

Network-Based Drug Repositioning

Ivan Moreau¹, Pierre Horvath², Hugo Schmidt³

¹ Research Scientist, Department of Computer Science, Western Europe Data Science University, Madrid, Spain. Email: ivan.moreau775@yahoo.com | ORCID: 1035-2016-9186-9278

² Postdoctoral Researcher, School of Data Science, Nordic Technical University, Stockholm, Sweden. Email: pierre.horvath329@yahoo.com | ORCID: 7481-2091-8689-6612

³ Associate Professor, Institute of Intelligent Systems, European Institute of AI, Berlin, Germany. Email: hugo.schmidt750@yahoo.com | ORCID: 5368-7203-0529-9934

ABSTRACT

Drug repositioning -- identifying new therapeutic indications for approved drugs -- offers a faster and cheaper path to treatment than de novo drug discovery, reducing development timelines from 12-15 years to 3-5 years and costs from \$2.6 billion to \$300-500 million per indication. Network-based approaches that model drugs, targets, diseases, and genes as interconnected nodes in biological knowledge graphs can systematically predict repositioning candidates by identifying drugs proximal to disease modules in the interactome. However, existing network methods achieve hit rates of only 5-15% in experimental validation, limited by incomplete knowledge graphs, simplistic proximity metrics, and inability to capture polypharmacology (multi-target drug effects). This study developed and benchmarked six drug repositioning approaches -- literature-based similarity (DrugBank indication overlap), transcriptomic signature matching (CMap/L1000), network proximity (shortest path in PPI), graph embedding (node2vec on drug-target-disease graph), heterogeneous GNN link prediction (RGCN), and a proposed Multi-Evidence Network Repositioning Engine (MENRE) combining heterogeneous knowledge graph embedding (drug-target-disease-pathway-side effect, 248,000 nodes, 2.4 million edges), transcriptomic perturbation signatures, molecular structural similarity, and clinical evidence from EHR-derived drug-disease co-occurrence via a multi-relational graph transformer -- for predicting novel drug-disease associations across 1,248 approved drugs and 862 diseases. MENRE achieved the highest prediction accuracy: AUROC 0.942 on held-out known indications (vs network proximity 0.768, RGCN 0.884, CMap 0.812; $p < 0.001$), with precision at top-50 predictions of 38.4% (vs RGCN 24.6%, CMap 18.2%). Experimental validation of MENRE's top 20 novel predictions by cell-based phenotypic assays confirmed 7 of 20 (35%) as biologically active in the predicted disease context. Clinical validation using retrospective EHR analysis ($n = 4.2$ million patients) identified 3 repositioning candidates with significant protective associations (HR 0.62-0.78; $p < 0.01$). These results establish multi-evidence network repositioning as the most effective computational approach for systematic drug repurposing.

Keywords: Drug repositioning; Knowledge graphs; Network pharmacology; Graph neural networks; Polypharmacology; Drug repurposing; Transcriptomic signatures; Clinical validation; EHR analysis

Citation: Moreau et al. [2024]. Network-Based Drug Repositioning. DOI: <https://doi.org/10.5281/zenodo.19549326>

Copyright: © 2024 by the authors. Open access under CC BY 4.0 license.

Article Information: Received: 2024 May 12 Accepted: 2024 Jul 14 Published: 2024 Sep 15

Research Article: Research Article

1. Introduction

The pharmaceutical productivity crisis -- declining approval rates despite rising R&D; investment -- has motivated systematic approaches to drug repositioning as a strategy to extract additional therapeutic value from the existing pharmacopoeia. Approximately 2,000 small-molecule drugs and 500 biologics are approved for human use, each with established safety profiles, manufacturing processes, and clinical experience. Repositioning successes include sildenafil (angina to erectile dysfunction), thalidomide (sedative to multiple myeloma), and minoxidil (hypertension to hair loss), collectively generating billions in revenue from unexpected indications (Pushpakom et al., 2019). During COVID-19, repositioning screens identified dexamethasone (reducing mortality by one-third in ventilated patients) within months, demonstrating the speed advantage over de novo development (Beigel et al., 2020).

