Pharmacophore

ISSN-2229-5402

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

Artificial Intelligence for Drug Repurposing Across Molecular and Clinical Evidence: A Scoping Review of Knowledge Graphs, Transcriptomic Signatures, Real-World Data, and Translational Validation Pathways

Maria Lopez[1]*, David Thompson[2], Omar Haddad[3], Lisa Johansson[4]

1. ***Department of Drug Repurposing and AI-Based Discovery, IE University, Madrid, Spain.***
2. ***Department of Pharmaceutical Data Science and Real-World Evidence, University of Auckland, Auckland, New Zealand.***
3. ***Department of Computational Drug Repurposing and Knowledge Graphs, University of Jordan, Amman, Jordan.***
4. ***Department of Pharmaceutical Bioinformatics and Translational Medicine, Lund University, Lund, Sweden.***

ARTICLE INFO

Received:
12 June 2026
Received in revised form:
04 October 2026
Accepted:
06 October 2026
Available online:
28 October 2026

Keywords: Artificial intelligence, Drug repurposing, Knowledge graphs, Transcriptomics, Real-world data, Translational validation

ABSTRACT

Artificial intelligence has expanded drug repurposing from similarity-based candidate searches toward heterogeneous systems that combine molecular networks, transcriptomic states, genetics, biomedical knowledge graphs, clinical records, and other real-world data. The resulting literature is difficult to interpret because computational prediction, mechanistic plausibility, experimental confirmation, and patient-level validation represent different forms of evidence but are often discussed under the same language of "validation." This scoping review maps how AI-enabled drug-repurposing studies construct and connect these evidence forms and examines where translation strengthens, remains conditional, or stops. A reproducible search and screening process identified 37 eligible evidence units from 2018–2026. Evidence was mapped by dominant data modality, computational method, repurposing objective, biomedical scale, and validation stage rather than pooled across incomparable performance measures. Knowledge-graph and network approaches formed a substantial part of the mapped literature, while transcriptomic and increasingly single-cell methods provided context-specific disease-state representations. Multimodal studies linked some of these molecular representations to genetics or patient data, whereas real-world evidence supported both hypothesis generation and post-nomination evaluation. Only a subset of computational programmes traversed more than one validation setting. The evidence therefore supports viewing AI drug repurposing as a set of heterogeneous translational pathways rather than a single prediction task. Progress depends less on maximizing an isolated model score than on establishing whether successive evidence sources provide sufficiently independent, biologically relevant, and clinically interpretable support for the proposed new use.

*This is an **open-access** article distributed under the terms of the Creative Commons Attribution-Non Commercial-Share Alike 4.0 License, which allows others to remix, and build upon the work non commercially.*

To Cite This Article: Lopez M, Thompson D, Haddad O, Johansson L. Artificial Intelligence for Drug Repurposing Across Molecular and Clinical Evidence: A Scoping Review of Knowledge Graphs, Transcriptomic Signatures, Real-World Data, and Translational Validation Pathways. Pharmacophore. 2026;17(5):112-20. https://doi.org/10.51847/gQ4z9weufP

Introduction

Drug repurposing seeks new therapeutic uses for existing or previously investigated compounds and can exploit prior knowledge of pharmacology, formulation, exposure, toxicology, or human use. These advantages do not eliminate the central problem of indication-specific evidence. A compound that is already characterized for one purpose still requires a defensible biological and clinical rationale for another, while computational approaches differ substantially in what they predict and what their outputs can establish. Contemporary repurposing therefore encompasses similarity methods, network approaches, transcriptomic reversal, machine learning, target-based strategies, and combinations of these methods rather than a single discovery paradigm [1, 2].

Three evidence families illustrate this heterogeneity particularly clearly. Knowledge graphs organize relationships among drugs, diseases, genes, proteins, phenotypes, pathways, and other biomedical entities and can support link prediction or

Corresponding Author: Maria Lopez; Department of Drug Repurposing and AI-Based Discovery, IE University, Madrid, Spain. E-mail: maria.lopez@ie.edu.

mechanistic path reasoning. Transcriptomic approaches instead ask whether a compound may alter a disease-associated molecular state, with the validity of the inference depending on tissue, cell state, perturbational context, and analytical construction of the signature. Real-world data can contribute yet another evidence form by identifying treatment–outcome patterns or evaluating nominated candidates in routinely collected populations. Each source can strengthen a repurposing argument, but none is interchangeable with another [3-5].

Recent work has also blurred former methodological boundaries. Data-driven repurposing increasingly integrates drug–target resources, pathway information, graph structure, molecular features, genetics, and clinical information, while knowledge-graph research has moved from conventional similarity and embedding methods toward graph neural networks and other representation-learning architectures [6]. These developments enlarge the information accessible to a model but also create additional questions about provenance, dependence among data sources, interpretability, negative-sample construction, external validation, and the difference between recovering known associations and identifying genuinely transportable new ones [7].

