated.

Integration of Gene, Pathway, and Regulatory Sources

After normalization, ingestion workflows would populate gene, pathway, drug, disease, and regulatory nodes from curated databases and agency resources. Reactome [4], KEGG, and WikiPathways would provide complementary pathway structures, while DrugBank can contribute drug-target and mechanism-oriented relationships [5]. Rare disease resources such as GARD-derived knowledge graphs demonstrate how disease information can be transformed into structured graph form for discovery [12]. Automated quality checks would examine identifier consistency, duplicate entities, missing relationship types, and conflicts between source assertions without claiming that such checks remove all biological uncertainty.

Graph Database and Update Cadence

The knowledge graph could be stored in Neo4j or a comparable graph platform that supports typed nodes, typed edges, provenance properties, and path queries. OREGANO illustrates how computational drug-repurposing knowledge can be distributed as a structured graph resource [9], while rare disease research discovery graphs show how new evidence streams can be incorporated as additional relationships [13]. Clinical-trial knowledge graph work also suggests that rare disease translational contexts can benefit from updateable graph infrastructure rather than static tables [14]. The proposed framework would therefore require scheduled refreshes of gene–disease annotations, pathway resources, drug-target knowledge, and orphan designation records to remain useful.

GRAPH Enrichment with Genes, Pathways, and Orphan Drug Labels

Encoding Rare Disease Genetics

Rare disease genetics would be encoded through explicit edges linking genes, variants where available, and disease entities. DisGeNET can support association-level relationships [3], while rare disease knowledge graph resources show how curated disease concepts can be represented in a graph structure suitable for computational discovery [12]. Ontology-enriched rare disease clustering further suggests that genetic evidence can be interpreted alongside disease similarity and semantic context rather than in isolation [18]. For framework use, each genetic edge should distinguish causal, contributory, and associative evidence so that inference does not treat all gene–disease links as equivalent, and negative-sampling considerations from rare disease prediction pipelines illustrate why absent links should not be assumed to mean absence of biology [26].

Representing Biological Pathways and Drug Targets

Biological pathways provide the intermediate layer that connects rare disease genes to druggable mechanisms. Reactome [4], KEGG, and WikiPathways can each contribute pathway membership and process-level relationships, while DrugBank [5] and ChEMBL [6] can link drugs to protein targets, mechanisms, and bioactivity evidence. This structure enables a multi-step mechanistic path in which a drug modulates a target, the target participates in a pathway, and the pathway is implicated by a rare disease gene. The framework should represent pathway evidence as supportive and interpretable rather than determinative, because pathway membership alone does not prove that a drug will modify disease-relevant biology.

Incorporating Orphan Drug Labels as Regulatory Evidence

Orphan drug labels and designations would be encoded as regulatory evidence nodes connected to drug entities, disease entities, approval status, indication language, and relevant provenance. Studies of orphan drug designations show that these records capture development patterns and disease focus areas that are directly relevant to rare disease translation [1, 2]. Rare disease research knowledge graphs demonstrate that non-molecular evidence can be integrated with disease concepts, supporting the idea that regulatory records can be represented alongside mechanistic biology [13]. When combined with literature-based rare disease repurposing embeddings [11] and rare-disease-specific drug discovery graphs [27], this regulatory layer could help distinguish candidates with mechanistic plausibility alone from candidates that also have orphan-development precedent.

Drug Repurposing Inference and Hypothesis Generation

Graph-Based Hypothesis Generation Algorithms

The reasoning layer would use knowledge graph completion, path-based inference, and graph neural network approaches to identify drug–disease pairs that are not explicitly connected but are supported by mechanistic paths. Literature-based embedding for rare disease repurposing demonstrates how textual and graph-derived relationships can be used to propose candidate therapeutic links [11], while multimodal knowledge graph embedding illustrates how disease-specific reasoning can combine heterogeneous evidence [21]. General drug-repurposing frameworks such as KG-Predict show how graph computation can infer candidate drug indications from structured biomedical relationships [22]. In this framework, candidate hypotheses would be generated when a drug connects to a rare disease through target genes, pathways, disease similarity, or orphan designation evidence, but these links would remain hypotheses requiring expert review.

