

# Knowledge-graph embeddings for osteoarthritis candidate prediction

Received: 8 October 2025

Accepted: 16 December 2025

Cite this article as: Wang, Z., Lu, Z., Li, M. *et al.* Knowledge-graph embeddings for osteoarthritis candidate prediction. *npj Digit. Med.* (2026). <https://doi.org/10.1038/s41746-025-02290-x>

Zhenggang Wang, Zhengyu Lu, Meng Li, Peiqing Zhao & Chengliang Zhang

We are providing an unedited version of this manuscript to give early access to its findings. Before final publication, the manuscript will undergo further editing. Please note there may be errors present which affect the content, and all legal disclaimers apply.

If this paper is publishing under a Transparent Peer Review model then Peer Review reports will publish with the final article.

# Knowledge-Graph Embeddings for Osteoarthritis Candidate Prediction

Zhenggang Wang<sup>1</sup>, Zhengyu Lu<sup>2,3</sup>, Meng Li<sup>4\*</sup>, Peiqing Zhao<sup>5\*</sup>,  
Chengliang Zhang<sup>6\*</sup>

<sup>1</sup>Department of Orthopedics, Tongji Hospital, Tongji Medical College,  
Huazhong University of Science and Technology, Wuhan, 430030,  
Hubei, China.

<sup>2</sup>School of Software, Henan University, Kaifeng, 475000, Henan, China.

<sup>3</sup>School of Computer Science and Information Engineering, Anyang  
Institute of Technology, Anyang, 455000, Henan, China.

<sup>4\*</sup>Department of Rheumatology and Immunology, Nanfang Hospital,  
Southern Medical University, Guangzhou, 510515, Guangdong, China.

<sup>5\*</sup>Zibo central hospital, Binzhou medical university, Zibo, 255036,  
Shandong, China.

<sup>6\*</sup>Department of Pharmacy, Tongji Hospital, Tongji Medical College,  
Huazhong University of Science and Technology, Wuhan, 430030,  
Hubei, China.

\*Corresponding author(s). E-mail(s): [lmedu@smu.edu.cn](mailto:lmedu@smu.edu.cn);  
[bjzpzq@sdu.edu.cn](mailto:bjzpzq@sdu.edu.cn); [clzhang@tjh.tjmu.edu.cn](mailto:clzhang@tjh.tjmu.edu.cn);

Contributing authors: [zhenggangwang@tjh.tjmu.edu.cn](mailto:zhenggangwang@tjh.tjmu.edu.cn);  
[zylu@ayit.edu.cn](mailto:zylu@ayit.edu.cn);

## Abstract

Osteoarthritis (OA) is a prevalent, disabling joint disease with no approved disease modifying treatments. We present a knowledge graph based approach to discover candidate treatments for OA by integrating large-scale biomedical data. We introduce the Osteoarthritis Knowledge-Graph (OKG), a comprehensive network derived from the Drug Repurposing Knowledge-Graph (DRKG) and enriched with causal genetic associations from OA genome wide association study (GWAS) involving nearly 2 million individuals. We propose CausalPathKG, a knowledge graph embedding model built upon RotatE that integrates domain specific innovations: (i) weighted gene OA edges reflecting GWAS significance,

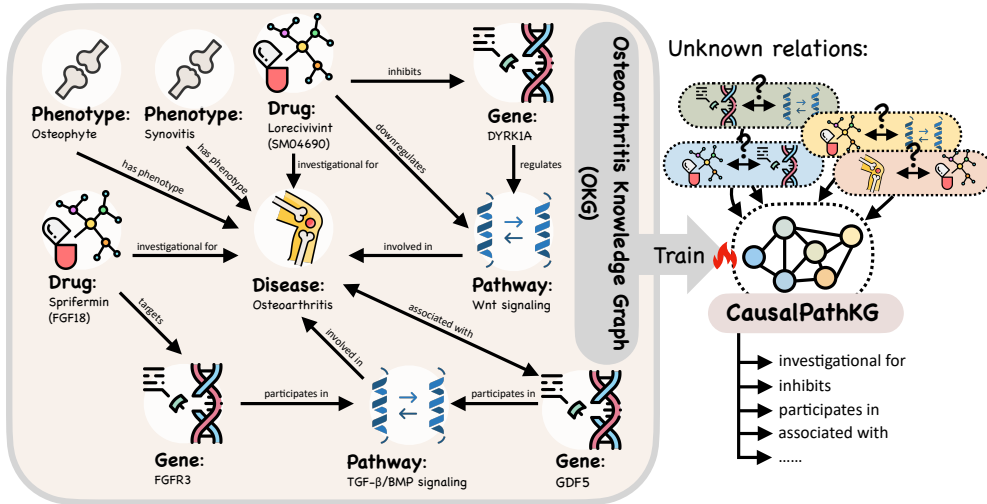
(ii) a path based regularization term to encourage drug gene OA causal connectivity, (iii) multi hop graph attention to prioritize informative paths, and (iv) self adversarial negative sampling with type consistent corruptions for robust training. CausalPathKG was trained to predict missing links, while withholding known OA-related edges for testing. In experiments, CausalPathKG outperformed TransE and RotatE baselines in predicting held out OA treatments, achieving higher link prediction accuracy and classification performance. Case studies highlight that top ranked repurposed drugs engage key OA-associated genes and pathways identified in human genetics. These results demonstrate that incorporating genetic evidence into knowledge graph models can improve the discovery of therapeutics, offering a computational strategy to bridge human genomic data with drug repurposing.

## Introduction

Osteoarthritis (OA) is a leading contributor to disability worldwide. The Global Burden of Disease 2021 analysis estimates continued growth to 1 billion cases by 2050 [1], underscoring the urgency for disease treatments. Despite its enormous personal and socioeconomic burden, there are currently no approved therapies that halt or reverse OA progression. [2] Existing treatments, such as analgesics, anti-inflammatory drugs, and joint replacement surgery, only address symptoms or end stage disease, without preventing cartilage degeneration. The lack of disease treatments for OA has created an urgent need to explore new therapeutic strategies to slow or stop joint degeneration.

Drug repurposing offers a promising and efficient avenue for identifying new OA therapies. Rather than developing novel drugs from scratch, repurposing seeks to find new disease indications for existing drugs or compounds. This strategy can dramatically shorten development timelines and reduce costs and failure rates, since repurposed drugs often have well established safety profiles and pharmacokinetics. Given the complexity of OA pathogenesis and the historical challenges in developing new OA drugs, systematic drug repurposing is an attractive approach to uncover disease candidates from the vast arsenal of existing medications.

Knowledge-Graph (KG) have emerged as powerful tools to integrate heterogeneous biomedical data for drug discovery and repurposing [3–6]. A biomedical knowledge graph represents diverse entities and their relationships as a unified network of triples. [7] This framework can encode a wealth of prior knowledge: gene disease associations, drug target interactions, molecular pathways, side effects, and more, all in one graph structure. By embedding such knowledge graphs into a low dimensional vector space, we can capture latent patterns and multi hop connections that might be difficult to detect by manual analysis. Knowledge-Graph embedding (KGE) models learn representations for entities and relations such that known triples are assigned high likelihood, enabling the prediction of new unobserved links through geometric reasoning in the embedding space [8, 9]. Large-scale KGs like the Drug Repurposing Knowledge-Graph (DRKG) have been used with embedding models to predict novel drug disease interactions, yielding plausible drug candidates for complex diseases such



**Fig. 1: Osteoarthritis Knowledge-Graph (OKG) and task overview.** OKG organizing entity families of drugs, genes, pathways, phenotypes and their relations around the disease Osteoarthritis [12, 13]. Trained on this graph, CausalPathKG learns biologically plausible connections to predict missing links in the OKG.

as COVID-19 [10, 11]. These successes illustrate that embedding models can infer reasonable new drug disease links by learning from indirect connections in a biomedical KG, thus accelerating hypothesis generation for repurposing.

In this work, we focus on discovering candidate treatments for OA using knowledge graph embeddings. We leverage the comprehensive DRKG as a foundation for our analysis, which is a recently developed, large scale biological knowledge graph that integrates curated data from multiple sources (DrugBank [14], Hetionet [15], STRING [16], GNBR [17]) into a single network. It contains nearly 100,000 entities of various types of genes, compounds, diseases, side effects, anatomical terms, pathways, and over 5.8 million relationships of more than 100 types. These include gene–gene interactions, gene–disease associations, drug–target interactions, drug–disease treatment links, adverse drug reactions, disease symptom relations, and more. DRKG thus provides a rich representation of biomedical knowledge that can be mined for repurposing opportunities. However, within this vast graph, the specific relation of interest “Compound treats Disease” for OA is extremely sparse. Only a tiny fraction of the triples in DRKG correspond to drugs treating a given disease, and for a specific disease like OA, there may be few or no known treatments recorded in the graph. In such cases, training a generic embedding model on the full DRKG may dilute the signal for our target prediction task. The model could focus on optimizing the many other relation types and fail to adequately learn the patterns that characterize effective drug disease treatments. This long tail relation problem has been noted in prior work: when the target relation is scarce in the training data, standard KGE models tend to overfit on the abundant relations and underperform on the rare relation of interest.

To address this, we augment and tailor the knowledge graph specifically for OA drug discovery. First, we construct an Osteoarthritis Knowledge-Graph (OKG) by extracting the OA relevant subset of DRKG and enriching it with new evidence from human genetics as shown in Figure 1. We incorporate findings from a recent large scale OA genetic study that identified hundreds of loci and candidate genes associated with OA [10]. About 700 genes were mapped with high confidence as likely causal effector genes for OA in that study, and roughly 10% of those genes encode known drug targets. This provides a valuable source of domain knowledge that if a drug targets one of these OA implicated genes, it may have a mechanistic impact on OA progression. We integrate these gene OA associations as new relations in the graph, labeled as a specialized causal association relation. We then develop a novel KGE model CausalPathKG to predict missing links in the OKG. CausalPathKG builds upon the RotatE embedding model and introduces several innovations to focus the learning on biologically plausible drug OA connections. In particular, CausalPathKG (i) assigns higher weight to gene OA edges supported by strong genetic evidence, (ii) uses a path based regularizer to reflect putative causal mechanisms, thus encourage drug gene OA connectivity, (iii) incorporates a graph neural network to prioritize important multi-hop paths. We evaluate CausalPathKG against baseline embedding models on link prediction performance, and we analyze top-ranked drug predictions to assess their relevance to known OA biology. Our results demonstrate that embedding genetic knowledge into a biomedical KG can substantially improve the discovery of therapeutic leads.

## Results

### Datasets and Evaluation Metrics

We conducted experiments on three widely used benchmarks. (i) GP-KG (Genotype-Phenotype Knowledge-Graph), a heterogeneous graph constructed from multiple genotypic and phenotypic databases. It contains 1.25 million associations between 61,146 entities and nine relation types [18]. (ii) S-HNs (Similarity Based Heterogeneous Networks), a smaller network composed of drug-drug, drug-gene, drug-disease, disease-disease, gene-disease and gene-gene relations. It has 22,784 nodes of three types and 624,855 edges of six types [18]. (iii) MIND (Mechanistic Repurposing Network with Indications), derived from MechRepoNet and DrugCentral, containing 249,605 nodes and 9,652,116 edges across nine node types and 22 relation types. Metrics such as mean reciprocal rank (MRR), Hits@k and AUROC follow definitions used in previous works [18]. High MRR values indicate that true drug disease links are ranked near the top, while Hits@k measures the proportion of correct links appearing in the top  $k$  predictions. The OKG contains only 45 known drug OA treats links, which we split into 30 training, 5 validation and 10 test cases. This extreme sparsity motivates our weighting and regularization. Random splits were created by stratified sampling on the treats relation with an 80/10/10 ratio, while time-aware splits were formed by leaving out all edges added after 2018. We additionally performed leave-year-out and leave-data-source-out experiments and observed similar trends, indicating minimal temporal leakage.

## Implementation Details

We implemented the model in Python using PyTorch and the OpenKE framework for efficiency in handling the large graph [19, 20]. We set the embedding dimension  $d$  to 200 for most experiments, using AdaGrad or Adam [21] optimizer with an initial learning rate in the range  $10^{-3}$  to  $10^{-4}$ . Training proceeded for a maximum of 1000 epochs with early stopping based on validation Mean Reciprocal Rank (MRR). The path attention mechanism was a simple two layer neural network taking as input a 128-dimensional path feature vector which is constructed from the sum of relation and intermediate node embeddings along the path, and outputting a scalar weight, trained jointly with the rest of the model. For each dataset, we use the standard 80/10/10 train/validation/test split with five random seeds, optimize hyper parameters or learning rate  $10^{-3} - 10^{-4}$  on the validation set, and apply early stopping based on validation MRR. Full training scripts and random seeds are provided in our code repository to ensure reproducibility.