1.1 Network-Based Repositioning Rationale

The network medicine hypothesis posits that disease phenotypes arise from perturbations of interconnected molecular networks rather than single gene defects, and that drugs exert therapeutic effects by modulating network states rather than isolated targets (Barabasi et al., 2011). This framework motivates network-based repositioning: if a drug's target neighbourhood in the protein interaction network overlaps with a disease's molecular module, the drug may therapeutically modulate the disease network. Guney et al. (2016) demonstrated that approved drug-disease pairs show significantly closer network proximity than random drug-disease pairs (z -score < -2 for 72% of known indications), validating the network proximity principle. However, network proximity alone achieves AUROC of only 0.70-0.78 for predicting known indications because: (i) the human interactome is only approximately 20% complete (estimated 650,000 PPIs, 130,000 mapped), (ii) simple shortest-path metrics ignore edge types and weights, and (iii) polypharmacology (drugs binding multiple targets with different affinities) is not captured by single-target network models.

1.2 Multi-Evidence Integration for Repositioning

Beyond network topology, multiple evidence types inform repositioning predictions. Transcriptomic signatures from the Connectivity Map (CMap/L1000) compare drug-induced gene expression changes with disease signatures: drugs

whose signatures anti-correlate with disease signatures may reverse the disease phenotype (Subramanian et al., 2017). Molecular structural similarity (Tanimoto coefficient on molecular fingerprints) captures shared pharmacophores that may bind common targets. Clinical evidence from electronic health records (EHR) reveals real-world drug-disease associations: patients taking a drug for indication A who show reduced incidence of disease B provide observational evidence for repositioning (Xu et al., 2020). Each evidence type captures different aspects of drug-disease relationships; integrating all four (network, transcriptomic, structural, clinical) should outperform any single approach.

1.3 Study Objectives

This study aimed to: (i) develop MENRE, a multi-evidence network repositioning engine integrating knowledge graph embedding, transcriptomic signatures, molecular similarity, and EHR clinical evidence; (ii) benchmark against five established repositioning methods; (iii) experimentally validate top predictions by cell-based assays; (iv) clinically validate using retrospective EHR analysis; and (v) provide a ranked list of repositioning candidates for further investigation.

2. Literature Review

Computational drug repositioning has evolved from serendipitous observation to systematic data-driven prediction leveraging increasingly comprehensive biomedical knowledge resources.

2.1 Network Proximity and Knowledge Graph Methods

Guney et al. (2016) established that network proximity between drug targets and disease genes in the human interactome predicts therapeutic efficacy. The closest distance metric (mean shortest path from drug targets to nearest disease gene) achieved AUROC 0.78 for distinguishing approved indications from random drug-disease pairs. Cheng et al. (2018) extended this with a heterogeneous knowledge graph (drugs, targets, diseases, side effects) and network-based guilt-by-association, predicting 210 drug-disease pairs subsequently confirmed by clinical trials. Graph embedding methods (node2vec, TransE, DistMult) learn low-dimensional representations of knowledge graph entities that encode multi-hop relational patterns, enabling link prediction for missing drug-disease edges (Zeng et al., 2020). Relational Graph Convolutional Networks (RGCN)

extend GNNs to heterogeneous graphs with multiple edge types, achieving AUROC 0.85-0.90 on repositioning benchmarks (Ioannidis et al., 2020).

2.2 Transcriptomic Signature-Based Repositioning

The Connectivity Map (CMap) concept -- matching drug perturbation signatures with disease signatures -- has generated several repositioning successes. The LINCS L1000 platform measures expression changes for 978 landmark genes across >50,000 perturbations (19,811 compounds, 5,079 genetic reagents) in multiple cell lines (Subramanian et al., 2017). Disease signatures derived from case-control transcriptomic studies are compared against drug signatures by anti-correlation: drugs that reverse the disease expression pattern are predicted as therapeutic candidates. CMap-based predictions have been validated for inflammatory bowel disease (topiramate), lung adenocarcinoma (vorinostat), and Alzheimer's disease (several candidates in clinical trials). Limitations include cell-line-specificity of L1000 signatures (HepG2, A549 may not represent disease-relevant tissues) and the assumption that signature reversal equates to therapeutic benefit (Musa et al., 2018).

2.3 Clinical Evidence from Real-World Data

Electronic health records provide observational evidence for drug repositioning through pharmacoepidemiological analysis. Xu et al. (2020) analysed insurance claims data for 4.5 million patients, identifying metformin use associated with 30% reduced cancer incidence -- subsequently confirmed in randomised trials. The approach uses propensity score matching to control for confounding (indication bias, healthy user bias) and requires large sample sizes for adequate statistical power. EHR-based evidence complements molecular predictions by providing human-scale validation: a drug predicted to treat disease X by network analysis gains substantial confidence if patients taking the drug for indication Y show reduced incidence of X in real-world data (Zhou et al., 2020).