The translational question becomes most visible when molecular and clinical evidence are connected within the same programme. Multimodal work in Alzheimer's disease, for example, has combined molecular profiling and systems-level analysis with electronic health record or claims-based investigation, illustrating how computational nomination can be followed by evidence at a different biological or clinical scale [8]. Such designs are important because agreement across genuinely distinct evidence sources may be more informative than repeated evaluation within the same computational representation. At the same time, cross-modal agreement cannot be assumed to be independent when modalities share databases, disease labels, training signals, or downstream outcome definitions.

The purpose of this scoping review is therefore not to determine which AI method has the highest reported performance or to calculate a pooled estimate across heterogeneous prediction tasks. It is to map the forms of evidence used in AI-enabled drug repurposing and examine how knowledge graphs, transcriptomic signatures, multimodal sources, real-world data, and translational validation are connected. The review asks which approaches dominate the literature, how molecular and clinical evidence are combined, what validation states are actually reached, and where apparent methodological progress stops short of stronger translational inference.

Materials and Methods

This review was conducted as a scoping review because the primary objective was to characterize the breadth, configuration, and translational depth of a heterogeneous evidence base rather than estimate a common treatment effect. Reporting was structured according to PRISMA-ScR principles [9]. The unit of inclusion was a distinct peer-reviewed report. Database mirrors, duplicate DOI records, and superseded versions of the same report were collapsed; distinct analyses using overlapping underlying datasets remained separate evidence units but were treated as potentially dependent during interpretation.

Searches conducted on 23 September 2026 covered publications from 2018–2026 and combined terms for drug repurposing or repositioning with knowledge graphs, graph neural networks, transcriptomics, gene-expression signatures, single-cell data, multimodal or multi-omics evidence, genetics, electronic health records, claims, real-world evidence, trial emulation, validation, and related AI or computational methods. Studies were eligible when drug repurposing was a substantive analytical purpose and the report provided extractable information on computational approach, evidence source, or validation pathway. Q1 status was not an evidence-eligibility criterion.

The search returned 121 records before normalization. Forty-three duplicate or alternate-version records were removed, leaving 78 records for title and abstract screening. Thirty-two were excluded at that stage. Forty-six reports were sought and retrieved for full assessment; nine were excluded after detailed review, leaving 37 included evidence units. Exclusions at full assessment comprised studies without a sufficiently relevant repurposing operation, companion or less complete versions, generic resources without an extractable repurposing analysis, and records peripheral to the molecular-to-clinical evidence question.

Extraction distinguished the infrastructure supplying information from the algorithm operating on it. This was particularly important for biomedical knowledge graphs, because resources such as PrimeKG integrate molecular, disease, treatment and phenotype relations within a precision-medicine graph [10], whereas PharMeBINet constructs a heterogeneous pharmacological, medical and biochemical network from a different set of sources and ontological relationships [11]. The Bioteque provides another representation strategy by producing reusable biomedical knowledge-graph embeddings derived from heterogeneous data [12]. Accordingly, graph size or breadth was not treated as a validation outcome.

For every eligible report, extraction recorded the dominant evidence modality, biomedical scale, AI or computational method, repurposing objective, validation design, external-validation status, applicability domain, reported performance information where relevant, and the strongest inference supported by the study design. Missing elements were retained as not reported rather than inferred. Because performance measures represented different prediction targets and evaluation designs, AUROC, AUPRC, ranking measures, enrichment statistics, observational effect estimates, docking results, and experimental outcomes were not standardized or quantitatively pooled.

The synthesis used a configurational evidence map. Each report was first coded by its dominant evidence-entry class and then by the furthest validation state actually documented. Validation was interpreted as a change in evidentiary setting rather than simply an additional metric: internal benchmark evaluation, external computational testing, mechanistic or docking support,

experimental confirmation, observational patient-level evaluation, and trial-emulation approaches were therefore kept analytically distinct.

The resulting methodological strata are summarized in **Table 1**.

Table 1. Evidence-entry strata used to characterize the 37 included AI drug-repurposing reports

Dominant evidence-entry stratum	Included evidence units	Typical source material	Characteristic computational operation	Validation question retained for synthesis
Knowledge graph or network dominant	18	Drug–disease–target relations, pathways, phenotypes, heterogeneous biomedical entities	Embedding, graph neural networks, metapaths, path reasoning, network propagation	Does the predicted relation generalize beyond the graph and benchmark structure that produced it?
Transcriptomic or single-cell dominant	9	Bulk expression, perturbational signatures, cell states, cell-type programmes	Signature reversal, feature learning, cell-state matching, pathway-linked modelling	Does molecular-state reversal persist across biological context and progress beyond expression-based nomination?
Real-world-data dominant	4	Electronic health records, claims, registries, medication and outcome histories	Association screening, comparative modelling, target-trial emulation	Does patient-level evidence strengthen causal interpretation or only identify a new observational signal?
Multimodal or genetics-led	3	Multi-omics, structural information, genetics, combined molecular and clinical evidence	Multimodal representation learning, integrative systems modelling, genotype-first prioritization	Are the modalities independent and do they add distinct evidence rather than correlated signals?
General computational/data-driven	2	Multi-database drug, target and disease information	Comparative data integration and candidate prioritization	How strongly does database construction determine the nominated candidates?
Computational-to-experimental bridge	1	Cross-species target similarity plus antipathogenic assays	Bioinformatic prioritization followed by laboratory testing	Does experimental activity support the predicted repurposing hypothesis within the tested biological system?