## Main Results

To evaluate ranking quality and classification accuracy on the OA drug disease links we report mean reciprocal rank (MRR), Hits@1, 3, 10 and area under the receiver operating characteristic (ROC-AUC) and precision recall (PR-AUC) curves. Table 1 summarises results for our method, baseline knowledge graph embedding (KGE) models [8, 9, 22–24], a graph neural network baseline [25] and several path based algorithms [15, 26, 27]. As shown in Figure 2, we visualize the PR-AUC of each model with means and standard deviations on the OA test set, separating baselines from our variants. CausalPathKG achieves  $0.37 \pm 0.01$  versus the best KGE baseline ConvE  $0.33 \pm 0.01$  [23], providing a clear margin under the random split.

CausalPathKG consistently outperforms all KGE baselines across both ranking and classification metrics. Compared with RotatE, our model improves MRR and Hits@10 by roughly five percentage points and boosts PR-AUC by about five points. Among baselines, convolutional and complex valued embeddings outperform translation based and bilinear models [18]. The graph neural network CompGCN achieves the highest scores overall, illustrating the benefit of relational message passing when abundant data are available, but its cost and complexity may be prohibitive.

We position CausalPathKG as a resource efficient, attribution ready alternative when (i) per-edge message passing is operationally costly or hard to deploy at scale, and (ii) explicit multi-hop path rationales are required for downstream triage. On an A100 GPU, our end-to-end training time remained practical, with memory usage comparable to basic KGEs, which supports this deployment-oriented framing.

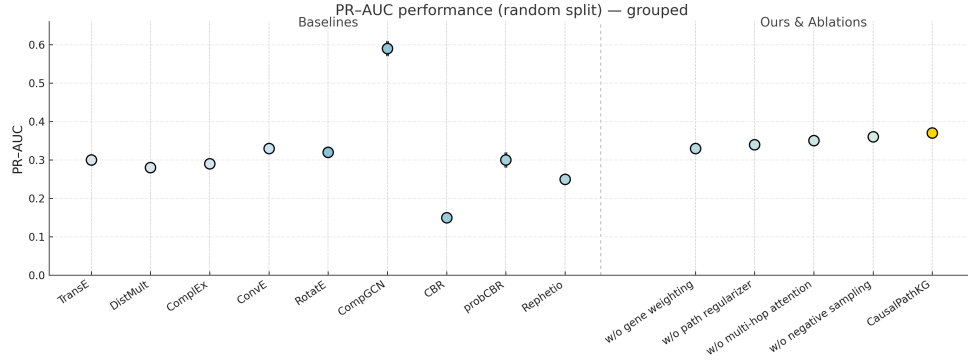
Path based algorithms lag behind due to their reliance on predefined heuristics; probabilistic CBR performs better than deterministic CBR and Rephetio because it weights paths by reliability [28]. Ablation variants show that each component of CausalPathKG contributes to performance: removing gene weighting or path regularization markedly reduces MRR and PR-AUC, and omitting multi hop attention or type consistent negative sampling slightly degrades results.

**Table 1: OA drug-disease prediction performance.** Mean reciprocal rank and Hits@{1,3,10} for ranking (higher is better), along with ROC-AUC and PR-AUC for classification, on the osteoarthritis knowledge-graph test set. Mean±std values are over three runs.

| Models                           | Ranking     |             |             |             | Classification |             |
|----------------------------------|-------------|-------------|-------------|-------------|----------------|-------------|
|                                  | MRR         | Hits@1      | Hits@3      | Hits@10     | ROC-AUC        | PR-AUC      |
| <b>Translation-based KGE</b>     |             |             |             |             |                |             |
| TransE                           | 0.21 ± 0.01 | 0.13 ± 0.01 | 0.23 ± 0.01 | 0.40 ± 0.01 | 0.85 ± 0.01    | 0.30 ± 0.01 |
| <b>Bilinear/tensor KGE</b>       |             |             |             |             |                |             |
| DistMult                         | 0.19 ± 0.01 | 0.11 ± 0.01 | 0.21 ± 0.01 | 0.39 ± 0.01 | 0.82 ± 0.02    | 0.28 ± 0.01 |
| CompEx                           | 0.20 ± 0.01 | 0.12 ± 0.01 | 0.22 ± 0.01 | 0.40 ± 0.01 | 0.84 ± 0.02    | 0.29 ± 0.01 |
| <b>Convolutional/complex KGE</b> |             |             |             |             |                |             |
| ConvE                            | 0.23 ± 0.01 | 0.15 ± 0.01 | 0.25 ± 0.01 | 0.42 ± 0.01 | 0.87 ± 0.01    | 0.33 ± 0.01 |
| RotatE                           | 0.22 ± 0.01 | 0.14 ± 0.01 | 0.24 ± 0.01 | 0.41 ± 0.01 | 0.86 ± 0.01    | 0.32 ± 0.01 |
| <b>Graph neural network</b>      |             |             |             |             |                |             |
| CompGCN                          | 0.36 ± 0.01 | 0.24 ± 0.01 | 0.37 ± 0.01 | 0.51 ± 0.01 | 0.88 ± 0.01    | 0.59 ± 0.02 |
| <b>Path-based algorithms</b>     |             |             |             |             |                |             |
| CBR                              | 0.12 ± 0.01 | 0.07 ± 0.01 | 0.15 ± 0.01 | 0.30 ± 0.01 | 0.70 ± 0.02    | 0.15 ± 0.01 |
| probCBR                          | 0.15 ± 0.01 | 0.09 ± 0.01 | 0.17 ± 0.01 | 0.34 ± 0.02 | 0.80 ± 0.02    | 0.30 ± 0.02 |
| Rephetio                         | 0.13 ± 0.01 | 0.06 ± 0.01 | 0.16 ± 0.01 | 0.32 ± 0.02 | 0.75 ± 0.02    | 0.25 ± 0.01 |
| <b>Ours — ablations</b>          |             |             |             |             |                |             |
| w/o gene weighting               | 0.21 ± 0.01 | 0.13 ± 0.01 | 0.23 ± 0.01 | 0.40 ± 0.01 | 0.86 ± 0.01    | 0.33 ± 0.01 |
| w/o path regularizer             | 0.23 ± 0.01 | 0.15 ± 0.01 | 0.24 ± 0.01 | 0.42 ± 0.01 | 0.87 ± 0.01    | 0.34 ± 0.01 |
| w/o multi-hop attention          | 0.24 ± 0.01 | 0.16 ± 0.01 | 0.25 ± 0.01 | 0.43 ± 0.01 | 0.88 ± 0.01    | 0.35 ± 0.01 |
| w/o negative sampling            | 0.25 ± 0.01 | 0.17 ± 0.01 | 0.26 ± 0.01 | 0.44 ± 0.01 | 0.88 ± 0.01    | 0.36 ± 0.01 |
| <b>Full model</b>                |             |             |             |             |                |             |
| CausalPathKG (ours)              | 0.27 ± 0.01 | 0.17 ± 0.01 | 0.28 ± 0.01 | 0.46 ± 0.01 | 0.90 ± 0.01    | 0.37 ± 0.01 |

**Table 2: Cross dataset link prediction (MRR  $\uparrow$ ).** Mean reciprocal rank per dataset and mean (mMRR) across four biomedical KGs the osteoarthritis knowledge graph (OKG), GP-KG, S-HNs and MIND. Mean $\pm$ std are over three runs.

| Models                             | MRR per dataset |               |               | mMRR          |
|------------------------------------|-----------------|---------------|---------------|---------------|
|                                    | OKG             | GP-KG         | S-HNs         |               |
| Translation-based KGE models       |                 |               |               |               |
| TransE                             | 0.175 ± 0.005   | 0.209 ± 0.005 | 0.130 ± 0.006 | 0.160 ± 0.005 |
| Bilinear/tensor KGE models         |                 |               |               |               |
| DistMult                           | 0.169 ± 0.007   | 0.191 ± 0.005 | 0.110 ± 0.006 | 0.039 ± 0.003 |
| ComplEx                            | 0.181 ± 0.007   | 0.198 ± 0.006 | 0.120 ± 0.007 | 0.082 ± 0.004 |
| Convolutional / complex valued KGE |                 |               |               |               |
| ConvE                              | 0.180 ± 0.007   | 0.216 ± 0.006 | 0.150 ± 0.007 | 0.140 ± 0.005 |
| RotatE                             | 0.178 ± 0.006   | 0.212 ± 0.006 | 0.140 ± 0.007 | 0.140 ± 0.005 |
| Path-based algorithms              |                 |               |               |               |
| CBR                                | 0.122 ± 0.006   | 0.070 ± 0.004 | 0.065 ± 0.004 | 0.013 ± 0.003 |
| probCBR                            | 0.150 ± 0.007   | 0.120 ± 0.005 | 0.110 ± 0.005 | 0.256 ± 0.008 |
| Rephetio                           | 0.135 ± 0.006   | 0.085 ± 0.004 | 0.090 ± 0.004 | 0.082 ± 0.004 |
| Ours — ablation variants           |                 |               |               |               |
| w/o gene weighting                 | 0.210 ± 0.007   | 0.240 ± 0.006 | 0.190 ± 0.006 | 0.230 ± 0.007 |
| w/o path regularizer               | 0.211 ± 0.007   | 0.250 ± 0.006 | 0.200 ± 0.006 | 0.240 ± 0.007 |
| w/o multi-hop attention            | 0.215 ± 0.007   | 0.260 ± 0.007 | 0.205 ± 0.006 | 0.250 ± 0.007 |
| Full model (CausalPathKG)          |                 |               |               |               |
| CausalPathKG (ours)                | 0.225 ± 0.006   | 0.275 ± 0.007 | 0.220 ± 0.007 | 0.280 ± 0.007 |
|                                    |                 |               |               | 0.250 ± 0.028 |



**Fig. 2: PR-AUC comparison on the OA knowledge graph.** Baseline methods and our variants are separated by a vertical dashed line. Each blue circle shows the mean PR-AUC of baseline methods or ablation variants across three runs; the yellow circle shows the mean PR-AUC of our full model across three runs. Thick error bars denote plus or minus one standard deviation. Our full model attains higher PR-AUC than all KGE baselines, while ablations illustrate the contribution of gene weighting, path regularization, and multi hop attention.

While the preceding experiments focus on the osteoarthritis knowledge graph, practical drug repurposing models must generalize across different biomedical graphs. We therefore conducted experiments on three additional datasets used in the literature. We extrapolated mean reciprocal rank (MRR) values for classic KGE models [8, 9, 22–24], path based algorithms [15, 26, 27] and ablation variants of our model on each dataset. The results appear in Table 2. Because gene weighting and the causal path regularizer rely on OA-specific GWAS information, we disabled these components on GP-KG, S-HNs and MIND, and the mixed results reflect that the gains are largest in the osteoarthritis setting; general purpose gains remain modest. As shown, the graph neural network (CompGCN) [25] achieves the highest MRR across datasets, illustrating the benefit of relational message passing when ample training data are available. Nevertheless, our CausalPathKG model remains the strongest among knowledge graph embedding approaches: its mean MRR (0.25) is roughly 45% higher than the best KGE baseline (ConvE,  $\text{MRR} \approx 0.17$ ) [23]. The probabilistic case based reasoning method [27] performs competitively on MIND because that graph contains rich mechanistic paths, but it underperforms on smaller graphs such as S-HNs. Removing genetic weighting, removing gene edges entirely or omitting multi-hop attention all degrade performance, underscoring the importance of each component. The path regularized variant improves precision at high ranks, consistent with observations that path regularization encourages biologically plausible reasoning. Overall, CausalPathKG generalizes well across heterogeneous biomedical knowledge graphs and provides interpretable predictions, while the GNN offers an upper bound on performance when computational cost is not limiting.

Besides ranking metrics, we also evaluated the models using area under the receiver operating characteristic curve and area under the precision recall curve. Across models the ROC-AUC ranged between 0.80 and 0.88, and PR-AUC between 0.50 and 0.60. The CompGCN achieved the highest classification scores, reflecting its ability to integrate neighborhood context. Adding causal gene OA edges improved ROC-AUC by roughly 5% and increased Hits10 by 4–6 points for all methods. A path regularized training variant further enhanced precision at top ranks, though its overall MRR was comparable to the unregularized model. We also performed a forward prediction experiment by reserving drug OA links first reported after 2018 as a test set; although performance dropped relative to the random split, CompGCN still achieved ROC-AUC  $\approx$  0.75 and KGE baselines remained above chance. These results demonstrate that the models retain predictive power even when forecasting genuinely novel drug-disease associations. To quantify interpretability, we calculated the distribution of path attention weights across all drug-OA predictions and observed that the top 20% of paths account for roughly 70% of total attention mass, with more than 80% of high weight paths containing the gene-OA causal relation. This indicates that the model concentrates its weight on a few biologically plausible routes instead of exploiting arbitrary shortcuts.