Ranking and Filtering Candidates

Candidate ranking would combine mechanistic plausibility, graph proximity, regulatory relevance, and evidence provenance rather than relying on a single numerical prediction. Biomedical knowledge graph learning can extend guilt-by-association reasoning across multiple network layers [10], and network-based deep learning approaches have shown how drug repositioning models can integrate complex biomedical relationships [16]. Graph neural network systems for drug repurposing further suggest that drugs, targets, and diseases can be jointly modeled to prioritize plausible therapeutic links [17, 24]. The proposed framework would filter out drug–disease pairs already represented as approved or extensively studied indications, while flagging candidates with related orphan designations as potentially more translationally actionable.

Table 2 provides a ranking and explainability framework that separates mechanistic support, regulatory precedent, provenance strength, novelty, and human-review readiness in rare disease repurposing.

Table 2. Candidate-ranking and explainability framework for rare disease drug repurposing hypotheses

Ranking domain	What the framework evaluates	Example graph evidence	How it strengthens candidate prioritization	Explanation shown to user	Failure mode controlled
Mechanistic plausibility	Whether the drug modulates a target, gene product, or biological process plausibly connected to the rare disease	Drug → target → pathway → rare disease gene → disease	Prevents prioritization based only on superficial disease similarity or regulatory coincidence	Display the shortest and strongest mechanistic paths, including target, pathway, and disease-gene links	Overranking drugs with orphan precedent but weak biological relevance
Pathway convergence	Whether multiple disease genes, targets, or related conditions converge on the same pathway or biological process	Multiple genes associated with a rare disease map to a shared pathway that includes a druggable target	Identifies hypotheses supported by convergent biology rather than a single fragile edge	Show a pathway-centered explanation panel listing genes, targets, and pathway memberships	Treating isolated or weak annotations as robust mechanistic support
Regulatory precedent	Whether the drug has orphan designation, approved indication evidence, or label history in the same or a related disease area	Drug → orphan designation → related rare condition; drug → label indication → phenotype overlap	Helps distinguish mechanistically plausible candidates from candidates with more actionable development history	Show regulatory-status node, indication language, jurisdiction, and date	Treating regulatory history as proof of efficacy in the target disease
Evidence provenance strength	Whether supporting edges come from curated databases, regulatory documents, peer-reviewed literature, or weakly mapped sources	Edge-level provenance records connected to each drug–disease path	Allows users to prioritize candidates supported by auditable and higher-quality sources	Display source badges, timestamps, database versions, and conflict flags	Black-box ranking without traceable evidence
Novelty and non-redundancy	Whether the drug–disease pair is already approved, heavily studied, or directly encoded as a known indication	Candidate pair compared with existing indication and orphan-designation edges	Focuses researcher attention on underexplored but plausible repurposing opportunities	Mark candidates as "known," "related precedent," "indirect support," or "novel hypothesis"	Producing trivial recommendations that rediscover already established uses
Rare disease specificity	Whether evidence truly applies to the target rare disease rather than only to a broad disease family or neighboring ontology class	Disease ontology paths; phenotype overlap; gene–disease specificity; related-condition links	Reduces inappropriate transfer of evidence from loosely related conditions	Show the level of disease mapping: exact disease, subtype, related condition, or phenotype-level match	Overgeneralizing from common or better-studied diseases to ultra-rare diseases
Safety and development feasibility context	Whether the drug's known pharmacology, label history, or development status creates practical constraints or opportunities	Drug → label indication; drug → safety-related label text; drug → development status	Supports translational triage by identifying candidates that may be plausible but impractical or high-risk	Provide a development-context card summarizing approved use, orphan precedent, and cautionary evidence	Prioritizing candidates that are mechanistically attractive but unsuitable for development
Human-review readiness	Whether the candidate can be reviewed through clear paths, interpretable evidence, and editable expert annotations	Candidate hypothesis node → evidence paths → provenance → reviewer status	Converts AI output into a research workflow rather than an automated decision	Show candidate rationale, competing explanations, weak links, and reviewer notes	Presenting graph-generated candidates as final therapeutic conclusions