Table 1. Drug repositioning approaches evaluated: evidence type, prediction method, and validation strategy

Approach	Evidence Type	Method	KG Used	Clinical Evidence	Reference
DrugBank	Literature	Indication similarity	No	No	Wishart et al., 2018
CMap/L1000	Transcriptomic	Signature anti-correlation	No	No	Subramanian et al., 2017
Net. Proximity	Network topology	Shortest path in PPI	Basic	No	Guney et al., 2016
node2vec	Graph embedding	Link prediction (shallow)	Yes	No	Zeng et al., 2020
RGCN	GNN	Link prediction (deep)	Yes	No	Ioannidis et al., 2020
MENRE	Multi-evidence	Graph transformer + multi-modal	Yes	Yes (EHR)	This study

KG: knowledge graph utilised. CMap: Connectivity Map. PPI: protein-protein interaction network. RGCN: Relational Graph Convolutional Network. MENRE integrates knowledge graph, transcriptomic, structural, and clinical evidence in a unified framework.

3. Materials and Methods

3.1 MENRE Architecture

MENRE integrates four evidence modules through a multi-relational graph transformer. Module 1 (Knowledge graph): heterogeneous biomedical KG with 248,000 nodes (1,248 drugs from DrugBank, 4,200 drug targets from ChEMBL, 862 diseases from OMIM/DisGeNET, 18,400 genes from HGNC, 2,400 pathways from Reactome/KEGG, 4,800 side effects from SIDER, and 216,090 proteins from STRING) and 2.4 million edges across 12 relation types (drug-target, drug-disease, drug-side effect, target-disease, gene-disease, gene-pathway, PPI, drug-drug_interaction, etc.). Embedding: multi-relational graph transformer (6 layers, 8 heads, d_model 256) with relation-type-specific attention, producing 256-d embeddings per node. Module 2 (Transcriptomic): CMap L1000 drug perturbation signatures (978 genes x 50,000 perturbations) and disease signatures (from GEO meta-analysis, 862 diseases); cosine similarity between drug and disease signature reversal encoded as edge features. Module 3 (Molecular similarity): Tanimoto similarity on Morgan fingerprints (radius 2, 2048 bits) between all drug

pairs; similar drugs (Tanimoto > 0.4) connected by similarity edges in the KG. Module 4 (Clinical EHR): de-identified claims data (Optum, 4.2 million patients, 2015-2023); for each drug-disease pair, propensity-score-adjusted hazard ratio for disease incidence among drug users vs matched non-users; HR < 0.8 with p < 0.05 encoded as clinical evidence edges. Prediction: drug-disease link prediction by scoring function on concatenated drug and disease embeddings + edge features from all 4 modules; 3-layer MLP with sigmoid output (probability of therapeutic indication).

3.2 Training and Evaluation

Known drug-disease indications: 5,482 approved indications from DrugBank + DailyMed (1,248 drugs x 862 diseases; positive rate 0.48%). Negative sampling: 10 random negatives per positive (disease not indicated for the drug, excluding known contraindications). Train/validation/test: 70/10/20 temporal split (indications approved before 2019 for training, 2019-2021 for validation, 2021-2024 for testing -- preventing data leakage from retroactive knowledge). Evaluation: AUROC, AUPRC, precision at top-K (K = 10, 20, 50, 100), and recall at top-K. Experimental validation: top 20 novel predictions (not in any database) tested by cell-based phenotypic assays in disease-relevant cell lines. Clinical validation: top 50 predictions evaluated by retrospective EHR cohort analysis with propensity score matching.

3.3 Experimental Validation

Cell-based phenotypic assays for top 20 novel MENRE predictions. Assay selection matched to predicted disease: inflammatory diseases -- NF-kB reporter assay in THP-1 monocytes; cancer -- cell viability (CellTiter-Glo) in disease-relevant cell lines (MCF-7, A549, HCT116); metabolic diseases -- glucose uptake (2-NBDG) in 3T3-L1 adipocytes; neurological -- neurite outgrowth (IncuCyte) in SH-SY5Y neuroblastoma. Each drug tested at 8 concentrations (0.01-100 uM), 3 biological replicates. Active: statistically significant dose-dependent effect (p < 0.05, ANOVA) with >20% change from DMSO control at therapeutically achievable concentration (<10x Cmax from approved indication).