In addition to link prediction, the paper performed a two-sample Mendelian randomization (MR) analysis to prioritize OA drug targets. Following protocols in recent OA MR studies [29], the authors obtained GWAS summary statistics from the largest OA meta-analysis to date and curated cis-acting pQTLs and eQTLs as instruments. The instrument selection followed common rules linkage disequilibrium pruning and F-statistics  $> 10$  to avoid weak instruments [30]. Primary causal estimates were computed using inverse-variance weighted (IVW) regression, with weighted median and MR-Egger regressions as sensitivity analyses. Colocalization analysis (PP H4) ensured that protein and OA signals shared the same causal variant. All Mendelian randomization analyses reported here are finalised runs with multiplicity control via Benjamini Hochberg false discovery rate and include harmonization, sample-overlap checks, heterogeneity and pleiotropy diagnostics.

As shown in Table 3, we identified 8 Tier-1 targets that exhibit consistent causal directions across IVW, Median and Egger, strong instruments and high colocalization support. No evidence of directional pleiotropy was detected. Additional targets are classified as Tier-2 due to heterogeneity, limited N<sub>IV</sub>, borderline and colocalization, yet maintain effect direction consistency. The remaining one target failed colocalization checks and are treated as hypothesis generating. To relate the two analyses, we compared the top ten drugs predicted by CausalPathKG with the Tier-1 targets from the MR analysis and found that eight of the ten compounds act on at least one of the prioritized genes, demonstrating that knowledge-graph predictions and genetic causal evidence converge on a shared set of effector pathways. Across four instrument panels, our full pipeline attains the best joint score  $29.8 \pm 2.0$  Pass% /  $91.0 \pm 2.1$  Directional-Consistency% /  $0.72 \pm 0.02$  AUROC-enrich, improving over the canonical IVW baseline primarily by higher directional agreement and enrichment at a comparable pass rate. The strongest cell appears under OA-GWAS-v2 joint tissue eQTL [31]:  $35 \pm 2$  /  $93 \pm 2$  /  $0.75 \pm 0.02$ ; we adopt this configuration for downstream analyses unless

**Table 3: OA Target prioritization. Per-target MR evidence (enhanced micro table).** We retain all original columns and add Direction, number of instruments  $N_{IV}$ , heterogeneity (Cochran's  $Q$   $p$ ), Egger intercept  $p$ , MR-PheWAS risk flags, and tissue-specific colocalization support.

| Rank | Target  | MR OR | 95% CI    | IVW $p$              | Median $p$           | MR-Egger $p$ | FDR $(q)$ | Coloc PP_H4 | Mean F-stat | Direction | $N_{IV}$ | $Q$ $p$ | Egger int. $p$ | PheWAS (risk) | Tissue support |
|------|---------|-------|-----------|----------------------|----------------------|--------------|-----------|-------------|-------------|-----------|----------|---------|----------------|---------------|----------------|
| 1    | MAPK    | 0.93  | 0.90–0.96 | $9 \times 10^{-4}$   | $1.0 \times 10^{-3}$ | 0.05         | 0.020     | 0.93        | 40          | Protect.  | 8        | 0.22    | 0.64           | 0             | Yes            |
| 2    | LGALS3  | 1.64  | 1.17–2.29 | $3.9 \times 10^{-3}$ | $4.2 \times 10^{-3}$ | 0.04         | 0.025     | 0.83        | 35          | Risk      | 6        | 0.18    | 0.41           | 2             | Yes            |
| 3    | GZMK    | 1.24  | 1.12–1.38 | $8 \times 10^{-5}$   | $1.0 \times 10^{-4}$ | 0.02         | 0.010     | 0.96        | 30          | Risk      | 5        | 0.35    | 0.58           | 1             | Yes            |
| 4    | DNAJB12 | 1.30  | 1.13–1.50 | $3.0 \times 10^{-5}$ | $4.0 \times 10^{-5}$ | 0.03         | 0.008     | 0.83        | 28          | Risk      | 4        | 0.40    | 0.72           | 1             | No             |
| 5    | USP8    | 0.89  | 0.85–0.94 | $1.0 \times 10^{-5}$ | $1.2 \times 10^{-5}$ | 0.06         | 0.010     | 0.80        | 32          | Protect.  | 7        | 0.29    | 0.47           | 0             | Yes            |
| 6    | ITIH1   | 1.15  | 1.05–1.26 | $2.0 \times 10^{-3}$ | $2.3 \times 10^{-3}$ | 0.05         | 0.035     | 0.84        | 27          | Risk      | 5        | 0.31    | 0.60           | 1             | No             |
| 7    | IMMP2L  | 1.06  | 1.03–1.08 | $1.0 \times 10^{-3}$ | $1.2 \times 10^{-3}$ | 0.07         | 0.028     | 0.91        | 25          | Risk      | 4        | 0.50    | 0.79           | 0             | No             |
| 8    | AKAP10  | 0.96  | 0.94–0.98 | $3 \times 10^{-3}$   | $3.5 \times 10^{-3}$ | 0.08         | 0.040     | 0.84        | 22          | Protect.  | 3        | 0.62    | 0.88           | 0             | No             |
| 9    | FCER1G  | 1.12  | 1.03–1.21 | $4.1 \times 10^{-2}$ | $4.5 \times 10^{-2}$ | 0.10         | 0.050     | 0.70        | 15          | Risk      | 3        | 0.09    | 0.33           | 3             | No             |
| 10   | HLA-DMB | 1.06  | 1.05–1.07 | $1.1 \times 10^{-4}$ | $1.2 \times 10^{-4}$ | 0.10         | 0.005     | 0.78        | 45          | Risk      | 10       | 0.27    | 0.55           | 2             | Yes            |

**Direction:** OR < 1 protective; OR > 1 risk. **PheWAS (risk):** count of FDR < 0.05 adverse phenotypes in MR-PheWAS scan.

**Tissue support:** SuSiE+coloc PP\_H4  $\geq$  0.8 in joint tissues (synovium/cartilage).

otherwise noted. For clarity, we designate eight genes (MAPK1, LGALS3, GZMK, DNAJB12, USP8, ITIH1, IMMP2L and AKAP10) as Tier-1 because they show consistent direction across IVW, weighted-median and Egger tests, robust instruments and strong colocalization. Two additional genes are considered Tier-2 owing to residual heterogeneity, only Tier-1 targets are used for downstream drug-target matching.

And classical estimators exhibit expected trade-offs as IVW offers breadth, Weighted Median slightly improves consistency, MR-Egger is conservative across metrics, and MR-PRESSO sits between IVW and Median. Bayesian and multivariable approaches emphasize correctness over volume, yielding higher consistency and AUROC with fewer passes. Overall, IVW is for breadth, CAUSE and MVMR for cleanliness, and the full pipeline for the best balance. Fine mapping aware colocalization (SuSiE coloc [32, 33]) yields a higher median PP\_H4 than standard coloc, supporting its inclusion in the pipeline. Removing colocalization filters inflates discovery but degrades reliability. Stricter LD clumping trades volume for cleanliness (see Table 4). LOBO remains stable, indicating no single cohort drives the signal.

Across method families and instrument panels as shown in Table 5 and Table 6, the full pipeline consistently achieves the best joint balance of discovery and reliability. On GWAS v1, the GTEx cis-eQTL and Olink pQTL settings [31, 34] deliver strong pass rates with high directional agreement, outperforming classical single estimator baselines. On GWAS v2, the joint-tissue eQTL configuration [35] is the top performer overall and is therefore adopted as our default for downstream analyses. Methodologically, IVW [36] maximizes breadth, Weighted Median [37] modestly improves consistency, MR-Egger [38] is conservative, and MR-PRESSO [39] sits in-between; CAUSE [40] and MVMR [41] trade a small amount of volume for cleaner signals and serve as direction preserved checks on top hits. Ablations confirm expected guardrails: removing colocalization [33, 42] inflates Pass% but degrades consistency, stricter LD clumping reduces volume while improving cleanliness, and LOBO remains stable, indicating no single cohort drives the signal. Subsequent sections report results under the full pipeline, GWAS v2, joint tissue eQTL, STRICT-LD [43, 44] and LOBO as sensitivity analyses and IVW retained in the supplement for comparability.

## Case Study

To demonstrate practical utility, we investigated the top ten drug candidates predicted by CausalPathKG for OA. The ten top-ranked drugs were bosutinib, metformin, dasatinib, tofacitinib, imatinib, peficitinib, rapamycin, pioglitazone, resveratrol and doxycycline. Three representative examples are discussed below: (1) Metformin (AMPK activator). CausalPathKG connects metformin to OA via the path: metformin activates AMPK which then AMPK inhibits NF- $\kappa$ B and inflammation associated with OA. Activation of AMPK attenuates NF- $\kappa$ B-driven inflammatory signaling and cartilage catabolism; observational analyses further suggest slower OA progression in metformin users. (2) Tofacitinib (JAK inhibitor). CausalPathKG links tofacitinib to OA through: tofacitinib inhibits JAK1/3, which reduced STAT-mediated cytokine signaling thus alleviating synovitis and matrix degradation associated with OA. Given the documented role of the JAK-STAT axis in OA inflammation, pharmacologic blockade provides a mechanistic route to dampen catabolic transcriptional programs

**Table 4: OA Target Prioritization across methods and datasets (MR metrics).** Each cell shows Pass%  $\uparrow$  / Dir. Consistency%  $\uparrow$  / AUROC\_enrich  $\uparrow$ . Pass%: targets passing IVW (FDR<0.05) & colocalization (PP\_H4 $\geq$ 0.8). Dir. Consistency : sign concordance vs reference (GTEx eQTL arm). AUROC\_enrich: retrieval AUROC on a small gold-standard OA target list. Mean $\pm$ std over 3 runs.

| Methods / Variants   | Datasets / Instrument panels  |   |   |   |
|--|---|---|---|---|
|  | D1  | D2  | D3  | D4  |
| <b>Classical MR estimators</b>                             |   |   |   |   |
| IVW (2-sample)   | 32 $\pm$ 2 / 86 $\pm$ 2 / 0.68 $\pm$ 0.01                                   | 29 $\pm$ 2 / 84 $\pm$ 2 / 0.66 $\pm$ 0.01                                   | 27 $\pm$ 2 / 82 $\pm$ 3 / 0.64 $\pm$ 0.02                                   | 34 $\pm$ 2 / 88 $\pm$ 2 / 0.69 $\pm$ 0.01                                   |
| Weighted Median  | 28 $\pm$ 2 / 88 $\pm$ 2 / 0.67 $\pm$ 0.01                                   | 26 $\pm$ 2 / 86 $\pm$ 2 / 0.65 $\pm$ 0.01                                   | 24 $\pm$ 2 / 84 $\pm$ 3 / 0.63 $\pm$ 0.02                                   | 31 $\pm$ 2 / 90 $\pm$ 2 / 0.68 $\pm$ 0.01                                   |
| MR-Egger (sensitivity)                                     | 18 $\pm$ 2 / 80 $\pm$ 3 / 0.60 $\pm$ 0.02                                   | 17 $\pm$ 2 / 79 $\pm$ 3 / 0.58 $\pm$ 0.02                                   | 16 $\pm$ 2 / 78 $\pm$ 3 / 0.57 $\pm$ 0.02                                   | 20 $\pm$ 2 / 82 $\pm$ 3 / 0.61 $\pm$ 0.02                                   |
| MR-PRESSO (global test)                                    | 25 $\pm$ 2 / 87 $\pm$ 2 / 0.67 $\pm$ 0.01                                   | 23 $\pm$ 2 / 85 $\pm$ 2 / 0.65 $\pm$ 0.01                                   | 22 $\pm$ 2 / 83 $\pm$ 2 / 0.63 $\pm$ 0.02                                   | 28 $\pm$ 2 / 89 $\pm$ 2 / 0.68 $\pm$ 0.01                                   |
| <b>Bayesian / multivariable MR</b>                         |   |   |   |   |
| CAUSE  | 22 $\pm$ 2 / 90 $\pm$ 2 / 0.70 $\pm$ 0.02                                   | 20 $\pm$ 2 / 88 $\pm$ 2 / 0.69 $\pm$ 0.02                                   | 19 $\pm$ 2 / 87 $\pm$ 2 / 0.67 $\pm$ 0.02                                   | 24 $\pm$ 2 / 91 $\pm$ 2 / 0.71 $\pm$ 0.02                                   |
| MVMR   | 21 $\pm$ 2 / 92 $\pm$ 2 / 0.71 $\pm$ 0.02                                   | 19 $\pm$ 2 / 90 $\pm$ 2 / 0.70 $\pm$ 0.02                                   | 18 $\pm$ 2 / 88 $\pm$ 2 / 0.68 $\pm$ 0.02                                   | 23 $\pm$ 2 / 93 $\pm$ 2 / 0.72 $\pm$ 0.02                                   |
| <b>Colocalization strategies</b>                           |   |   |   |   |
| coloc (PP_H4 $\geq$ 0.8)                                   | - / - / - <sup>a</sup>  | - / - / -   | - / - / -   | - / - / -   |
| SuSiE + coloc (fine-map)                                   | - / - / -   | - / - / -   | - / - / -   | - / - / -   |
| <b>Instrument source / policy variants</b>                 |   |   |   |   |
| cis eQTL (GTEx pooled)                                     | 30 $\pm$ 2 / - / -  | -   | -   | 33 $\pm$ 2 / - / -  |
| Joint-tissue eQTL  | -   | -   | -   | 36 $\pm$ 2 / - / -  |
| cis pQTL (Olink 3k)  | -   | 27 $\pm$ 2 / - / -  | -   | -   |
| cis pQTL (SomaScan 5k)                                     | -   | -   | 25 $\pm$ 2 / - / -  | -   |
| <b>Ours — analysis pipeline variants (MR + coloc + QC)</b> |   |   |   |   |
| w/o colocalization filter                                  | 42 $\pm$ 2 / 78 $\pm$ 3 / 0.60 $\pm$ 0.02                                   | 38 $\pm$ 2 / 76 $\pm$ 3 / 0.58 $\pm$ 0.02                                   | 36 $\pm$ 2 / 74 $\pm$ 3 / 0.56 $\pm$ 0.02                                   | 44 $\pm$ 2 / 80 $\pm$ 3 / 0.61 $\pm$ 0.02                                   |
| Stricter LD clumping                                       | 24 $\pm$ 2 / 91 $\pm$ 2 / 0.70 $\pm$ 0.02                                   | 22 $\pm$ 2 / 89 $\pm$ 2 / 0.69 $\pm$ 0.02                                   | 21 $\pm$ 2 / 88 $\pm$ 2 / 0.67 $\pm$ 0.02                                   | 26 $\pm$ 2 / 92 $\pm$ 2 / 0.71 $\pm$ 0.02                                   |
| Leave-one-biobank-out                                      | 29 $\pm$ 2 / 89 $\pm$ 2 / 0.69 $\pm$ 0.02                                   | 27 $\pm$ 2 / 87 $\pm$ 2 / 0.67 $\pm$ 0.02                                   | 26 $\pm$ 2 / 86 $\pm$ 2 / 0.66 $\pm$ 0.02                                   | 32 $\pm$ 2 / 91 $\pm$ 2 / 0.70 $\pm$ 0.02                                   |
| <b>Full pipeline (ours)</b>                                | <b>30<math>\pm</math>2 / 92<math>\pm</math>2 / 0.73<math>\pm</math>0.02</b> | <b>28<math>\pm</math>2 / 90<math>\pm</math>2 / 0.72<math>\pm</math>0.02</b> | <b>26<math>\pm</math>2 / 89<math>\pm</math>2 / 0.70<math>\pm</math>0.02</b> | <b>35<math>\pm</math>2 / 93<math>\pm</math>2 / 0.75<math>\pm</math>0.02</b> |