The executed selection process is shown in **Figure 1**.

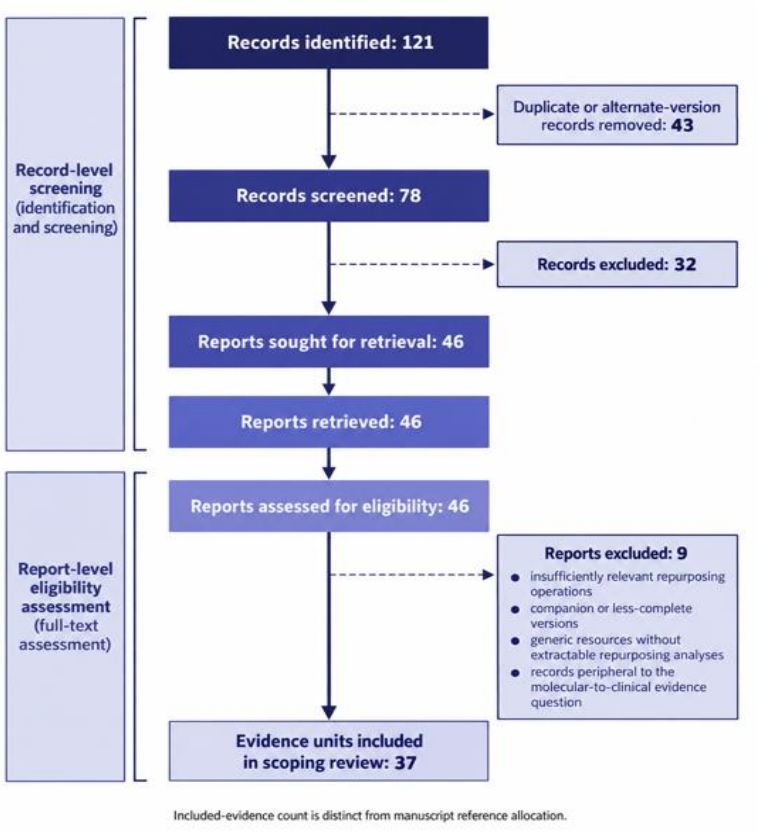


Figure 1. Executed search and selection of evidence for the AI drug-repurposing map

Results and Discussion

The 37 included evidence units did not form a single methodological progression. Instead, they entered the repurposing problem through different representations of disease, drug action, and patient context. Knowledge-graph and network approaches were the largest dominant stratum, but their internal aims ranged from general association recovery to disease-specific augmentation, mechanistic path construction, and patient-level prioritization.

Recent graph-based systems illustrate this range. A clinician-oriented foundation model uses a biomedical knowledge graph to support treatment prediction across disease contexts, including settings in which direct training examples are sparse [13]. Earlier drug–disease–target embedding work similarly demonstrated how joint representations can generate repurposing hypotheses from network structure [14]. More recent approaches add patient-level information: precision drug-repurposing models have incorporated biobank-derived polygenic risk, biomarkers, or clinical history into graph-based ranking [15], while disease-aware knowledge-graph augmentation has been developed to improve inductive reasoning for sparsely represented diseases [16]. Patient-specific loss functions extend the same principle by allowing characteristics of an individual to influence how graph relationships are learned [17]. These methods shift the unit of prediction, but patient-specific ranking remains distinct from evidence that a nominated treatment benefits a particular patient.

Other graph models place greater emphasis on mechanism or interpretability. Metapath-guided graph learning with adaptive negative sampling has been followed by molecular docking for selected candidates [18]. ExPDrug integrates transcriptomic pathway information with a knowledge graph so that candidate prioritization can be related to pathway-level features [19]. Mechanism-guided network embeddings similarly use biologically constructed paths to organize drug–disease relationships [20]. These designs can make the computational rationale more interpretable, yet the explanatory path generated by a model remains different from experimentally demonstrating the mechanism represented by that path.

The representational space is also expanding beyond conventional biomedical graphs. Prompt-based multimodal learning combines molecular and protein information within task-specific representations [21]. CAREPath combines graph structure with biomedical language-model information and semantic reasoning paths, including evaluation under cold-start conditions [22]. Genetics-informed approaches provide another entry point: genotype-first prioritization can begin with disease-associated genetic evidence and move toward candidate targets and drugs without requiring the same starting data as expression-reversal methods [23]. Expert biological knowledge can likewise be incorporated into graph neural networks to constrain or enrich learned drug–disease representations [24]. Across these approaches, added modalities alter the evidence available to a model, but they do not by themselves establish independent validation.