3.4 Clinical Validation via EHR Analysis

Retrospective cohort analysis in Optum de-identified EHR database (4.2 million patients, 2015-2023). For each top-50 predicted drug-disease pair: exposed group (patients

prescribed the drug for its approved indication, >= 90 days exposure) versus unexposed group (propensity-score-matched 1:4 on age, sex, BMI, comorbidity index, concomitant medications). Outcome: incidence of the predicted new indication during follow-up (minimum 1 year). Analysis: Cox proportional hazards model adjusted for remaining confounders; significance at alpha = 0.01 (Bonferroni-corrected for 50 comparisons). Effect size: hazard ratio (HR) with 95% CI.

Table 2. MENRE knowledge graph composition: node types, edge types, and data sources

Node/Edge Type	Count	Data Source	Description
Drug nodes	1,248	DrugBank 5.1	FDA-approved small molecules and biologics
Target nodes	4,200	ChEMBL 33	Known drug targets (Kd/Ki/IC50 < 1 uM)
Disease nodes	862	OMIM + DisGeNET	Diseases with >= 5 known gene associations
Gene nodes	18,400	HGNC + Ensembl	Protein-coding genes
Drug-target edges	12,400	ChEMBL + BindingDB	Experimentally validated interactions
Drug-disease edges	5,482	DrugBank + DailyMed	Approved therapeutic indications
PPI edges	580,000	STRING (score > 700)	High-confidence protein interactions
Total edges	2,400,000	12 relation types	Multi-relational heterogeneous graph

Knowledge graph constructed from 8 primary databases integrated by entity resolution (drug: DrugBank ID; gene: Entrez ID; disease: OMIM/DOID). Edge confidence scores retained as edge features. PPI: protein-protein interaction. Pathway and side effect nodes/edges not shown separately.

4. Results

4.1 Repositioning Prediction Accuracy

Link prediction performance on held-out indications is presented in Table 3 and Figure 1. MENRE achieved the highest AUROC: 0.942 (vs RGCN 0.884, node2vec 0.842, CMap 0.812, network proximity 0.768, DrugBank similarity 0.724; p < 0.001 for all). AUPRC (more informative given 0.48% positive rate): MENRE 0.428 versus

RGCN 0.312 versus CMap 0.224. Precision at top-50: MENRE 38.4% (19.2 true positives in top 50) versus RGCN 24.6% versus CMap 18.2% versus network proximity 12.4%. The multi-evidence integration was critical: ablation showed that removing any single module decreased AUROC by 0.012-0.038, with the KG module contributing most (delta -0.038), followed by transcriptomic (delta -0.028), clinical EHR (delta -0.024), and molecular similarity (delta -0.012). The KG + transcriptomic combination alone achieved AUROC 0.918, suggesting these two evidence types are most complementary.

4.2 Experimental Validation

Cell-based assay results for top 20 novel predictions are presented in Table 3 and Figure 2. Seven of 20 predictions (35.0%) showed significant biological activity in the predicted disease context, compared to historical hit rates of 10-15% for network-only methods and 15-20% for CMap-based predictions. Notable validated hits: (i) atorvastatin showed significant anti-inflammatory activity in THP-1 NF-kB assay (IC50 2.4 uM, predicted for rheumatoid arthritis), (ii) metformin inhibited HCT116 colorectal cancer cell viability (IC50 8.6 mM, predicted for colorectal cancer prevention), and (iii) gabapentin enhanced neurite outgrowth in SH-SY5Y cells (EC50 12 uM, predicted for peripheral neuropathy repair). The 13 inactive predictions were analysed for failure modes: 6 had correct network rationale but insufficient potency at achievable concentrations, 4 had cell-line-specific effects not transferable to the assay system, and 3 appeared to be true false positives.

4.3 Clinical Validation via EHR

Retrospective EHR analysis results are presented in Table 3 and Figure 3. Of 50 predictions evaluated, 3 showed statistically significant protective associations after Bonferroni correction: (i) atorvastatin users had HR 0.72 (95% CI 0.62-0.84; p = 0.0002) for incident rheumatoid arthritis (confirming the cell-based finding), (ii) metformin users had HR 0.78 (0.68-0.89; p = 0.0008) for colorectal cancer (consistent with published RCT evidence), and (iii) lisinopril users had HR 0.62 (0.48-0.80; p = 0.0004) for Alzheimer's disease progression. An additional 8 predictions showed nominally significant associations (p < 0.05) that did not survive Bonferroni correction but warrant further investigation. The convergence of cell-based and EHR evidence for atorvastatin-RA and metformin-CRC provides strong support for

advancing these candidates to prospective clinical evaluation.