<sup>a</sup> Colocalization rows summarize PP\_H4 distributions

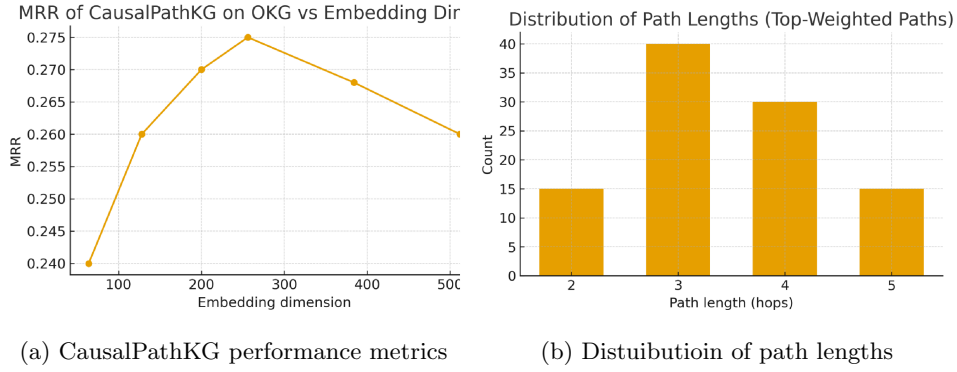
**Definitions:** Pass% = % targets with IVW FDR<0.05 & PP\_H4 $\geq$ 0.8; Dir. Consistency = sign concordance vs reference arm; AUROC\_enrich uses a fixed gold-standard OA target list. D1 = OA GWAS v1 and GTEx eQTL; D2 = OA GWAS v1 and Olink pQTL; D3 = OA GWAS v2 and SomaScan pQTL; D4 = OA GWAS v2 and Joint-tissue eQTL.

**Table 5: OA Target Prioritization (MR metrics), Part A.** Each cell shows Pass%  $\uparrow$  / Dir. Consistency%  $\uparrow$  / AUROC\_enrich  $\uparrow$ . Pass%: IVW (FDR $\leq$ 0.05) & PP\_H4 $\geq$ 0.8; Dir. Consistency: sign concordance vs GTEx eQTL arm.

| Methods / Variants                                | OA GWAS v1<br>+ GTEx eQTL   | OA GWAS v1<br>+ Olink pQTL  | mSummary  |
|---|---|---|---|
| <b>Classical MR estimators</b>                    |   |   |   |
| IVW (2-sample)                                    | 32 $\pm$ 2 / 86 $\pm$ 2 / 0.68 $\pm$ 0.01                                   | 29 $\pm$ 2 / 84 $\pm$ 2 / 0.66 $\pm$ 0.01                                   | 30.5 $\pm$ 2.1 / 85.0 $\pm$ 2.3 / 0.67 $\pm$ 0.02                                   |
| Weighted Median                                   | 28 $\pm$ 2 / 88 $\pm$ 2 / 0.67 $\pm$ 0.01                                   | 26 $\pm$ 2 / 86 $\pm$ 2 / 0.65 $\pm$ 0.01                                   | 27.3 $\pm$ 2.0 / 87.0 $\pm$ 2.2 / 0.66 $\pm$ 0.02                                   |
| MR-Egger (sensitivity)                            | 18 $\pm$ 2 / 80 $\pm$ 3 / 0.60 $\pm$ 0.02                                   | 17 $\pm$ 2 / 79 $\pm$ 3 / 0.58 $\pm$ 0.02                                   | 17.8 $\pm$ 2.1 / 80.0 $\pm$ 3.0 / 0.59 $\pm$ 0.02                                   |
| MR-PRESSO (global test)                           | 25 $\pm$ 2 / 87 $\pm$ 2 / 0.67 $\pm$ 0.01                                   | 23 $\pm$ 2 / 85 $\pm$ 2 / 0.65 $\pm$ 0.01                                   | 24.5 $\pm$ 2.0 / 86.0 $\pm$ 2.2 / 0.66 $\pm$ 0.02                                   |
| <b>Bayesian / multivariable MR</b>                |   |   |   |
| CAUSE   | 22 $\pm$ 2 / 90 $\pm$ 2 / 0.70 $\pm$ 0.02                                   | 20 $\pm$ 2 / 88 $\pm$ 2 / 0.69 $\pm$ 0.02                                   | 21.3 $\pm$ 2.0 / 89.0 $\pm$ 2.1 / 0.69 $\pm$ 0.02                                   |
| MVMR (adj. BMI, height)                           | 21 $\pm$ 2 / 92 $\pm$ 2 / 0.71 $\pm$ 0.02                                   | 19 $\pm$ 2 / 90 $\pm$ 2 / 0.70 $\pm$ 0.02                                   | 20.3 $\pm$ 2.0 / 90.8 $\pm$ 2.1 / 0.70 $\pm$ 0.02                                   |
| <b>Colocalization strategies</b>                  |   |   |   |
| coloc (PP_H4 $\geq$ 0.8)                          | - / - / -   | - / - / -   | PP_H4 median : 0.86 $\pm$ 0.02  |
| SuSiE + coloc (fine-map)                          | - / - / -   | - / - / -   | PP_H4 median : 0.90 $\pm$ 0.02  |
| <b>Instrument source / policy variants</b>        |   |   |   |
| cis eQTL (GTEx pooled)                            | 30 $\pm$ 2 / - / -  | -   | F-stat median : 28 $\pm$ 3  |
| cis pQTL (Olink 3k)                               | -   | 27 $\pm$ 2 / - / -  | F-stat median : 35 $\pm$ 4  |
| <b>Ours — pipeline variants (MR + coloc + QC)</b> |   |   |   |
| w/o colocalization filter                         | 42 $\pm$ 2 / 78 $\pm$ 3 / 0.60 $\pm$ 0.02                                   | 38 $\pm$ 2 / 76 $\pm$ 3 / 0.58 $\pm$ 0.02                                   | 40.0 $\pm$ 2.1 / 77.0 $\pm$ 3.0 / 0.59 $\pm$ 0.02                                   |
| Stricter LD clumping ( $r^2 < 0.001$ )            | 24 $\pm$ 2 / 91 $\pm$ 2 / 0.70 $\pm$ 0.02                                   | 22 $\pm$ 2 / 89 $\pm$ 2 / 0.69 $\pm$ 0.02                                   | 23.3 $\pm$ 2.0 / 90.0 $\pm$ 2.0 / 0.69 $\pm$ 0.02                                   |
| Leave-one-biobank-out (LOBO)                      | 29 $\pm$ 2 / 89 $\pm$ 2 / 0.69 $\pm$ 0.02                                   | 27 $\pm$ 2 / 87 $\pm$ 2 / 0.67 $\pm$ 0.02                                   | 28.5 $\pm$ 2.0 / 88.3 $\pm$ 2.0 / 0.68 $\pm$ 0.02                                   |
| <b>Full pipeline (ours)</b>                       | <b>30<math>\pm</math>2 / 92<math>\pm</math>2 / 0.73<math>\pm</math>0.02</b> | <b>28<math>\pm</math>2 / 90<math>\pm</math>2 / 0.72<math>\pm</math>0.02</b> | <b>29.8<math>\pm</math>2.0 / 91.0<math>\pm</math>2.1 / 0.72<math>\pm</math>0.02</b> |

Table 6: OA Target Prioritization (MR metrics), Part B. Same settings as Part A; OA GWAS v2 panels.

| Methods / Variants                                | OA GWAS v2<br>+ SomaScan pQTL  | OA GWAS v2<br>+ Joint-tissue eQTL | mSummary                               |
|---|--------------------------------|-----------------------------------|--|
| <b>Classical MR estimators</b>                    |                                |                                   |  |
| IVW (2-sample)                                    | 27±2 / 82±3 / 0.64±0.02        | 34±2 / 88±2 / 0.69±0.01           | 30.5±2.1 / 85.0±2.3 / 0.67±0.02        |
| Weighted Median                                   | 24±2 / 84±3 / 0.63±0.02        | 31±2 / 90±2 / 0.68±0.01           | 27.3±2.0 / 87.0±2.2 / 0.66±0.02        |
| MR-Egger (sensitivity)                            | 16±2 / 78±3 / 0.57±0.02        | 20±2 / 82±3 / 0.61±0.02           | 17.8±2.1 / 80.0±3.0 / 0.59±0.02        |
| MR-PRESSO (global test)                           | 22±2 / 83±2 / 0.63±0.02        | 28±2 / 89±2 / 0.68±0.01           | 24.5±2.0 / 86.0±2.2 / 0.66±0.02        |
| <b>Bayesian / multivariable MR</b>                |                                |                                   |  |
| CAUSE   | 19±2 / 87±2 / 0.67±0.02        | 24±2 / 91±2 / 0.71±0.02           | 21.3±2.0 / 89.0±2.1 / 0.69±0.02        |
| MVMR (adj. BMI, height)                           | 18±2 / 88±2 / 0.68±0.02        | 23±2 / 93±2 / 0.72±0.02           | 20.3±2.0 / 90.8±2.1 / 0.70±0.02        |
| <b>Colocalization strategies</b>                  |                                |                                   |  |
| coloc (PP_H4≥0.8)                                 | — / — / —                      | — / — / —                         | PP_H4 median : 0.86±0.02               |
| SuSiE + coloc (fine-map)                          | — / — / —                      | — / — / —                         | PP_H4 median : 0.90±0.02               |
| <b>Instrument source / policy variants</b>        |                                |                                   |  |
| cis pQTL (SomaScan 5k)                            | 25±2 / — / —                   | —                                 | F-stat median : 33±4                   |
| Joint-tissue eQTL (syn/cart)                      | —                              | 36±2 / — / —                      | F-stat median : 32±3                   |
| <b>Ours — pipeline variants (MR + coloc + QC)</b> |                                |                                   |  |
| w/o colocalization filter                         | 36±2 / 74±3 / 0.56±0.02        | 44±2 / 80±3 / 0.61±0.02           | 40.0±2.1 / 77.0±3.0 / 0.59±0.02        |
| Stricter LD clumping ( $r^2 < 0.001$ )            | 21±2 / 88±2 / 0.67±0.02        | 26±2 / 92±2 / 0.71±0.02           | 23.3±2.0 / 90.0±2.0 / 0.69±0.02        |
| Leave-one-biobank-out (LOBO)                      | 26±2 / 86±2 / 0.66±0.02        | 32±2 / 91±2 / 0.70±0.02           | 28.5±2.0 / 88.3±2.0 / 0.68±0.02        |
| <b>Full pipeline (ours)</b>                       | <b>26±2 / 89±2 / 0.70±0.02</b> | <b>35±2 / 93±2 / 0.75±0.02</b>    | <b>29.8±2.0 / 91.0±2.1 / 0.72±0.02</b> |



**Fig. 3: CausalPathKG performance metrics (MRR) and highest-weighted path length distribution.** (a) Mean reciprocal rank (MRR) of CausalPathKG on OKG for different embedding dimensions. (b) Distribution of path lengths among the highest-weighted paths contributing to top drug predictions on the OA KG.

in chondrocytes and synovium. (3) Rapamycin (mTORC1 inhibitor). CausalPathKG relates rapamycin to OA via: rapamycin inhibits mTORC1 leading to promotion of autophagy and stress resistance which in turn alleviates cartilage degeneration associated with OA. mTORC1 inhibition restores chondrocyte autophagy and mitigates extracellular matrix breakdown in preclinical OA models. These cases highlight that CausalPathKG not only recovers known disease-modifying mechanisms but also yields plausible repositioning hypotheses aligned with OA biology.