Repurposing platforms add a further distinction between analytical capability and evidentiary strength. Drugst.One integrates multiple systems-medicine and network-based tools within an accessible environment for drug-repurposing analyses [25]. RepurposeDrugs uses information on drug indications and clinical-development histories to support prediction of mono- and combination-therapy opportunities [26]. Such platforms broaden access to repurposing methods and data, but deployment of an integrated tool is analytically separate from validation of any specific candidate emerging from it.

Transcriptomic repurposing shows a parallel shift from population-average signatures toward cell-specific disease states. DrugReSC focuses on disease-critical cellular subpopulations in cancer rather than treating a tumour expression profile as homogeneous [27]. scDrugLink links cell-type-specific targets with perturbational drug signatures for central nervous system diseases [28], while scKAN applies interpretable nonlinear modelling to single-cell data for cell-type-specific gene discovery and drug prioritization [29]. SuperFeat likewise learns quantitative features from single-cell RNA sequencing and reports downstream biological validation in selected disease settings [30]. Together, these studies show how increasing biological resolution changes the unit against which a drug is judged: the relevant question may become whether a compound counteracts a pathological cell programme rather than whether it reverses an aggregate disease signature.

Alzheimer's disease provides particularly clear examples of this transition. Cell-type-specific meta-programs have been used to define disease-relevant molecular states for repurposing [31]. In another programme, single-cell transcriptomic network analysis nominated ketorolac and was followed by evaluation in two real-world patient databases, creating a rare connection between cellular-state modelling and independent patient-level observational evidence [32]. The latter design reaches farther across evidentiary settings than internal computational validation, although its real-world associations remain observational rather than randomized evidence of treatment effect.

Mechanistic knowledge representation and expression-network integration occupy intermediate positions within the map. An interpretable pharmacodynamic mechanism knowledge graph has been used to predict drug effects while exposing paths that connect drugs with proposed mechanisms [33]. Breast-cancer repurposing work has integrated drug-exposure expression profiles with drug–drug links in a graph-neural-network architecture [34]. Large heterogeneous graph neural networks have likewise been used to screen drug–disease relationships across broad biomedical networks [35]. The common feature is integration; the important difference is what constitutes validation. A recovered known link, an external literature-supported candidate, a mechanistically plausible path, and a candidate tested in another biological or clinical setting do not provide equivalent evidence.

Table 2 maps these differing evidence modalities according to their principal analytical role and the validation question that remains after a candidate is produced.

Table 2. How evidence modalities contribute differently to AI-enabled drug-repurposing decisions

Evidence modality or configuration	Representation supplied to the repurposing model	Characteristic AI operation	Immediate repurposing objective	Validation state commonly reached in the mapped evidence	Translational question left unresolved
Biomedical knowledge graphs	Drug–disease–target–phenotype–pathway relations	Link prediction, embedding, GNN propagation, path reasoning	Recover or nominate therapeutically plausible associations	Internal or external computational evaluation; case-study support	Will the relationship survive testing outside the graph's source structure and benchmark?
Mechanism- or pathway-augmented graphs	Curated biological pathways, pharmacodynamic relations, transcriptomic pathway states	Mechanism-guided embedding or interpretable graph reasoning	Prioritize candidates with an explanatory biological route	Computational evaluation, sometimes docking or literature corroboration	Is the inferred pathway experimentally active in the relevant disease system?
Bulk transcriptomic signatures	Disease and drug-induced expression states	Signature reversal, similarity or learned feature comparison	Identify drugs predicted to oppose a disease-associated state	Computational and context-specific molecular validation	Does reversal persist across tissue, dose, timing, cell type and disease heterogeneity?
Single-cell transcriptomics	Cell types, cell states and disease-critical subpopulations	Cell-state feature learning, programme matching, target–signature linkage	Identify compounds aligned with pathogenic cellular programmes	Computational evaluation; selected experimental or patient-level follow-up	Does cell-specific molecular correction translate into organism- or patient-level benefit?
Multimodal/genetics-informed evidence	Molecular structure, proteins, genetics, pathways and/or clinical attributes	Joint representation learning or staged evidence integration	Prioritize candidates supported across heterogeneous evidence sources	Predominantly computational; occasional cross-domain follow-up	Are modalities genuinely independent and concordant, or do they reproduce shared upstream information?
Real-world clinical data	EHR, claims, registries, medication exposure and outcomes	Association screening, comparative modelling, causal-design emulation	Generate or evaluate candidate indications in treated populations	Observational association or target-trial emulation	Can residual confounding, selection and measurement processes be sufficiently controlled to support causal interpretation?

Real-world evidence constitutes its own heterogeneous field rather than the final stage of every repurposing programme. A recent survey of EHR-based repurposing shows that clinical data have been used both to discover new indications and to validate hypotheses originating elsewhere, with substantial methodological variation in data preprocessing, representation, comparator construction, and validation [36]. Target-trial emulation provides a more explicitly causal design strategy. High-throughput emulation in Alzheimer's disease has demonstrated how candidate drugs can be evaluated using large observational datasets while making treatment assignment, comparators, follow-up and confounding adjustment more explicit [37]. Even within such designs, model choice and covariate balance remain consequential; emulation strengthens the observational design but does not create randomization.