Case Studies and Retrospective Validation

Retrospective validation would conceptually test whether known rare disease repurposing examples could be recovered after temporarily hiding the direct drug–disease relationship from the graph. This strategy is consistent with prior graph-based repurposing work in which known associations are used to evaluate whether a model can rediscover established therapeutic relationships [7, 22]. Rare disease studies using knowledge graph embedding and rare-disease-specific graph resources provide suitable methodological precedents for examining whether sparse disease evidence can still support meaningful candidate recovery [11, 27]. The framework should report such evaluation qualitatively as a validation strategy, not as artificial performance claims or unsupported experimental results.

Explainability and User Interaction with the Knowledge Graph

Evidence Path Visualization

Each predicted drug–disease hypothesis would be accompanied by one or more evidence paths showing how the drug is connected to the rare disease through targets, genes, pathways, phenotypes, or orphan designation records. Path-based knowledge graph completion is particularly relevant because it emphasizes explanations that can be inspected as sequences of biomedical relationships rather than only as opaque scores [19]. Case-based explainable graph neural network methods also suggest that users can better interpret predictions when they are connected to mechanistic analogues or precedent cases [20]. In the proposed interface, an orphan designation edge would be visually distinct from a molecular edge so that users can distinguish regulatory precedent from biological mechanism.

Interactive Querying and Hypothesis Refinement

The user interface would allow translational researchers to query the graph by disease, gene, pathway, drug target, orphan designation status, or evidence source. Clinician-centered drug repurposing frameworks show the importance of designing AI outputs around the needs of human decision-makers rather than treating ranking alone as the final product [25]. Rare disease clinical trial knowledge graphs further indicate that structured graph interfaces can support exploration of study-related and disease-specific evidence [14]. Human-in-the-loop refinement would allow researchers to add newly reviewed literature, mark weak paths as less persuasive, or prioritize hypotheses that align with a foundation's clinical development strategy.

Integration Into Rare Disease Research And Clinical Translation

Supporting Academic and Industry Rare Disease Translational Programs

The framework could be deployed as a web portal, internal graph platform, or API for rare disease consortia, academic groups, patient foundations, and industry translational teams. Rare disease research discovery graphs show how funding, scientific evidence, and disease entities can be connected to support research navigation [13], while clinical trial knowledge graph work demonstrates the value of structured rare disease information for translational planning [14]. Precision-medicine knowledge graphs also provide a precedent for organizing heterogeneous biomedical knowledge into resources that can support decision-making across disease contexts [8]. In practice, the framework would help teams generate evidence-backed repurposing proposals rather than replace clinical, regulatory, or biological judgment.

Prioritization of Repurposing Candidates for Clinical Development

By incorporating orphan designation status and label information, the framework would allow users to identify candidates with both mechanistic plausibility and regulatory precedent. Analyses of orphan drug designations show that designation records contain meaningful information about rare disease development trends and therapeutic focus [1, 2]. DrugBank and ChEMBL can complement this regulatory layer with pharmacological and target information, helping users interpret whether a regulatory precedent is biologically relevant to the disease under consideration [5, 6]. This combination could guide translational teams toward candidates whose safety, mechanism, or development history makes them more suitable for formal evaluation.

Evaluation Strategy

Retrospective Repurposing Recovery

A retrospective evaluation strategy would hide known drug–disease links and examine whether the graph can recover them through indirect paths involving genes, pathways, targets, and orphan regulatory evidence. Hetionet provides a precedent for evaluating repurposing hypotheses using integrated biomedical networks [7], while rare disease embedding work demonstrates how sparse disease contexts can still be explored computationally [11]. Rare-disease-specific drug discovery graphs could support similar recovery exercises when they encode relationships among rare conditions, drugs, and mechanistic evidence [27]. Such evaluation should be framed as a stress test of framework logic rather than proof that a candidate will succeed clinically.