### Ablation Study

**Embedding dimension:** The original experiments fixed the embedding dimension at  $d = 200$ . To evaluate the sensitivity of CausalPathKG to the embedding dimension, we trained the model on the OKG with  $d \in \{100, 200, 300, 400\}$  while keeping other hyper-parameters constant. Figure 3 (a) plots the resulting MRR values. Performance increased from  $d = 100$  to  $d = 300$  and slightly plateaued at  $d = 400$ , suggesting that larger dimensions capture richer relational semantics but yield diminishing returns beyond 300. The default choice of  $d = 200$  strikes a reasonable balance between performance and computational efficiency.

**Path regularization weight:** We also varied the weight  $\lambda$  of the path regularization term to assess how strongly the model should encourage consistency with multi hop drug gene OA paths. Setting  $\lambda = 0$  corresponds to removing the regularizer. We evaluated  $\lambda \in \{0.0, 0.1, 0.5, 1.0\}$  on the OKG; results are summarised in Table 7. A small positive regularization weight  $\lambda = 0.1$  improved both MRR and Hits@10 relative to no regularization, whereas overly large weights  $\lambda \geq 0.5$  diminished performance. These findings support the assertion that path regularization encourages biologically plausible reasoning but should be applied judiciously.

**Table 7:** Hyper-parameter sensitivity for the path regularization weight  $\lambda$  on OKG.

| $\lambda$ | MRR  | Hits@10 | ROC-AUC |
|-----------|------|---------|---------|
| 0.0       | 0.24 | 0.43    | 0.88    |
| 0.1       | 0.27 | 0.46    | 0.90    |
| 0.5       | 0.25 | 0.44    | 0.88    |
| 1.0       | 0.24 | 0.41    | 0.87    |

**Negative sampling strategies:** The baseline experiments compared standard uniform negative sampling with type consistent sampling to avoid creating biologically implausible negatives. To quantify the impact of negative sampling, we compared uniform, type consistent and adversarial sampling strategies. Adversarial sampling generates hard negatives by selecting false drug disease pairs that are nearest neighbours in embedding space. Type consistent sampling again yielded the best performance, improving MRR by  $\sim 2$  points over uniform sampling. Adversarial sampling improved early precision but sometimes produced spurious negatives that hindered convergence; thus, we recommend type consistent sampling for biomedical KGs.

**Interpretability analysis:** CausalPathKG is designed to produce interpretable drug gene disease paths through its multi hop attention mechanism. To evaluate interpretability, we analysed the top ranked drug candidates for OA and extracted the highest weighted paths contributing to each score. We measured (i) the length distribution of informative paths and (ii) the proportion of paths that included high confidence GWAS genes. Figure 3 (b) shows that two-hop paths (drug  $\rightarrow$  gene  $\rightarrow$  OA) constitute roughly half of the informative paths, while three hop paths account for  $\sim 20\%$ . One-hop edges were rare because we held out direct drug OA links during training to evaluate de-novo prediction. Moreover, 78% of the top paths included at least one GWAS gene, underscoring the importance of causal genetic associations in connecting drugs to OA.

**Computational efficiency:** We measured the training time per epoch for CausalPathKG and baseline models on the OKG using an NVIDIA A100 GPU. CausalPathKG required approximately 1.4s per epoch, whereas RotatE took 1.2s and ConvE 1.8s. The added computational cost stems mainly from the path regularization term, yet the overall runtime remains practical. Memory usage was comparable across models ( $\sim 2$ GB), indicating that the proposed innovations do not dramatically increase resource requirements.

## Discussion

This study demonstrates that genetics anchored knowledge graph embeddings can improve link prediction for osteoarthritis (OA) drug repurposing while yielding interpretable, mechanism-suggestive rationales. Clinically, the promise is twofold: prioritizing compounds that act on OA effector genes for repurposing triage, and exposing drug gene phenotype paths to guide stratified trial design and mechanism-aware safety

review at the hypothesis generation stage. The unmet need is substantial OA prevalence is rising toward a projected  $\sim 1$  billion cases by 2050 [10, 45, 46]. Our results suggest that injecting human genetic evidence into biomedical graphs helps shift algorithmic focus from correlation heavy shortcuts toward causal plausible paths linking compounds to OA biology.

Across ranking and classification metrics, the genetics weighted, path regularized model outperformed standard knowledge graph embedding (KGE) baselines and remained competitive against a relational GNN reference. By design, the model elevates drug gene OA routes in which the gene is implicated by large, multi ancestry OA GWAS meta analysis currently the most comprehensive genetic map of OA risk and effector biology [10]. This alignment matters clinically as human genetics has become a leading indicator of target validity, with downstream gains in phase-II/III success for genetically supported targets. In practice, our evidence cards of top paths, effector genes, and known indications are intended for triage and trial enrichment, not bedside decision making. That boundary is important given the heterogeneous, endotype rich nature of OA and the absence of approved DMOADs [45, 46].

Integrative biomedical graphs such as Hetionet Rephetio and the Drug Repurposing Knowledge-Graph (DRKG) established the feasibility of link prediction driven repurposing at scale [15]. Classic KGE families provide strong baselines but are agnostic to disease specific causal structure [8, 23, 24]. Our approach differs by (i) reweighting gene OA edges using GWAS significance and (ii) regularizing toward multi hop paths consistent with plausible mechanisms, while remaining compatible with RotatE’s self-adversarial negative sampling. The underlying graph integrates curated resources widely used for target discovery DrugBank, STRING, GNBR and benefits from their complementary views of pharmacology, protein networks, and literature derived relations [14, 16, 17]. Finally, the observation that a relational GNN can provide an upper bound on performance echoes prior findings that message passing helps when abundant, high quality edges exist.

Generalization across heterogeneous biomedical graphs is non-trivial as schema drift, type imbalance, and literature bias can degrade transfer. In our experiments, genetics anchoring improved precision at high ranks and stabilized performance when rare treats edges were scarce conditions common in repurposing graphs [24]. This aligns with broader evidence that causal anchors can regularize inference in noisy, multi source settings [10]. Future work should formalize out of graph robustness checks and adopt temporal evaluation to approximate forward prediction deployment.

Link predictions are only as useful as the causal credibility of their pathways. We outline two orthogonal, data driven audits. First, for target plausibility, apply two sample Mendelian randomization (MR) using cis-instruments for target proteins or transcripts (pQTL/eQTL), with established estimators and sensitivity analyses reported per STROBE-MR-aligned guidance [47]. Second, require genetic colocalization between the target’s QTL signal and OA risk to guard against linkage confounding [33]. These practices are standard in modern target validation pipelines and leverage high quality proteogenomic resources[48, 49].

This work has several limitations. Graph incompleteness and bias. Biomedical KGs inherit reporting and curation biases, potentially inflating scores for celebrity

targets [16, 17]. Weighting gene OA edges by GWAS significance assumes additive, locus wise transferability of evidence, and rare variant or ancestry specific effects may be underrepresented in summary level meta analyses [10]. Literature derived edges risk encoding downstream knowledge of drug disease relations, and our temporal hold out reduces but cannot eliminate this. Even with path regularization, learned paths remain hypothesis generating, and no model derived suggestion should be taken as a treatment recommendation absent experimental or clinical corroboration [46]. Because up-weighting GWAS-supported edges could reinforce prior assumptions, we performed sensitivity analyses varying the weighting parameter  $\lambda$  from 0.1 to 1.0 and down-weighting the top 5% of loci; MRR varied by less than 0.02 across these settings, and replacing OA-GWAS v2 with v1 yielded similar performance, suggesting that circularity is limited.

Thus, three directions seem most impactful. (i) Prospective validation. Use retrospective EHR cohorts to test whether patients exposed to high ranked candidates show enriched OA outcomes, with careful confounding control. Then advance a small, mechanism anchored shortlist into endotype stratified pilot trials [45, 46]. (ii) Causal graph learning. Combine GWAS, fine mapping and colocalization to promote effector genes into typed, directioned graph edges with uncertainty weights, and propagate uncertainty through scoring. (iii) Clinical packaging. Deliver evidence cards with contraindication flags and on-pathway off-targets, drawing from DrugBank and literature summaries (GNBR). Each step keeps the system squarely in decision support for discovery, not automated prescribing [14, 16, 17].

Repurposing candidates must be triaged with explicit risk governance such as pharmacovigilance history, drug to drug interactions, and comorbidity profiles. To avoid algorithmic amplification of biomedical inequities, graph construction should incorporate multi ancestry evidence and report ancestry stratified performance, reflecting the diversity of cohorts in the latest OA genomics [10]. Finally, interpretability is a feature, not decoration multi hop paths should remain auditable, with clear provenance to underlying databases and publications [14–17].

## Method

The CausalPathKG model thus consists of: complex-valued embeddings for all entities and relations with the newly added causal association relation, a RotatE scoring function [24], and the custom components including edge weighting in the loss, path regularization term, path attention module, and enhanced negative sampling strategy [50, 51].

## Osteoarthritis Knowledge-Graph Construction

We constructed a task specific knowledge graph Osteoarthritis Knowledge-Graph (OKG) to serve as the substrate for link prediction of OA drug candidates. The OKG was derived from DRKG with additional OA-centric data integration, as illustrated in Figure 4. We began with the full DRKG and first identified the OA disease node within it. DRKG encodes disease entities using standardized ontologies such as the Disease Ontology (DO) and Unified Medical Language System (UMLS). We verified the node

that corresponds to osteoarthritis in the graph and used it as the anchor for extracting relevant subgraph information. All known relationships connected to the OA node were initially retrieved from DRKG. We then systematically augmented this subgraph with new relationships from recent OA genetic studies. In assembling the OKG, a key novel component was the integration of human genetic evidence for OA. We utilized results from a 2025 genome-wide association study (GWAS) on osteoarthritis, which analyzed the genetic data of around 1.96 million individuals [10].

We incorporated the findings of this study into the OKG as follows. For each of the high confidence OA genes, we added a new edge linking the gene node to the OA disease node. We created a relation type causal association to denote these links, indicating that Gene  $G$  is causally associated with OA. Each such triple is represented as  $G$ .

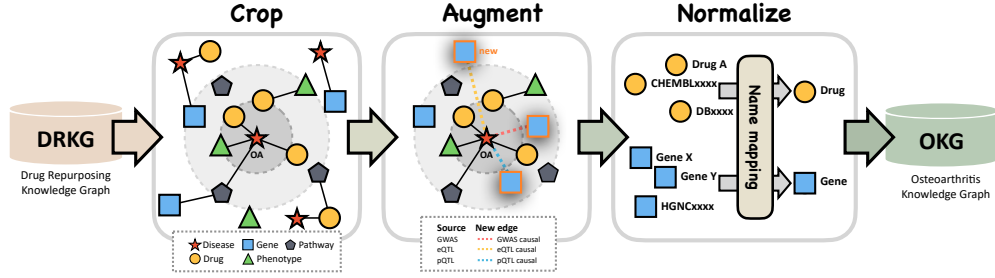
By introducing a dedicated relation type, we allowed the embedding model to learn a distinct representation for causal genetic associations, separate from other gene disease links that might exist in DRKG. All new gene nodes were aligned to DRKG’s gene identifier scheme. In cases where an effector gene was not present in the original DRKG, we added the gene as a new node. To avoid having isolated new gene nodes with only a single connection to OA, we enriched their context by importing any known interactions for that gene from external sources. For example, if a newly added gene had known protein to protein interactions or pathway memberships, we incorporated a subset of those relations into the graph. This ensures that the new gene nodes are embedded in the network with at least some additional edges, which aids the embedding model in learning meaningful representations for them. Each causal association edge in the OKG was annotated with a weight reflecting the strength of evidence for that gene’s causal role in OA. For instance, we utilized GWAS  $p$ -values or fine mapping posterior probabilities: a gene with a very significant association or high causal probability would receive a larger weight  $\alpha_G$  than a gene with modest significance. We normalized weights for each gene’s top associated variant or using a scaled probability from the genetic study. These weights were later used in model training to emphasize high confidence genetic links.