Conversely, real-world data can operate at an earlier hypothesis-generation stage. Outcome-wide analysis of sodium–glucose cotransporter-2 inhibitor users has been used to identify associations that could motivate investigation of additional therapeutic effects [38]. Such signals are useful precisely because they expose patterns in treated populations, but multiplicity, confounding and health-system measurement processes mean that an unexpected association is not itself a validated new indication.

Experimental follow-up offers a different route across the prediction–validation boundary. In malaria, protein-target similarity was used as a computational repurposing strategy and related to in-vitro antipathogenic activity against *Plasmodium falciparum* [39]. Here, the evidence changes setting from computational similarity to laboratory activity. The transition provides stronger biological support for the nominated relationship while remaining bounded by the assay, organism, exposure conditions and absence of human therapeutic evaluation.

Across the mapped literature, translational depth therefore cannot be represented by one ordered model-performance scale. Some studies strengthen a candidate by testing it in a biologically different system; some move from molecular data to patient-level observation; others improve internal generalization or interpretability without changing evidentiary setting. **Figure 2** organizes these pathways as a non-geographic evidence field rather than a universal pipeline.

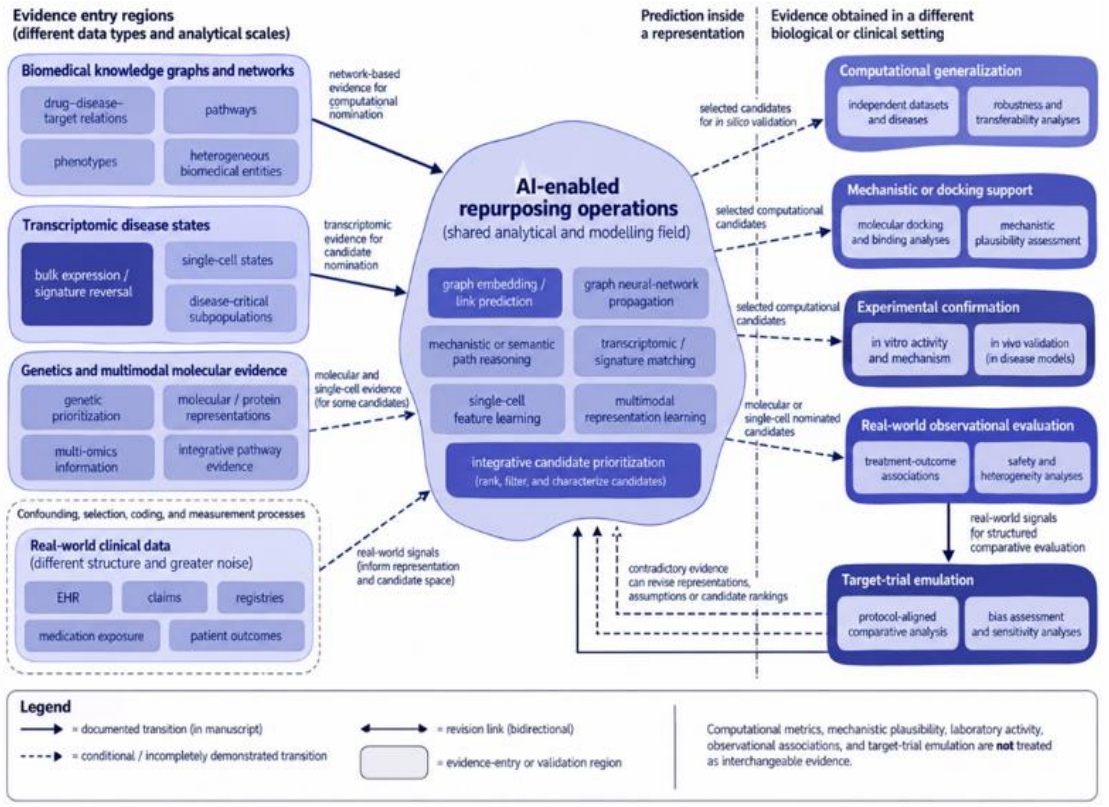


Figure 2. Where AI drug-repurposing evidence changes form before becoming translationally informative

The evidence map shows that AI-enabled drug repurposing is more mature as a system for generating and organizing hypotheses than for establishing translational validity. Knowledge-graph, graph-learning, and EHR literatures remain methodologically diverse, but many studies evaluate candidates within the same representational environment that produced them [3, 7, 36]. Internal benchmarks can establish ranking or held-out generalization; they cannot determine whether a drug–disease relationship persists in a different biological system, population, or clinical setting.

This boundary is important for graph models. Foundation models and patient-specific graph learning can change how evidence is represented and how candidates are ranked [13]. Biobank-informed prioritization can incorporate patient characteristics [15], while disease-aware augmentation can improve inference for sparsely represented conditions [16]. These advances strengthen modelling but do not themselves create clinical evidence. Translational gain occurs when a candidate is challenged by information that is sufficiently independent of the graph structure, labels, or data used for nomination.