Novel Hypothesis Yield and Plausibility

Novel hypotheses should be assessed by rare disease experts for mechanistic coherence, disease relevance, safety considerations, and regulatory opportunity. Explainable drug repurposing methods support this review process by exposing the paths that connect a predicted drug to a disease [19], while clinician-centered AI approaches emphasize that model outputs should be interpretable and actionable for expert users [25]. Biomedical graph learning and graph neural network approaches can generate candidate associations, but expert review remains essential because graph connectivity may reflect research bias,

incomplete annotation, or indirect biological relevance [10, 17]. The most useful output would therefore be a prioritized and explainable hypothesis set, not an automated claim of therapeutic efficacy.

Limitations

Data Incompleteness and Bias in Public Databases

The proposed graph would inherit incompleteness, annotation bias, and uneven disease coverage from its source databases. DisGeNET, Reactome, KEGG, and WikiPathways each provide valuable structured knowledge, but their coverage differs across genes, diseases, pathways, and curation contexts [3, 4]. Rare disease ontology-enriched graph work highlights that sparse information can sometimes be mitigated through semantic relatedness, but such relatedness must not be mistaken for direct biological evidence [18]. Negative-sampling research in rare disease repurposing also underscores that missing graph links are not equivalent to negative facts, which is especially important for ultra-rare conditions [26].

Static Nature of the Graph

A knowledge graph reflects the sources and update cycle from which it is built, so stale regulatory or biomedical data could mislead users if not refreshed. DrugBank, ChEMBL, Reactome, KEGG, and WikiPathways all evolve over time, making update governance essential for any repurposing framework that depends on these resources [4-6]. Rare disease research and clinical trial knowledge graphs further show that translational evidence changes as studies, designations, and research priorities shift [13, 14]. The framework should therefore include versioning, timestamped provenance, and scheduled review of regulatory records, while acknowledging that it cannot capture unpublished evidence or clinical judgment.

Conclusion

The proposed AIF framework positions rare disease drug repurposing as an integrative knowledge problem rather than a single-model prediction task. By connecting genes, pathways, drugs, targets, orphan designations, and label indications, the framework would allow researchers to examine candidate therapies through both mechanistic and regulatory lenses.

Its main strength is the explicit treatment of orphan drug labels as computable evidence within the graph. This regulatory-aware design would make it possible to distinguish candidates supported only by biological similarity from candidates that also have relevant development precedent, safety context, or indication history.

Important challenges remain, including sparse evidence for ultra-rare diseases, variable database coverage, entity-normalization errors, and the need to keep the graph synchronized with new biomedical and regulatory information. Any AI-generated hypothesis would still require expert interpretation, experimental investigation, and prospective clinical evaluation before it could influence patient care.

Sustaining such a framework would require collaboration among rare disease foundations, regulatory scientists, informaticians, clinicians, and biomedical database maintainers. With shared governance and transparent provenance, a regulatory-aware biomedical knowledge graph could become a practical infrastructure for rare disease translational discovery.

Acknowledgments: None

Conflict of interest: None

Financial support: None

Ethics statement: None

References

1. Fermaglich LJ, Miller KL. A comprehensive study of the rare diseases and conditions targeted by orphan drug designations and approvals over the forty years of the Orphan Drug Act. Orphanet J Rare Dis. 2023;18(1):163.
2. Miller KL, Fermaglich LJ, Maynard J. Using four decades of FDA orphan drug designations to describe trends in rare disease drug development: substantial growth seen in development of drugs for rare oncologic, neurologic, and pediatric-onset diseases. Orphanet J Rare Dis. 2021;16(1):265.
3. Piñero J, Ramírez-Anguita JM, Saüch-Pitarch J, Ronzano F, Centeno E, Sanz F, et al. The DisGeNET knowledge platform for disease genomics: 2019 update. Nucleic Acids Res. 2020;48(D1):D845-55.
4. Milacic M, Beavers D, Conley P, Gong C, Gillespie M, Griss J, et al. The Reactome pathway knowledgebase 2024. Nucleic Acids Res. 2024;52(D1):D672-8.
5. Wishart DS, Feunang YD, Guo AC, Lo EJ, Marcu A, Grant JR, et al. DrugBank 5.0: a major update to the DrugBank database for 2018. Nucleic Acids Res. 2018;46(D1):D1074-82.
6. Zdrazil B, Felix E, Hunter F, Manners EJ, Blackshaw J, Corbett S, et al. The ChEMBL database in 2023: a drug discovery platform spanning multiple bioactivity data types and time periods. Nucleic Acids Res. 2024;52(D1):D1180-92.