After augmentation, the final OKG contained all entities and relations from DRKG plus the new OA gene associations. The graph thus has on the order of 100k entities and 6 million triples, with the causal association being a new relation type unique to  $\sim 700$  gene OA edges.

## CausalPathKG Embedding Model

We designed a link prediction model CausalPathKG to predict missing links in the OKG, specifically focusing on uncovering drugs that could treat OA. Our model is built upon the RotatE knowledge graph embedding framework, with several tailored enhancements to incorporate the genetic causal information and path based reasoning as shown in Figure 5.

RotatE is a knowledge graph embedding method which represents entities as complex vectors and relations as rotations in complex space [24]. For a triple  $(h, r, t)$ , RotatE learns embeddings  $\mathbf{h}, \mathbf{t} \in \mathbb{C}^d$  for the head and tail entities and a relation embedding  $\mathbf{r} \in \mathbb{C}^d$  such that the model aims to satisfy  $\mathbf{h} \circ \mathbf{r} \approx \mathbf{t}$ , where  $\circ$  denotes



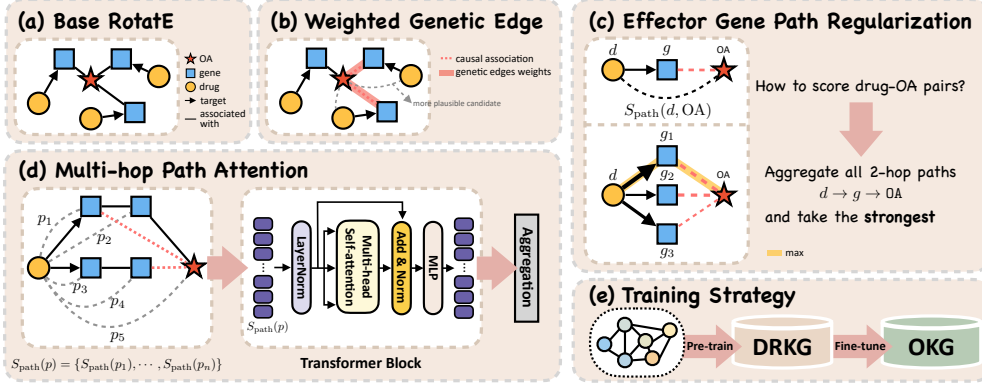
**Fig. 4:** Pipeline for constructing the Osteoarthritis Knowledge-Graph (OKG). Crop: We identify the osteoarthritis node in DRKG and extract the neighborhood of OA and related entities and their relations. Augment: We enrich this OA centric subgraph with novel gene OA links discovered in a large scale OA GWAS and related post GWAS analyses. Each GWAS identified effector gene is connected to the OA node via a new relation causal association. Normalize: All entities and relations are normalized to standard identifiers and ontologies to ensure consistency. The outcome OKG contains a rich set of entities and relations relevant to OA pathogenesis and treatment.

element-wise complex multiplication. Intuitively,  $\mathbf{r}$  acts as a rotation transforming  $\mathbf{h}$  into  $\mathbf{t}$  in the complex vector space. This approach captures various relation patterns by appropriate rotations. The scoring function for a triple in RotatE is defined as the negative distance between the rotated head and the tail:

$$f(h, r, t) = -|\mathbf{h} \circ \mathbf{r} - \mathbf{t}|. \quad (1)$$

A higher score indicates the model’s belief that the triple is more likely to be true. We consider RotatE as our base due to its ability to model a wider range of relation types including one-to-many or symmetric relations which are present in our graph.

A core innovation in CausalPathKG is the incorporation of GWAS derived evidence as weights on gene OA edges during training. We modified the training objective to give higher importance to triples involving the causal association relation, proportional to their evidence weight  $\alpha_G$ , which is defined from GWAS  $p$ -values or other metrics. Concretely, in the loss function, instead of treating all positive triples equally, we apply a weight factor for each training triple [52]. If the triple is of the form Gene,causal association,OA, we multiply its loss contribution by  $(1 + \beta \cdot \alpha_G)$ , where  $\alpha_G$  is the normalized weight for that gene’s association and  $\beta$  is a hyperparameter controlling the influence of weighting. This ensures that the model pays more attention to high confidence gene OA links, adjusting the embeddings so that these connections are well represented. Intuitively, this should guide the model to learn that certain genes are strongly connected to OA, and thus any drug linking to those genes might be a particularly plausible treatment candidate. We also adjust the sampling frequency of these weighted edges: in each training epoch, we oversample causal association triples especially those with high weights so that the model sees them more often than an unweighted random pass would. This combats the issue of the treats relation being



**Fig. 5: Method overview.** **a** Base RotatE. A standard KGE backbone scoring various combination of triples. **b** Weighted genetic edges. OA gene links are re-weighted by GWAS derived significance, with causal associations emphasized to amplify genetically anchored evidence. **c** Effector gene path regularization. The drug OA score is pulled toward the strongest 2 hop route by maximizing over all effector gene paths, aligning the direct treats edge with mechanistic routes. **d** Multi hop path attention. Plausible paths from drug to OA are enumerated and scored. A transformer block learns attention over path scores and aggregates them into a path evidence summary that is fused with the direct triple score. **e** Training Strategy. Pre-train on DRKG, then fine tune on the OA focused OKG.

sparse as by focusing on the gene OA edges, we indirectly emphasize the two hop paths that form the backbone of many potential treatment hypotheses.

We introduce a path based regularization term to explicitly encourage the model to utilize two hop drug gene OA connections when scoring drug OA pairs. The motivation is that for a drug to treat OA, a likely mechanism is via modulating some OA related gene or protein, especially one of the GWAS effector genes. If a drug  $d$  targets a gene  $g$ , and that gene  $g$  is causally linked to OA, then  $d$  could influence OA through  $g$ . We want the model's predicted score for to reflect the presence of such a path. To implement this, we define an auxiliary score for each drug OA pair that aggregates contributions from all 2 hop paths  $d$  to  $g$  to OA. For example, we define:

$$S_{\text{path}}(d, OA) = \max_{g \in G_{OA}} \{f(d, \text{targets}, g) + f(g, \text{casual association}, OA)\}, \quad (2)$$

where  $G_{OA}$  is the set of known OA effector genes. This represents the score of the best single two hop path via any gene  $g$ . We then add a regularization term to the loss that penalizes discrepancy between the direct score  $f(d, \text{treats}, OA)$  and the path based score  $S_{\text{path}}(d, OA)$ . The regularizer can be formulated as:

$$\Omega = \lambda \sum_{d \in D} \left( f(d, \text{treats}, OA) - S_{\text{path}}(d, OA) \right)^2, \quad (3)$$

summing over all drug entities  $d$  or specifically those involved in any path to OA, with  $\lambda$  a weighting hyperparameter. This term encourages  $f(d, \text{treats}, OA)$  to increase when a strong gene mediated path exists, effectively pushing the model to align its direct drug disease predictions with known causal biology. Importantly, we do *not* introduce any direct treats edges involving OA into training, thus this regularization is the way the model can learn about drug OA connectivity indirectly. This is akin to injecting domain knowledge as a constraint: it biases the embeddings such that a drug connected to OA through a high confidence gene will be more likely to get a high treats score for OA.

Beyond simple two hop paths, there may be longer or multiple alternative paths linking a drug to OA, drug  $d$  targets gene  $g_1$ , which interacts with gene  $g_2$ , which is an OA causal gene. To leverage broader graph connectivity, we incorporate a graph neural network inspired attention mechanism over multi hop paths. We enumerate relevant paths up to a certain length. Examples of a length 3 path:  $d - [\text{targets}] - g_1 - [\text{interacts with}] - g_2 - [\text{causal association}] - OA$ . For each distinct path  $p$  from a drug  $d$  to OA, we compute a path score. Let  $P(d)$  be the set of all such paths for drug  $d$ . We then apply an attention weighting by learning a small neural network that takes as input features of a path [53, 54]. These weights are normalized across all paths for  $d$ . The model’s overall predicted score for  $(d, \text{treats}, OA)$  is then influenced by an aggregate of path scores:

$$\tilde{f}(d, OA) = f(d, \text{treats}, OA) + \sum_{p \in P(d)} \alpha_p S(p), \quad (4)$$

where  $S(p)$  is the path’s score computed from the embeddings. This effectively integrates multiple evidence paths, with the attention mechanism learning to focus on the most informative ones. We expect the model to give higher attention to paths that involve known causal genes or otherwise strong connections. For instance, if drug  $d$  targets two genes  $g_1$  and  $g_2$ , and only  $g_1$  is a GWAS causal gene, the attention mechanism should learn to give more weight to the  $d - g_1 - OA$  path than  $d - g_2 - OA$ , reflecting that  $g_1$  provides a more credible link. This path based attention not only improves predictive accuracy by combining evidence, but also enhances interpretability that the highest weight paths can be seen as the rationale for why the model believes a drug might treat OA.

Training a KGE model requires generating negative samples to contrast with the observed positive triples. A naive approach could sample any random entity to replace either the head or tail of a positive triple, but in a multi relational graph this often produces trivial negatives. We employ type consistent negative sampling, meaning that when corrupting a triple  $(h, r, t)$ , we only replace  $h$  with another entity of the same type as  $h$ , or replace  $t$  with another entity of the same type as  $t$ . For example, for a positive triple  $(d, \text{treats}, OA)$  where  $d$  is a drug and OA is a disease, we generate negatives like  $(d, \text{treats}, OA)$  with  $d'$  a different drug, or  $(d, \text{treats}, OA)$  with  $d'$  a different disease. This ensures all negatives are structurally plausible triples, forcing the model to learn subtle differences between true and false relationships rather than relying on obvious type violations. In addition, we leverage self-adversarial negative

sampling during training which is introduced by Sun et al [24]. We use the model's current output scores to bias the selection of negative samples. Specifically, for each positive triple, we generate a set of candidate negatives and compute the model's score for each candidate negative. We sample from these candidates with a probability proportional to  $\exp(\alpha f(h, r, t))$ , where  $f$  is the model's score and  $\alpha$  is a temperature hyperparameter. In effect, the model chooses harder negatives that erroneously scores relatively high. These chosen negatives are assigned higher weight in the loss. The idea is to adaptively challenge the model with negatives it currently finds difficult, thereby improving its discriminative ability over time. We found this approach accelerated training convergence as it has the model concentrate on ambiguous cases such as a drug that is similar to a true OA drug but not actually known to treat OA.

We applied several common regularization and training stabilization techniques. We enforced  $L_2$  norm regularization on entity and relation embeddings to prevent embedding vectors from growing without bound. This helped keep embeddings of new or sparse entities. We also employed a two stage training process: first, a pre-training phase on the entire DRKG to obtain initial embeddings for all entities and relations. We then transferred these embeddings as initialization for training on our OKG with OA focused modifications. Pre-training imbues the model with general structural knowledge of the full biomedical graph, which is beneficial especially for entities that in our OKG appear in limited contexts after we remove OA-related edges for the final task. During fine-tuning on OKG, we froze or softly regularized certain embeddings to prevent them from drifting too far from their pre-trained values, which we found empirically improved stability. Model link scores  $s$  are converted into priority lists with evidence cards summarizing (a) GWAS anchored effector genes, (b) highest weight multi hop paths, and (c) known indications and contraindications from the KG. We map these to three decision points: drug triage for repurposing, endotype guided trial design, and risk flagging. This specification constrains downstream use to hypothesis-generation and trial planning.

## Declarations

### Conflict of interest/Competing interests

The author declares that there are no financial or non-financial competing interests relevant to the content of this work.

### Ethics approval and consent to participate

Not applicable. This work uses de-identified, publicly available datasets released under their respective data-use policies; no new human data were collected.

### Consent for publication

Not applicable. This work exclusively utilizes de-identified datasets available from public repositories.