Transcriptomic approaches create a related trade-off. Single-cell methods sharpen biological context by locating disease-associated programmes within particular cell states [27], yet the resulting predictions may become more sensitive to tissue, disease stage, exposure, and preprocessing. Experimental follow-up therefore adds a different form of support from improved signature matching [30]. The ketorolac programme extends farther by following molecular nomination with analyses in independent real-world databases [32], although observational confirmation remains vulnerable to confounding.

Real-world evidence itself ranges from hypothesis generation to more explicitly causal designs. Target-trial emulation can formalize treatment assignment, comparators, follow-up, and confounding adjustment [37], whereas outcome-wide studies are well suited to detecting candidate associations [38]. Neither should be labelled simply “clinical validation.” Across modalities, the strongest translational inference arises when evidence changes setting and tests a distinct assumption rather than reproducing another correlated view of the same underlying information.

The same principle applies to apparent multimodal concordance. Combining molecular, pathway, graph, and clinical information can broaden the evidentiary field, as demonstrated by integrative Alzheimer’s disease work [8]. However, pathway-informed graph models may still derive multiple features from overlapping source databases [19], while multimodal representation learning can improve prediction without moving outside the original computational task [21]. Consequently, the review treats independence of evidence as a property to be demonstrated rather than assumed. A second modality is most informative when it introduces a genuinely different measurement process, population, biological scale, or validation setting.

Table 3 summarizes validation transitions that change the evidentiary meaning of an AI-generated repurposing candidate.

Table 3. Validation transitions that change the evidentiary meaning of an AI-generated repurposing candidate

Starting state	Next setting	Evidence added	Residual limitation
Graph ranking	External computational dataset	Generalization beyond the original split	Shared source bias may remain

Molecular hypothesis	Docking/mechanistic modelling	Structural or mechanistic plausibility	Still model-dependent
Transcriptomic/network nomination	Experimental system	Biological activity in a new setting	Preclinical context may not reproduce patients
Molecular nomination	EHR/claims/registry analysis	Patient-level association	Confounding and measurement remain
Observational signal	Target-trial emulation	Explicit comparator and causal-design assumptions	No randomization; residual confounding
Cross-domain support	Prospective clinical evaluation	Direct benefit–risk test	Population and endpoint generalizability remain

Research Gaps and Translational Implications

The principal gap is not candidate generation but independent evidence transition. Many studies improve representation, ranking, negative sampling, or mechanistic explanation while remaining within computational evaluation. Genetics-informed prioritization likewise requires independent experimental, pharmacological, or clinical confirmation [23]. Future programmes should therefore define prospectively what evidence could falsify a nominated hypothesis, rather than treating another concordant model score as sufficient validation.

Cross-scale combinations remain comparatively sparse. Studies integrating molecular and clinical information show that such bridges are feasible [8], and molecular-to-RWD validation provides one concrete route [32]. Future work should report data provenance, dependence among modalities, temporal separation of development and validation, applicability domains, and whether downstream testing used information unavailable during nomination. Clinical claims should also track design: target-trial emulation [37], outcome-wide association screening [38], and laboratory testing [39] resolve different uncertainties.

Reporting practice is therefore itself a translational bottleneck. Data-driven repurposing studies should make source provenance, candidate-selection rules, benchmark construction, temporal separation, negative-sample strategy, and external-validation status explicit [6]. Knowledge-graph studies additionally need to distinguish graph completion from evidence obtained outside the graph [7]. For clinical-data analyses, the information required to interpret treatment assignment, comparators, outcome ascertainment, missingness, and transportability should be reported in enough detail to establish what kind of validation was actually performed [36]. Without these distinctions, superficially similar claims of "validation" can describe substantially different evidentiary achievements.

Table 4 summarizes underrepresented evidence combinations and the translational questions they leave unresolved.

Table 4. Underrepresented evidence combinations and the translational questions they leave unresolved

Underrepresented combination	Missing bridge	Unresolved question	Priority
Knowledge graph + single-cell state + patient outcomes	Cross-scale validation	Does the predicted relation operate in pathogenic cells and treated patients?	Link cell-specific mechanisms to independent clinical datasets
Genetics-led nomination + pharmacological perturbation	Functional validation	Does target direction correspond to achievable drug action?	Add perturbation and exposure-relevant testing
Multimodal model + temporally separate cohort	Independent generalization	Does performance persist across sources and time?	Predefine external cohorts and prevent leakage
Transcriptomic reversal + prospective assessment	Clinical validation	Does molecular correction predict patient benefit?	Use prospectively defined endpoints
RWD candidate + mechanistic follow-up	Biological corroboration	Is the association mechanistically coherent?	Pair causal-design analysis with independent experiments

Conclusion

AI-enabled drug repurposing is better understood as a translation problem than as one prediction task. Knowledge graphs, transcriptomic signatures, multimodal molecular information, and real-world data contribute different forms of evidence. A candidate becomes more credible when it survives testing in a new setting that challenges assumptions made upstream. AI can narrow and structure the repurposing search space, but a new indication requires validation matched to the biological and clinical uncertainty that remains.