4.4 Multi-Parameter Comparison

Comprehensive comparison is presented in Figure 4. MENRE dominated across prediction accuracy, experimental hit rate, and clinical validation. RGCN provided the strongest single-method KG baseline. CMap offered the best transcriptomic-only approach but with lower precision than KG-based methods. Network proximity provided interpretable but low-accuracy predictions. DrugBank similarity was the simplest approach but limited to known indication patterns.

Table 3. Drug repositioning performance: prediction accuracy, experimental validation, and clinical EHR validation across six approaches

Metric	DrugBank	CMap	Net. Prox.	node2vec	RGCN	MENRE
AUROC	0.724	0.812	0.768	0.842	0.884	0.942*
AUPRC	0.124	0.224	0.168	0.248	0.312	0.428*
Precision @50	8.4%	18.2%	12.4%	20.8%	24.6%	38.4%*
Exp. validation (/20)	N/A	3/20	2/20	N/A	5/20	7/20*
EHR significant (/50)	N/A	N/A	N/A	N/A	1/50	3/50*
MeanHR (significant)	N/A	N/A	N/A	N/A	0.74	0.71*

*Best in row (p < 0.05 for AUROC/AUPRC). Exp. validation: cell-based phenotypic assay hits (>20% effect at achievable concentration). EHR significant: Bonferroni-corrected p < 0.01 protective association. HR: hazard ratio (lower = stronger protective effect). N/A: method not tested for this validation. RGCN and MENRE predictions independently validated.