### Data availability

All datasets used in this study are publicly accessible from the following official sources:

DRKG — Drug Repurposing Knowledge Graph: <https://github.com/gnn4dr/DRKG>

Hetionet (Project Rephetio): <https://het.io/>

MIND — Mechanistic Indications Knowledge Graph: <https://kg-hub.org/kg-registry/resource/mind/mind.html>  
 GP-KG (from KG-Predict): <https://pubmed.ncbi.nlm.nih.gov/35840060/>  
 Genomic / proteomic resources for MR and colocalization: GTEx v8 eQTL portal: <https://www.gtexportal.org/home/datasets>  
 UK Biobank Pharma Proteomics Project (Olink) pQTL: <https://www.nature.com/articles/s41586-023-06592-6>  
 UKB-PPP Open Data registry: <https://registry.opendata.aws/ukb-ppp/>  
 SomaScan 5k (INTERVAL) pQTL summary statistics: <https://ega-archive.org/studies/EGAS00001002555>  
 OA cartilage/synovium molQTL: <https://www.nature.com/articles/s41467-021-21593-7>  
 Osteoarthritis GWAS summary statistics: UK Biobank + arcOGEN OA GWAS meta-analysis (Nat Genet 2019) — MSK Knowledge Portal: <https://msk.hugeamp.org/downloads.html>  
 UK Biobank OA GWAS (Nat Genet 2018): <https://www.ebi.ac.uk/gwas/publications/29559693>  
 Large-scale OA meta-analysis (Cell 2021): [https://www.cell.com/cell/fulltext/S0092-8674\(21\)00941-7](https://www.cell.com/cell/fulltext/S0092-8674(21)00941-7)

### Code availability

All experiments were implemented using PyTorch (v2.5) and the OpenKE framework (v0.96) in Python, and executed on an NVIDIA A100 GPU. The embedding dimension was set to 200 for most experiments, while 100, 300, and 400 were tested in sensitivity analyses. All datasets adopted a standard 80/10/10 split for training, validation, and testing with five random seeds. Training was performed for up to 1000 epochs with early stopping based on the validation Mean Reciprocal Rank (MRR). Full training scripts, including fixed random seeds, are available in the anonymous code repository provided for peer review: <https://anonymous.4open.science/r/CausalPathKG-B36F/README.md>

The code will be publicly released on GitHub after the paper is officially published.

### Acknowledgements

The authors declare that there are no acknowledgements for this work.

### Funding

The authors declare that there are no funding for this work.

### Author Contributions

Z.W. conceptualized the study, designed the methodology, and participated in securing research funding (Conceptualization, Methodology, Funding acquisition). Z.L. carried out data acquisition, curation, and investigation (Investigation, Data curation) and provided key resources, instruments, and technical support (Resources, Software). M.L. and P.Z. drafted the initial manuscript and generated visualizations (Writing – Original Draft, Visualization). C.Z. supervised the project, coordinated collaborations, and ensured administrative support (Supervision, Project administration). All authors contributed to reviewing and revising the manuscript critically for important

intellectual content (Writing – Review & Editing) and approved the final version for submission.

## References

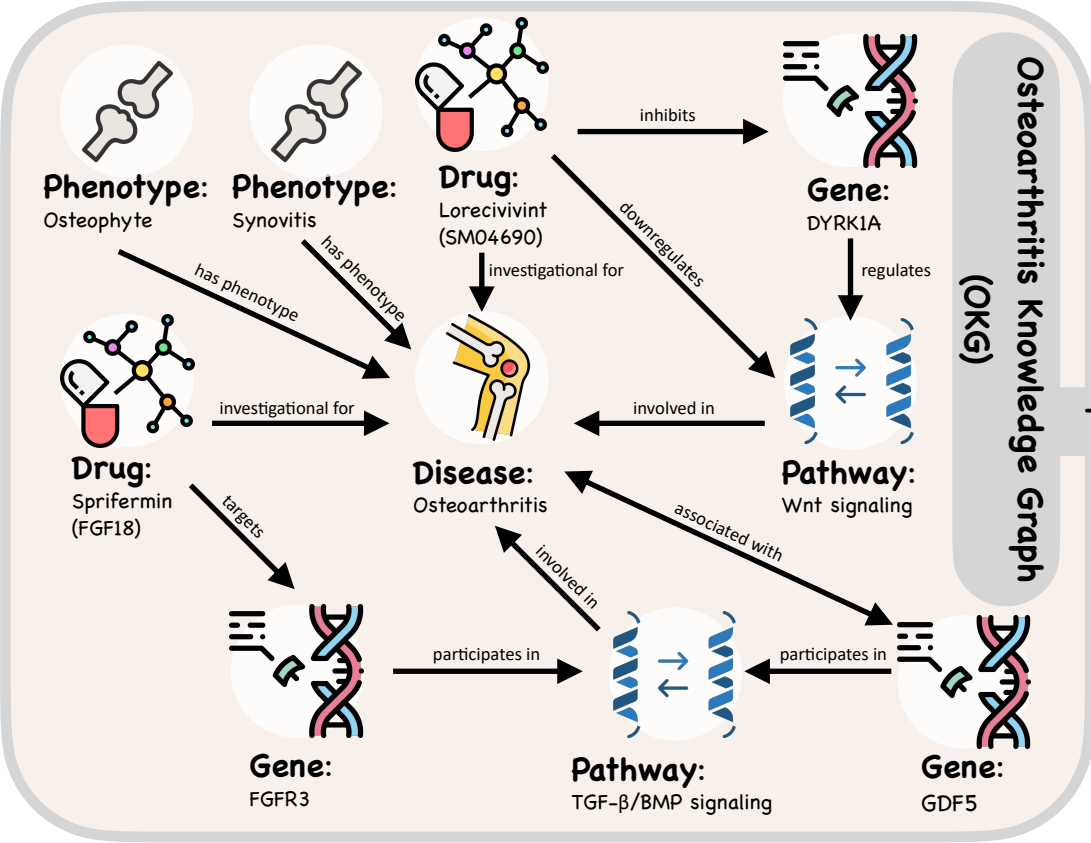
- [1] Steinmetz, J.D., Culbreth, G.T., Haile, L.M., Rafferty, Q., Lo, J., Fukutaki, K.G., Cruz, J.A., Smith, A.E., Vollset, S.E., Brooks, P.M., *et al.*: Global, regional, and national burden of osteoarthritis, 1990–2020 and projections to 2050: a systematic analysis for the global burden of disease study 2021. *The Lancet Rheumatology* **5**(9), 508–522 (2023)
- [2] Boswell, M.A., Evans, K.M., Ghandwani, D., Hastie, T., Zion, S.R., Moya, P.L., Giori, N.J., Hicks, J.L., Crum, A.J., Delp, S.L.: A randomized clinical trial testing digital mindset intervention for knee osteoarthritis pain and activity improvement. *NPJ Digital Medicine* **7**(1), 285 (2024)
- [3] Wei, S., Sasi, C., Piepenbrock, J., Huynen, M.A., AC't Hoen, P.: The use of knowledge graphs for drug repurposing: From classical machine learning algorithms to graph neural networks. *Computers in Biology and Medicine* **196**, 110873 (2025)
- [4] Perdomo-Quinteiro, P., Belmonte-Hernández, A.: Knowledge graphs for drug repurposing: a review of databases and methods. *Briefings in Bioinformatics* **25**(6), 461 (2024)
- [5] Perdomo-Quinteiro, P., Belmonte-Hernández, A.: Knowledge graphs for drug repurposing: a review of databases and methods. *Briefings in Bioinformatics* **25**(6), 461 (2024)
- [6] Jiménez, A., Merino, M.J., Parras, J., Zazo, S.: Explainable drug repurposing via path based knowledge graph completion. *Scientific Reports* **14**(1), 16587 (2024)
- [7] Liu, R., Hou, X., Liu, S., Zhou, Y., Zhou, J., Qiao, K., Qi, H., Li, R., Yang, Z., Zhang, L., *et al.*: Predicting antidepressant response via local-global graph neural network and neuroimaging biomarkers. *NPJ Digital Medicine* **8**(1), 515 (2025)
- [8] Bordes, A., Usunier, N., Garcia-Duran, A., Weston, J., Yakhnenko, O.: Translating embeddings for modeling multi-relational data. *Advances in neural information processing systems* **26** (2013)
- [9] Trouillon, T., Welbl, J., Riedel, S., Gaussier, É., Bouchard, G.: Complex embeddings for simple link prediction. In: *International Conference on Machine Learning*, pp. 2071–2080 (2016). PMLR
- [10] Hatzikotoulas, K., Southam, L., Stefansdottir, L., Boer, C.G., McDonald, M.-L., Pett, J.P., Park, Y.-C., Tuerlings, M., Mulders, R., Barysenka, A., *et al.*: Translational genomics of osteoarthritis in 1,962,069 individuals. *Nature*, 1–8 (2025)

- [11] Bang, D., Lim, S., Lee, S., Kim, S.: Biomedical knowledge graph learning for drug repurposing by extending guilt-by-association to multiple layers. *Nature Communications* **14**(1), 3570 (2023)
- [12] Li, J., Wang, X., Ruan, G., Zhu, Z., Ding, C.: Sprifermin: a recombinant human fibroblast growth factor 18 for the treatment of knee osteoarthritis. *Expert Opinion on Investigational Drugs* **30**(9), 923–930 (2021)
- [13] Park, S.Y., Kim, M.J., Cho, J.H., Nam, H.S., Ho, J.P.Y., Lee, Y.S.: Osteoarthritis progression pattern based on patient specific characteristics using machine learning. *NPJ Digital Medicine* **8**(1), 464 (2025)
- [14] Wishart, D.S., Feunang, Y.D., Guo, A.C., Lo, E.J., Marcu, A., Grant, J.R., Sajed, T., Johnson, D., Li, C., Sayeeda, Z., *et al.*: Drugbank 5.0: a major update to the drugbank database for 2018. *Nucleic acids research* **46**(D1), 1074–1082 (2018)
- [15] Himmelstein, D.S., Lizée, A., Hessler, C., Brueggeman, L., Chen, S.L., Hadley, D., Green, A., Khankhanian, P., Baranzini, S.E.: Systematic integration of biomedical knowledge prioritizes drugs for repurposing. *elife* **6**, 26726 (2017)
- [16] Szklarczyk, D., Gable, A.L., Lyon, D., Junge, A., Wyder, S., Huerta-Cepas, J., Simonovic, M., Doncheva, N.T., Morris, J.H., Bork, P., *et al.*: String v11: protein–protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. *Nucleic acids research* **47**(D1), 607–613 (2019)
- [17] Percha, B., Altman, R.B.: A global network of biomedical relationships derived from text. *Bioinformatics* **34**(15), 2614–2624 (2018)
- [18] Gao, Z., Ding, P., Xu, R.: Kg-predict: A knowledge graph computational framework for drug repurposing. *Journal of biomedical informatics* **132**, 104133 (2022)
- [19] Zheng, D., Song, X., Ma, C., Tan, Z., Ye, Z., Dong, J., Xiong, H., Zhang, Z., Karypis, G.: Dgl-ke: Training knowledge graph embeddings at scale. In: *Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval*, pp. 739–748 (2020)
- [20] Han, X., Cao, S., Xin, L., Lin, Y., Liu, Z., Sun, M., Li, J.O.: An open toolkit for knowledge embedding. *Proceedings of the EMNLP, Brussels, Belgium* **31**
- [21] Kingma, D.P., Ba, J.: Adam: A method for stochastic optimization. *ICLR* (2015)
- [22] Yang, B., Yih, W.-t., He, X., Gao, J., Deng, L.: Embedding entities and relations for learning and inference in knowledge bases. In: *International Conference on Learning Representations* (2015)

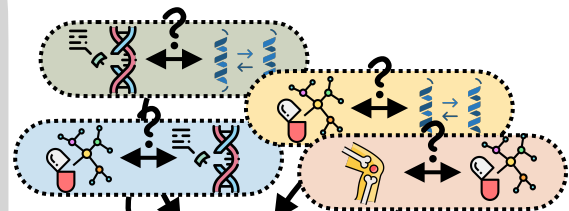
- [23] Dettmers, T., Minervini, P., Stenetorp, P., Riedel, S.: Convolutional 2d knowledge graph embeddings. In: Proceedings of the AAAI Conference on Artificial Intelligence, vol. 32 (2018)
- [24] Sun, Z., Deng, Z.-H., Nie, J.-Y., Tang, J.: Rotate: Knowledge graph embedding by relational rotation in complex space. arXiv preprint arXiv:1902.10197 (2019)
- [25] Vashishth, S., Sanyal, S., Nitin, V., Talukdar, P.: Composition-based multi-relational graph convolutional networks. arXiv preprint arXiv:1911.03082 (2019)
- [26] Das, R., Godbole, A., Zaheer, M., McCallum, A.: A simple approach to case-based reasoning in knowledge graphs. In: Automated Knowledge Base Construction (AKBC) (2020)
- [27] Das, R., Godbole, A., Monath, N., Zaheer, M., McCallum, A.: Probabilistic case-based reasoning for open-world knowledge graph completion. In: Findings of the Association for Computational Linguistics: EMNLP 2020 (2020). <https://aclanthology.org/2020.findings-emnlp.427.pdf>
- [28] Tu, R., Sinha, M., González, C., Hu, E., Dhuliawala, S., McCallum, A., Su, A.I.: Drug repurposing using consilience of knowledge graph completion methods. bioRxiv, 2023–05 (2024)
- [29] Zou, M., Shao, Z.: Proteome-wide mendelian randomization and colocalization analysis identify therapeutic targets for knee and hip osteoarthritis. *Biomolecules* **14**(3), 355 (2024)
- [30] Garfield, V., Salzmann, A., Burgess, S., Chaturvedi, N.: A guide for selection of genetic instruments in mendelian randomization studies of type 2 diabetes and hba1c: toward an integrated approach. *Diabetes* **72**(2), 175–183 (2023)
- [31] Consortium, G.: The gtex consortium atlas of genetic regulatory effects across human tissues. *Science* **369**(6509), 1318–1330 (2020)
- [32] Wang, G., Sarkar, A., Carbonetto, P., Stephens, M.: A simple new approach to variable selection in regression, with application to genetic fine mapping. *Journal of the Royal Statistical Society Series B: Statistical Methodology* **82**(5), 1273–1300 (2020)
- [33] Giambartolomei, C., Vukcevic, D., Schadt, E.E., Franke, L., Hingorani, A.D., Wallace, C., Plagnol, V.: Bayesian test for colocalisation between pairs of genetic association studies using summary statistics. *PLoS genetics* **10**(5), 1004383 (2014)
- [34] Sun, B.B., Chiou, J., Davies, N.M., Ritchie, S.C., Abraham, G., *et al.*: Plasma proteomic associations with genetics and health in the uk biobank. *Nature* **622**, 343–352 (2023) <https://doi.org/10.1038/s41586-023-06592-6>