Acknowledgments: None

Conflict of interest: None

Financial support: None

Ethics statement: None

References

1. 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
2. Jarada TN, Rokne JG, Alhajj R. A review of computational drug repositioning: Strategies, approaches, opportunities, challenges, and directions. J Cheminform. 2020;12(1):46. doi:10.1186/s13321-020-00450-7
3. Perdomo-Quinteiro P, Belmonte-Hernández A. Knowledge graphs for drug repurposing: A review of databases and methods. Brief Bioinform. 2024;25(6). doi:10.1093/bib/bbae461
4. Tan GSQ, Sloan EK, Lambert P, Kirkpatrick CMJ, Ilomäki J. Drug repurposing using real-world data. Drug Discov Today. 2023;28(1):103422. doi:10.1016/j.drudis.2022.103422
5. Koudijs KKM, Böhringer S, Guchelaar HJ. Validation of transcriptome signature reversion for drug repurposing in oncology. Brief Bioinform. 2023;24(1). doi:10.1093/bib/bbac490
6. Savander S, Curabaz NN, Abbasi AM, Khan A, Saeed K, Tanoli Z. Data-driven strategies for drug repurposing: Insights, recommendations, and case studies. Brief Bioinform. 2025;26(6). doi:10.1093/bib/bbaf625
7. Wei S, Sasi C, Piepenbrock J, Huynen MA, 't Hoen PAC. The use of knowledge graphs for drug repurposing: From classical machine learning algorithms to graph neural networks. Comput Biol Med. 2025;196(Pt C):110873. doi:10.1016/j.compbiomed.2025.110873
8. Mottaqi M, Zhang P, Xie L. Integrating explainable artificial intelligence with multiomics systems biology and electronic health record data mining for personalized drug repurposing in Alzheimer's disease. Brief Bioinform. 2025;26(6). doi:10.1093/bib/bbaf623
9. Tricco AC, Lillie E, Zarin W, O'Brien KK, Colquhoun H, Levac D, et al. PRISMA extension for scoping reviews (PRISMA-ScR): Checklist and explanation. Ann Intern Med. 2018;169(7):467-73. doi:10.7326/M18-0850
10. Chandak P, Huang K, Zitnik M. Building a knowledge graph to enable precision medicine. Sci Data. 2023;10(1):67. doi:10.1038/s41597-023-01960-3
11. Königs C, Friedrichs M, Dietrich T. The heterogeneous pharmacological medical biochemical network PharMeBINet. Sci Data. 2022;9(1):393. doi:10.1038/s41597-022-01510-3
12. Fernández-Torras A, Duran-Frigola M, Bertoni M, Locatelli M, Aloy P. Integrating and formatting biomedical data as pre-calculated knowledge graph embeddings in the Bioteque. Nat Commun. 2022;13(1):5304. doi:10.1038/s41467-022-33026-0
13. 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. doi:10.1038/s41591-024-03233-x
14. Moon C, Jin C, Dong X, Abrar S, Zheng W, Chirkova RY, et al. Learning drug-disease-target embedding (DDTE) from knowledge graphs to inform drug repurposing hypotheses. J Biomed Inform. 2021;119:103838. doi:10.1016/j.jbi.2021.103838
15. Oğuztüzün Ç, Gao Z, Li H, Xu R. Precision drug repurposing (PDR): Patient-level modeling and prediction combining foundational knowledge graph with biobank data. J Biomed Inform. 2025;163:104786. doi:10.1016/j.jbi.2025.104786
16. Oğuztüzün Ç, Gao Z, Li H, Xu R. KGiA: Drug repurposing through disease-aware knowledge graph augmentation. J Biomed Inform. 2025;168:104857. doi:10.1016/j.jbi.2025.104857
17. Oğuztüzün Ç, Gao Z, Li J, Qi X, Koyutürk M, Xu R. Learning to take it personally: Precision drug repurposing through patient-specific loss on knowledge graphs using Biobank data. J Biomed Inform. 2026;179:105039. doi:10.1016/j.jbi.2026.105039
18. Zhou Y, Shi X, Wang L, Xu J, Li D, Chen C. Drug repositioning with metapath guidance and adaptive negative sampling enhancement. J Biomed Inform. 2025;171:104916. doi:10.1016/j.jbi.2025.104916
19. Kim J, Jang H, Park Y, Jung I, Jo K. ExPDrug: Integration of an interpretable neural network and knowledge graph for pathway-based drug repurposing. Comput Biol Med. 2025;187:109729. doi:10.1016/j.compbiomed.2025.109729
20. Zeng Y, Wang L, Liu X. BioMNEDR: Mechanism-guided network embedding for drug repurposing. Brief Bioinform. 2026;27(2). doi:10.1093/bib/bbag101
21. Liu J, U K, Rana D, Zhang SM, Yu J, Yang S, et al. Prompt-based multimodal representation learning for drug repurposing. Brief Bioinform. 2025;26(6). doi:10.1093/bib/bbaf636
22. Song H, Bang D, Koo B, Kim S, Lee S. CAREPath: Semantic context-aware reasoning paths with mechanism-augmented embeddings for drug repurposing. Brief Bioinform. 2026;27(5). doi:10.1093/bib/bbag472
23. Muneeb M, Ascher DB. G2DR: A genotype-first framework for genetics-informed target prioritization and drug repurposing. Brief Bioinform. 2026;27(4). doi:10.1093/bib/bbag427
24. Wu Z, Yan C, Chen J, Xiao S, Gao J. Drug repositioning based on expert knowledge augmented graph neural network. IEEE J Biomed Health Inform. 2026;30(6):4786-97. doi:10.1109/JBHI.2025.3633239
25. Maier A, Hartung M, Abovsky M, Adamowicz K, Bader GD, Baier S, et al. Drugst.One—a plug-and-play solution for online systems medicine and network-based drug repurposing. Nucleic Acids Res. 2024;52(W1). doi:10.1093/nar/gkae388