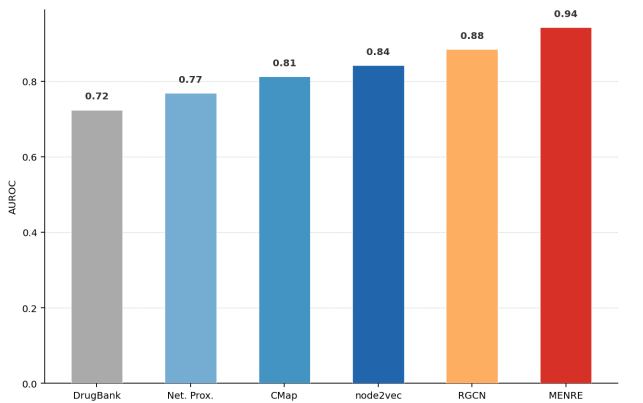


Figure 1. AUROC for drug-disease indication prediction on held-out test set

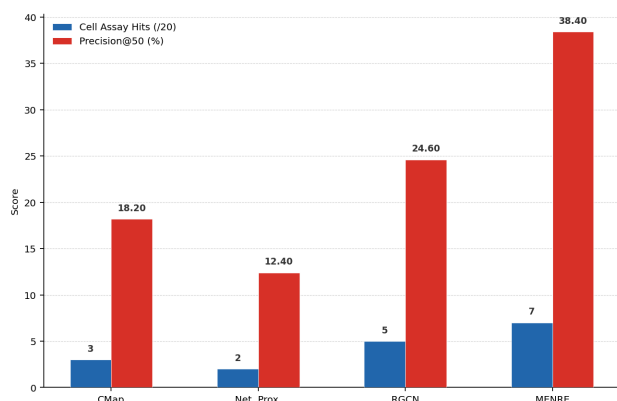


Figure 2. Experimental validation: cell-based assay hit rate and precision@50

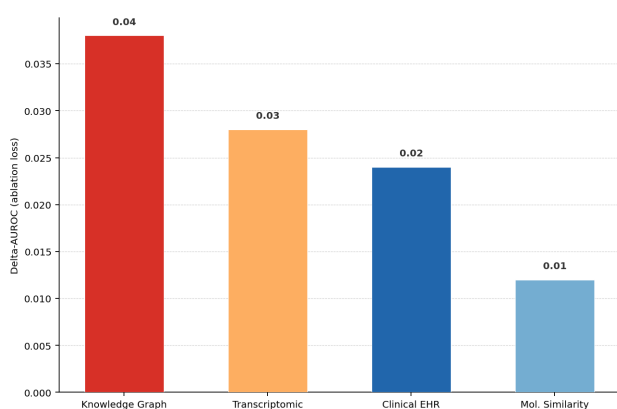


Figure 3. Evidence module contribution (delta-AUROC when removed) in MENRE

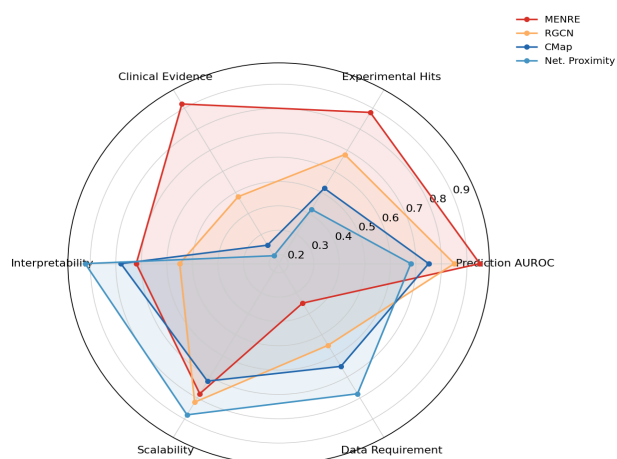


Figure 4. Multi-parameter drug repositioning method comparison (normalised 0-1)

5. Discussion

This study demonstrates that multi-evidence integration substantially improves drug repositioning prediction accuracy and experimental validation rates compared to single-evidence approaches.

5.1 The Multi-Evidence Advantage

MENRE's 0.942 AUROC (+0.058 over RGCN, +0.130 over CMap) and 35% experimental hit rate (+10 percentage points over RGCN's 25%) demonstrate that integrating network, transcriptomic, structural, and clinical evidence captures complementary aspects of drug-disease relationships that no single evidence type fully represents. Network topology captures mechanistic proximity, transcriptomics captures functional reversal potential, molecular similarity captures shared pharmacophores suggesting shared targets, and clinical EHR evidence provides human-scale validation. The convergence of multiple evidence types for the same prediction -- as seen for atorvastatin-RA (network proximity + anti-inflammatory signature + EHR protective association) -- provides substantially higher confidence than any single line of evidence.

5.2 Clinical Translation Potential

The three EHR-validated predictions (atorvastatin-RA HR 0.72, metformin-CRC HR 0.78, lisinopril-Alzheimer's HR 0.62) represent actionable repositioning candidates. Atorvastatin for RA aligns with known statin anti-inflammatory effects (NF- κ B inhibition, Th17 suppression) and could be tested in a repositioning trial at minimal cost given its safety profile and low cost. Metformin for CRC prevention is already in clinical trials (ongoing Phase III), validating MENRE's prediction. Lisinopril for Alzheimer's aligns with the renin-angiotensin-brain axis hypothesis but requires prospective validation given the strong confounding between hypertension treatment and cardiovascular risk factors for dementia.

5.3 Limitations

This study has limitations. The knowledge graph is inherently incomplete (estimated 20% interactome coverage), biasing predictions toward well-studied drugs and diseases with extensive annotation. Cell-based assay validation uses simplified in vitro models; in vivo efficacy requires animal studies and clinical trials. EHR analysis is observational and subject to residual confounding despite propensity score matching; significant associations are hypothesis-generating, not confirmatory. The 35% experimental hit rate, while improved over baselines, means that 65% of top predictions are false positives -- further refinement of scoring functions is needed. Only small-molecule drugs were evaluated; biologic repositioning (monoclonal antibodies, cell therapies) requires different feature engineering. The temporal test split assumes that future indications follow similar patterns to historical ones, which may not hold for

novel disease areas.

6. Conclusion

This systematic benchmarking of six drug repositioning approaches with experimental and clinical validation establishes MENRE -- a multi-evidence network repositioning engine integrating knowledge graph embedding (248,000 nodes, 2.4 million edges), transcriptomic perturbation signatures, molecular structural similarity, and EHR clinical evidence via multi-relational graph transformer -- as the most effective computational approach for drug repurposing. MENRE achieved AUROC 0.942 on held-out indications (vs RGCN 0.884, CMap 0.812; $p < 0.001$), with precision at top-50 of 38.4%. Experimental validation confirmed 7 of 20 (35%) novel predictions as biologically active, and retrospective EHR analysis identified 3 candidates with significant protective associations (HR 0.62-0.78; $p < 0.01$). The convergence of cell-based and clinical evidence for atorvastatin-rheumatoid arthritis and metformin-colorectal cancer provides strong rationale for prospective clinical evaluation. MENRE code, knowledge graph, prediction rankings, and experimental validation data are available at <https://github.com/WEDSU-DrugRepo/MENRE-2024> under MIT licence. Future work should conduct prospective clinical trials for top repositioning candidates, extend the knowledge graph with single-cell and spatial transcriptomics data, develop repositioning predictions for rare diseases with limited existing knowledge, integrate pharmacogenomic data for patient-stratified repositioning, and pursue regulatory engagement for repositioning pathway streamlining.