7. 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.
8. Chandak P, Huang K, Zitnik M. Building a knowledge graph to enable precision medicine. Sci Data. 2023;10(1):67.
9. Boudin M, Diallo G, Drancé M, Mougin F. The OREGANO knowledge graph for computational drug repurposing. Sci Data. 2023;10(1):871.
10. Bang D, Lim S, Lee S, Kim S. Biomedical knowledge graph learning for drug repurposing by extending guilt-by-association to multiple layers. Nat Commun. 2023;14(1):3570.
11. 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. Pac Symp Biocomput. 2020:463-74.
12. Zhu Q, Nguyen DT, Grishagin I, Southall N, Sid E, Pariser A. An integrative knowledge graph for rare diseases, derived from the Genetic and Rare Diseases Information Center (GARD). J Biomed Semantics. 2020;11(1):13.
13. Zhu Q, Nguyễn DT, Sheils T, Alyea G, Sid E, Xu Y, et al. Scientific evidence based rare disease research discovery with research funding data in knowledge graph. Orphanet J Rare Dis. 2021;16(1):483.
14. Yang JP, Leadman D, Ballew RM, Sid E, Xu Y, Mathé EA, et al. User centered rare disease clinical trial knowledge graph (RCTKG). Stud Health Technol Inform. 2024;310:94.
15. Perdomo-Quinteiro P, Belmonte-Hernández A. Knowledge graphs for drug repurposing: a review of databases and methods. Brief Bioinform. 2024;25(6):bbae461.
16. 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.
17. Ayuso-Munoz A, Prieto-Santamaria L, Ugarte-Carro E, Serrano E, Rodriguez-Gonzalez A. Uncovering hidden therapeutic indications through drug repurposing with graph neural networks and heterogeneous data. Artif Intell Med. 2023;145:102687.
18. Sanjak J, Binder J, Yadaw AS, Zhu Q, Mathé EA. Clustering rare diseases within an ontology-enriched knowledge graph. J Am Med Inform Assoc. 2024;31(1):154-64.
19. Jiménez A, Merino MJ, Parras J, Zazo S. Explainable drug repurposing via path based knowledge graph completion. Sci Rep. 2024;14(1):16587.
20. Gonzalez-Cavazos AC, Tu R, Sinha M, Su AI. A case-based explainable graph neural network framework for mechanistic drug repositioning. Bioinformatics. 2026:btag008.
21. Zhu C, Yang Z, Xia X, Li N, Zhong F, Liu L. Multimodal reasoning based on knowledge graph embedding for specific diseases. Bioinformatics. 2022;38(8):2235-45.
22. Gao Z, Ding P, Xu R. KG-Predict: a knowledge graph computational framework for drug repurposing. J Biomed Inform. 2022;132:104133.
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. Li W, Ma W, Yang M, Tang X. Drug repurposing based on the DTD-GNN graph neural network: revealing the relationships among drugs, targets and diseases. BMC Genomics. 2024;25(1):584.
25. Huang K, Chandak P, Wang Q, Havaldar S, Vaid A, Leskovec J, et al. A foundation model for clinician-centered drug repurposing. Nat Med. 2024;30(12):3601-13.
26. Bianchi N, Lefebvre AE, Wolstencroft KJ, Spruit M. Design and evaluation of semantically-valid negative samples integration techniques for scalable semi-automated drug repurposing prediction pipelines in rare disease research. BMC Bioinformatics. 2026;27(1).
27. Zhu C, Xia X, Li N, Zhong F, Yang Z, Liu L. RDKG-115: assisting drug repurposing and discovery for rare diseases by trimodal knowledge graph embedding. Comput Biol Med. 2023;164:107262.