26. Ianevski A, Kushnir A, Nader K, Miihkinen M, Xhaard H, Aittokallio T, et al. RepurposeDrugs: An interactive web-portal and predictive platform for repurposing mono- and combination therapies. Brief Bioinform. 2024;25(4). doi:10.1093/bib/bbae328
27. Liu C, Zhang Y, Liang Y, Zhang T, Wang G. DrugReSC: Targeting disease-critical cell subpopulations with single-cell transcriptomic data for drug repurposing in cancer. Brief Bioinform. 2024;25(6). doi:10.1093/bib/bbae490
28. Huang L, Lu X, Chen D. scDrugLink: Single-cell drug repurposing for CNS diseases via computationally linking drug targets and perturbation signatures. IEEE J Biomed Health Inform. 2026;30(6):4535-47. doi:10.1109/JBHI.2025.3552536
29. He H, Tang Z, Chen G, Xu F, Hu Y, Feng Y, et al. scKAN: Interpretable single-cell analysis for cell-type-specific gene discovery and drug repurposing via Kolmogorov-Arnold networks. Genome Biol. 2025;26(1):300. doi:10.1186/s13059-025-03779-0
30. Zhong J, Yang J, Song Y, Zhang Z, Wang C, Tong R, et al. SuperFeat: Quantitative feature learning from single-cell RNA-seq data facilitates drug repurposing. Genomics Proteomics Bioinformatics. 2024;22(3). doi:10.1093/gpbjnl/qzae036
31. Zhang C, Zhang Y, Wu Z, Zhang Y, Xue F, Tan Q, et al. Screening of cell-type-specific meta-programs for drug repurposing in Alzheimer's disease. Brief Bioinform. 2026;27(4). doi:10.1093/bib/bbag411
32. Xu J, Song W, Xu Z, Danziger MM, Karavani E, Zang C, et al. Single-microglia transcriptomic transition network-based prediction and real-world patient data validation identifies ketorolac as a repurposable drug for Alzheimer's disease. Alzheimers Dement. 2025;21(1). doi:10.1002/alz.14373
33. Tanaka T, Katayama T, Imai T. Predicting the effects of drugs and unveiling their mechanisms of action using an interpretable pharmacodynamic mechanism knowledge graph (IPM-KG). Comput Biol Med. 2025;184:109419. doi:10.1016/j.compbiomed.2024.109419
34. Cui C, Ding X, Wang D, Chen L, Xiao F, Xu T, et al. Drug repurposing against breast cancer by integrating drug-exposure expression profiles and drug-drug links based on graph neural network. Bioinformatics. 2021;37(18):2930-7. doi:10.1093/bioinformatics/btab191
35. Doshi S, Chepuri SP. A computational approach to drug repurposing using graph neural networks. Comput Biol Med. 2022;150:105992. doi:10.1016/j.compbiomed.2022.105992
36. Talukdar N, Zhang X, Paithankar S, Wang H, Chen B. A survey of the use of EHR as real-world evidence for discovering and validating new drug indications. Drug Discov Today. 2025;30(12):104552. doi:10.1016/j.drudis.2025.104552
37. Zang C, Zhang H, Xu J, Zhang H, Fouladvand S, Havaldar S, et al. High-throughput target trial emulation for Alzheimer's disease drug repurposing with real-world data. Nat Commun. 2023;14(1):8180. doi:10.1038/s41467-023-43929-1
38. Tan GSQ, Morton JI, Wood S, Shaw JE, Magliano DJ, Ilomäki J. SGLT-2 inhibitor use and cause-specific hospitalization rates: An outcome-wide study to identify novel associations of SGLT-2 inhibitors. Clin Pharmacol Ther. 2024;115(6):1304-15. doi:10.1002/cpt.3194
39. Mogire RM, Miruka SA, Juma DW, McNamara CW, Andagalu B, Burrows JN, et al. Protein target similarity is positive predictor of in vitro antipathogenic activity: A drug repurposing strategy for Plasmodium falciparum. J Cheminform. 2024;16(1):63. doi:10.1186/s13321-024-00856-7