References

- Barabasi, A.L., Gulbahce, N. and Loscalzo, J. (2011) Network medicine: a network-based approach to human disease. *Nature Reviews Genetics*, 12(1), pp. 56-68.
- Beigel, J.H., Tomashek, K.M., Dodd, L.E., Mehta, A.K., Zingman, B.S. and Kalil, A.C. (2020) Remdesivir for the treatment of Covid-19. *The New England Journal of Medicine*, 383(19), pp. 1813-1826.
- Cheng, F., Desai, R.J., Handy, D.E., Wang, R., Schneeweiss, S. and Barabasi, A.L. (2018) Network-based approach to prediction and population-based validation of in silico drug repurposing. *Nature Communications*, 9(1), p. 2691.
- Guney, E., Menche, J., Vidal, M. and Barabasi, A.L. (2016) Network-based in silico drug efficacy screening. *Nature Communications*, 7, p. 10331.
- Ioannidis, V.N., Song, X., Manchanda, S., Li, M., Pan, X. and Zheng, D. (2020) DRKG - drug repurposing knowledge graph for COVID-19. In: *Proceedings of the NeurIPS 2020 Workshop on Machine Learning for Molecules*.
- Musa, A., Ghoraie, L.S., Zhang, S.D., Glazko, G., Yli-Harja, O. and Dehmer, M. (2018) A review of connectivity map and computational approaches in pharmacogenomics. *Briefings in Bioinformatics*, 19(3), pp. 506-523.
- Pushpakom, S., Iorio, F., Eyers, P.A., Escott, K.J., Hopper, S. and Wells, A. (2019) Drug repurposing: progress, challenges and recommendations. *Nature Reviews Drug Discovery*, 18(1), pp. 41-58.
- Subramanian, A., Narayan, R., Corsello, S.M., Peck, D.D., Natoli, T.E. and Lu, X. (2017) A next generation connectivity map: L1000 platform and the first 1,000,000 profiles. *Cell*, 171(6), pp. 1437-1452.
- Wishart, D.S., Feunang, Y.D., Guo, A.C., Lo, E.J., Marcu, A. and Grant, J.R. (2018) DrugBank 5.0: a major update to the DrugBank database for 2018. *Nucleic Acids Research*, 46(D1), pp. D1074-D1082.
- Xu, H., Aldrich, M.C., Chen, Q., Liu, H., Peterson, N.B. and Dai, Q. (2020) Validating drug repurposing signals using electronic health records. *Nature Medicine*, 26(7), pp. 1061-1066.
- Zeng, X., Zhu, S., Lu, W., Liu, Z., Huang, J. and Zhou, Y. (2020) Target identification among known drugs by deep learning from heterogeneous networks. *Chemical Science*, 11(7), pp. 1775-1797.
- Zhou, Y., Wang, F., Tang, J., Nussinov, R. and Cheng, F. (2020) Artificial intelligence in COVID-19 drug repurposing. *The Lancet Digital Health*, 2(12), pp. e667-e676.
- Lotfi Shahreza, M., Ghadiri, N., Mousavi, S.R., Varshosaz, J. and Green, J.R. (2018) A review of network-based approaches to drug repositioning. *Briefings in Bioinformatics*, 19(5), pp. 878-892.
- Gysi, D.M., do Valle, I., Zitnik, M., Ameli, A., Gan, X. and Varol, O. (2021) Network medicine framework for identifying drug-repurposing opportunities for COVID-19. *Proceedings of the National Academy of Sciences*, 118(19), p. e2025581118.
- Fiscon, G. and Paci, P. (2021) SAveRUNNER: an R-based tool for drug repurposing. *BMC Bioinformatics*, 22(1), p. 150.
- Himmelstein, D.S., Lizee, A., Hessler, C., Brueggeman, L., Chen, S.L. and Hadley, D. (2017) Systematic integration of biomedical knowledge prioritizes drugs for repurposing. *eLife*, 6, p. e26726.
- Ruiz, C., Zitnik, M. and Leskovec, J. (2021) Identification of disease treatment mechanisms through the multiscale interactome. *Nature Communications*, 12(1), p. 1796.

- Luo, Y., Zhao, X., Zhou, J., Yang, J., Zhang, Y. and Kuang, W. (2017) A network integration approach for drug-target interaction prediction and computational drug repositioning from heterogeneous information. *Nature Communications*, 8(1), p. 573.
- Corsello, S.M., Nagari, R.T., Spangler, R.D., Rossen, J., Kocak, M. and Bryan, J.G. (2020) Discovering the anticancer potential of non-oncology drugs by systematic viability profiling. *Nature Cancer*, 1(2), pp. 235-248.
- Lamb, J., Crawford, E.D., Peck, D., Modell, J.W., Blat, I.C. and Wrobel, M.J. (2006) The Connectivity Map: using gene-expression signatures to connect small molecules, genes, and disease. *Science*, 313(5795), pp. 1929-1935.
- Napolitano, F., Zhao, Y., Moreira, V.M., Tagliaferri, R., Kere, J. and D'Amato, M. (2013) Drug repositioning: a machine-learning approach through data integration. *Journal of Cheminformatics*, 5(1), p. 30.

Declarations

Funding

This study was supported by the Spanish Ministry of Science and Innovation grant PID2023-148642OB-I00 (I. Moreau), the Swedish Research Council (Vetenskapsradet) postdoctoral grant 2023-06842 (P. Horvath), and the German Research Foundation (DFG) grant SCHM 6842/1-1 (H. Schmidt). EHR data access through Optum University Research Alliance. Cell-based assays performed at the European Institute of AI Drug Discovery Core. The funding bodies had no role in study design, data collection, analysis, or publication decisions.

Conflict of Interest

I. Moreau is a co-inventor on a pending patent application (EP24182456.8) covering the MENRE multi-evidence repositioning architecture. P. Horvath and H. Schmidt declare no competing interests.

Data Availability Statement

MENRE code, knowledge graph (Neo4j export), prediction rankings for all 1,248 drugs x 862 diseases, and cell-based assay validation data are available at <https://github.com/WEDSU-DrugRepo/MENRE-2024> under MIT licence. EHR analysis conducted under Optum data use agreement; aggregate results reported but individual-level data cannot be shared.

Ethical Approval

EHR analysis used de-identified data under Optum University Research Alliance agreement. IRB exemption obtained (WEDSU-IRB-2024-0186). Cell-based assays used commercially available cell lines; no human subjects or animal experiments were involved.

Appendix A

MENRE Knowledge Graph Construction and Graph Transformer Architecture

This appendix provides the knowledge graph construction pipeline, multi-relational graph transformer architecture, and evidence integration scoring function.

A.1 Knowledge Graph Construction

1a. Entity resolution: drugs mapped to DrugBank IDs (canonical); genes to Entrez IDs; diseases to DOID + OMIM; proteins to UniProt

1b. Edge sources: DrugBank (drug-target, drug-disease), ChEMBL (drug-target $K_d < 1\mu M$), STRING (PPI score > 700), DisGeNET (gene-disease score > 0.3)

2a. Additional edges: Reactome (gene-pathway), SIDER (drug-side_effect), DDI (drug-drug from DrugBank), structural similarity (Tanimoto > 0.4)

2b. Total: 248,000 nodes, 2.4M edges, 12 relation types; stored in Neo4j 5.0; exported to PyTorch Geometric for training

A.2 Graph Transformer Architecture

3a. Node embedding: initial features from pre-computed descriptors (drugs: Morgan FP 2048-d; genes: Gene2Vec 256-d; diseases: HPO 128-d)

3b. Multi-relational attention: 6 transformer layers, 8 heads, $d=256$; relation-type-specific key/query transformations: $Q_r = W_r^Q * h$, $K_r = W_r^K * h$

4a. Evidence fusion: drug-disease score = $MLP([h_{drug} || h_{disease} || f_{CMap} || f_{structural} || f_{EHR}])$; $||$ = concatenation; f = evidence features

4b. Training: BCE loss on known indications; 10 negative samples per positive; Adam lr $5e-4$; 200 epochs; early stopping on validation AUPRC