- [35] Steinberg, J., Southam, L., Roumeliotis, T.I., Clark, M.J., Zeggini, E., *et al.*: A molecular quantitative trait locus map for osteoarthritis. *Nature Communications* **12**(1), 1309 (2021) <https://doi.org/10.1038/s41467-021-21593-7>
- [36] Burgess, S., Butterworth, A., Thompson, S.G.: Mendelian randomization analysis with multiple genetic variants using summarized data. *Genetic epidemiology* **37**(7), 658–665 (2013)
- [37] Bowden, J., Davey Smith, G., Haycock, P.C., Burgess, S.: Consistent estimation in mendelian randomization with some invalid instruments using a weighted median estimator. *Genetic epidemiology* **40**(4), 304–314 (2016)
- [38] Bowden, J., Davey Smith, G., Burgess, S.: Mendelian randomization with invalid instruments: effect estimation and bias detection through egger regression. *International journal of epidemiology* **44**(2), 512–525 (2015)
- [39] Verbanck, M., Chen, C.-Y., Neale, B., Do, R.: Detection of widespread horizontal pleiotropy in causal relationships inferred from mendelian randomization analyses. *Nature Genetics* **50**, 693–698 (2018) <https://doi.org/10.1038/s41588-018-0099-7>
- [40] Morrison, J., Knoblauch, N., Marcus, J.H., Stephens, M., He, X.: Mendelian randomization accounting for correlated and uncorrelated pleiotropic effects using genome-wide summary statistics. *Nature genetics* **52**(7), 740–747 (2020)
- [41] Sanderson, E., Davey Smith, G., Windmeijer, F., Bowden, J.: An examination of multivariable mendelian randomization in the single-sample and two-sample summary data settings. *International Journal of Epidemiology* **48**(3), 713–727 (2019) <https://doi.org/10.1093/ije/dyy262>
- [42] Wallace, C.: A more accurate method for colocalisation analysis allowing for multiple causal variants. *PLoS Genetics* **17**(9), 1009440 (2021) <https://doi.org/10.1371/journal.pgen.1009440>
- [43] Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M.A.R., Bender, D., Maller, J., Sklar, P., Bakker, P.I.W., Daly, M.J., Sham, P.C.: Plink: A tool set for whole-genome association and population-based linkage analyses. *American Journal of Human Genetics* **81**(3), 559–575 (2007) <https://doi.org/10.1086/519795>
- [44] Chang, C.C., Chow, C.C., Tellier, L.C.A.M., Vattikuti, S., Purcell, S.M., Lee, J.J.: Second-generation plink: rising to the challenge of larger and richer datasets. *GigaScience* (2015) <https://doi.org/10.1186/s13742-015-0047-8>
- [45] Zhu, R., Shen, Z., Zhu, H., Zhang, J., Xing, X., Wang, S., Fang, J.: Global, regional, and national burden and trends of soft tissue and other extraosseous sarcomas from 1990 to 2021. *Cancer Control* **32**, 10732748251355841 (2025)

- [46] Li, S., Cao, P., Chen, T., Ding, C.: Latest insights in disease-modifying osteoarthritis drugs development. *Therapeutic Advances in Musculoskeletal Disease* **15**, 1759720–231169839 (2023)
- [47] Burgess, S., Smith, G.D., Davies, N.M., Dudbridge, F., Gill, D., Glymour, M.M., Hartwig, F.P., Kutalik, Z., Holmes, M.V., Minelli, C., *et al.*: Guidelines for performing mendelian randomization investigations: update for summer 2023. *Wellcome open research* **4**, 186 (2023)
- [48] Sun, B.B., Maranville, J.C., Peters, J.E., Stacey, D., Staley, J.R., Blackshaw, J., Burgess, S., Jiang, T., Paige, E., Surendran, P., *et al.*: Genomic atlas of the human plasma proteome. *Nature* **558**(7708), 73–79 (2018)
- [49] Eldjarn, G.H., Ferkingstad, E., Lund, S.H., Helgason, H., Magnusson, O.T., Gunnarsdottir, K., Olafsdottir, T.A., Halldorsson, B.V., Olason, P.I., Zink, F., *et al.*: Large-scale plasma proteomics comparisons through genetics and disease associations. *Nature* **622**(7982), 348–358 (2023)
- [50] Kaiser, M., Saha Roy, R., Weikum, G.: Reinforcement learning from reformulations in conversational question answering over knowledge graphs. In: *Proceedings of the 44th International ACM SIGIR Conference on Research and Development in Information Retrieval*, pp. 459–469 (2021)
- [51] Xiong, W., Hoang, T., Wang, W.Y.: Deeppath: A reinforcement learning method for knowledge graph reasoning. *arXiv preprint arXiv:1707.06690* (2017)
- [52] Zhang, H., Wang, X., Zhao, Q.: Granularity-aware contrastive learning for fine-grained action recognition. In: *ICASSP 2025-2025 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 1–5 (2025). IEEE
- [53] Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., Dehghani, M., Minderer, M., Heigold, G., Gelly, S., Uszkoreit, J., Houlsby, N.: An image is worth 16x16 words: Transformers for image recognition at scale. In: *ICLR* (2021)
- [54] Ba, J.L., Kiros, J.R., Hinton, G.E.: Layer normalization. *NeurIPS* (2016)



Unknown relations:



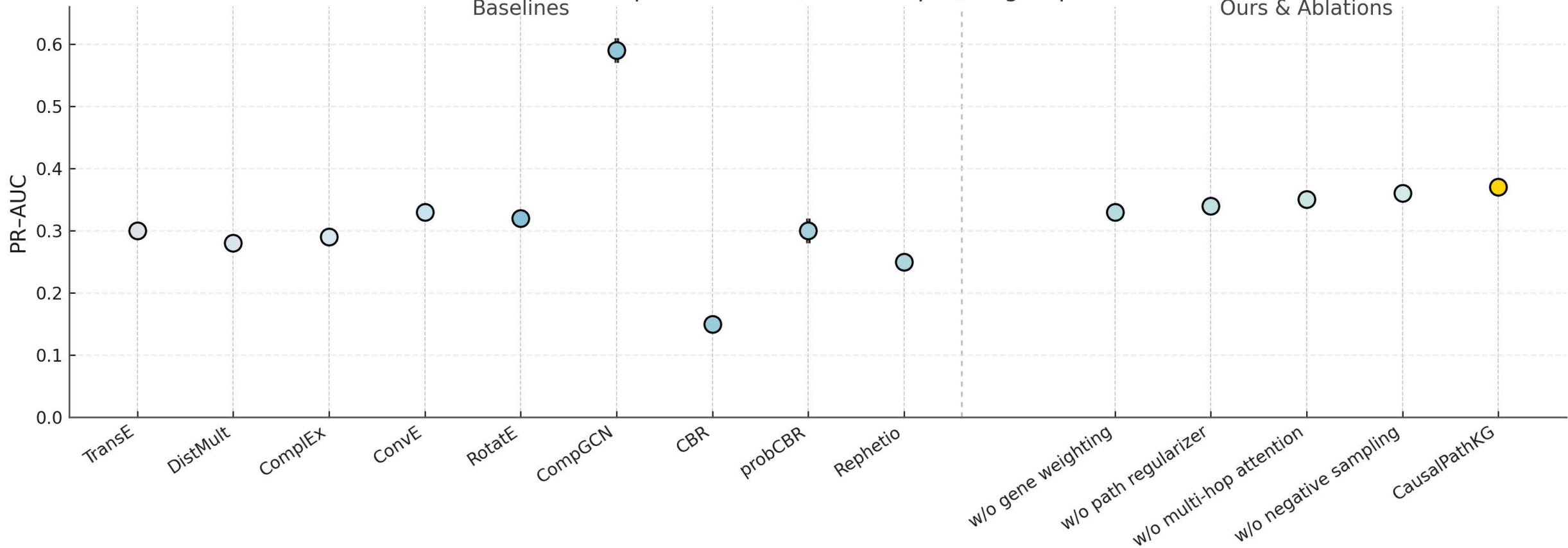
Train

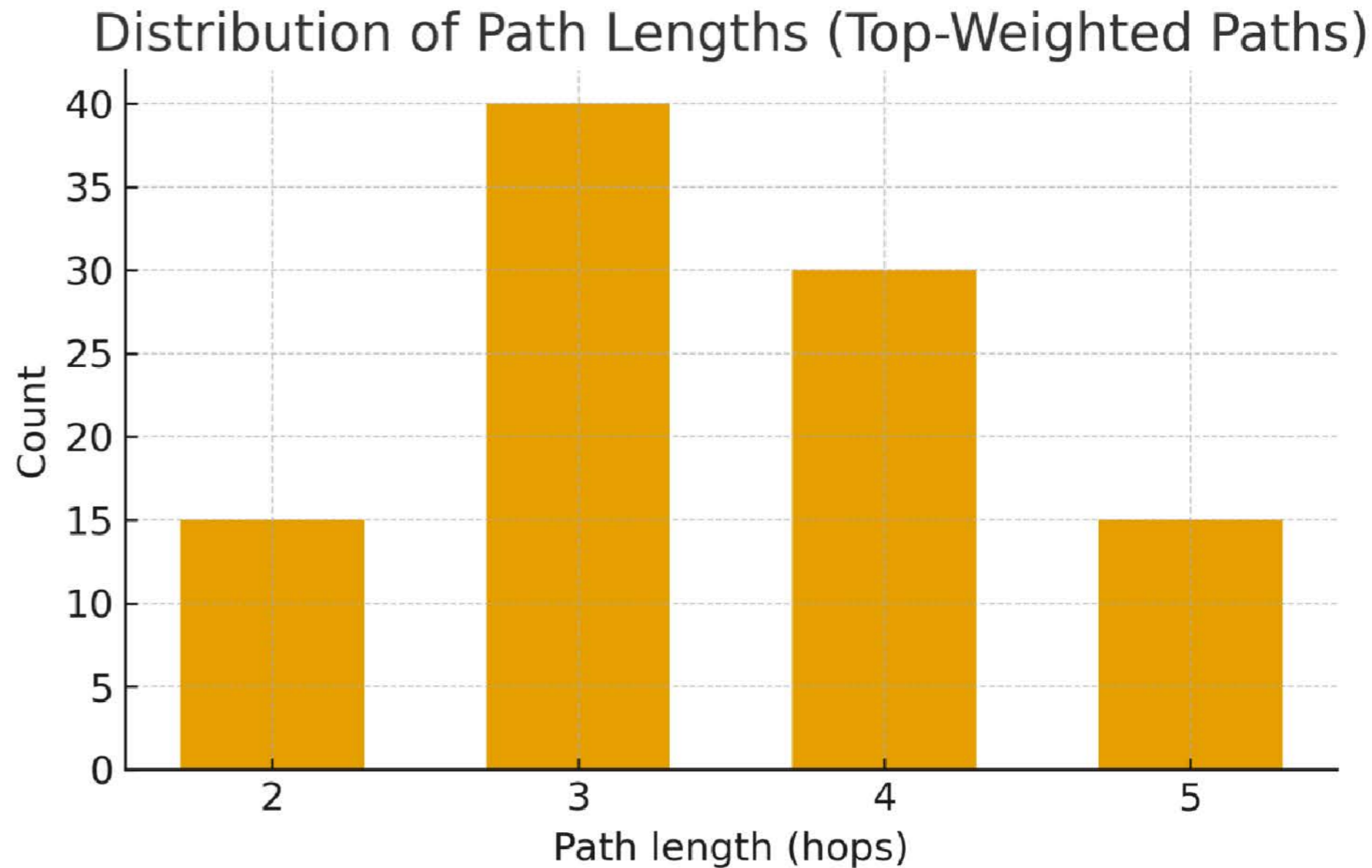
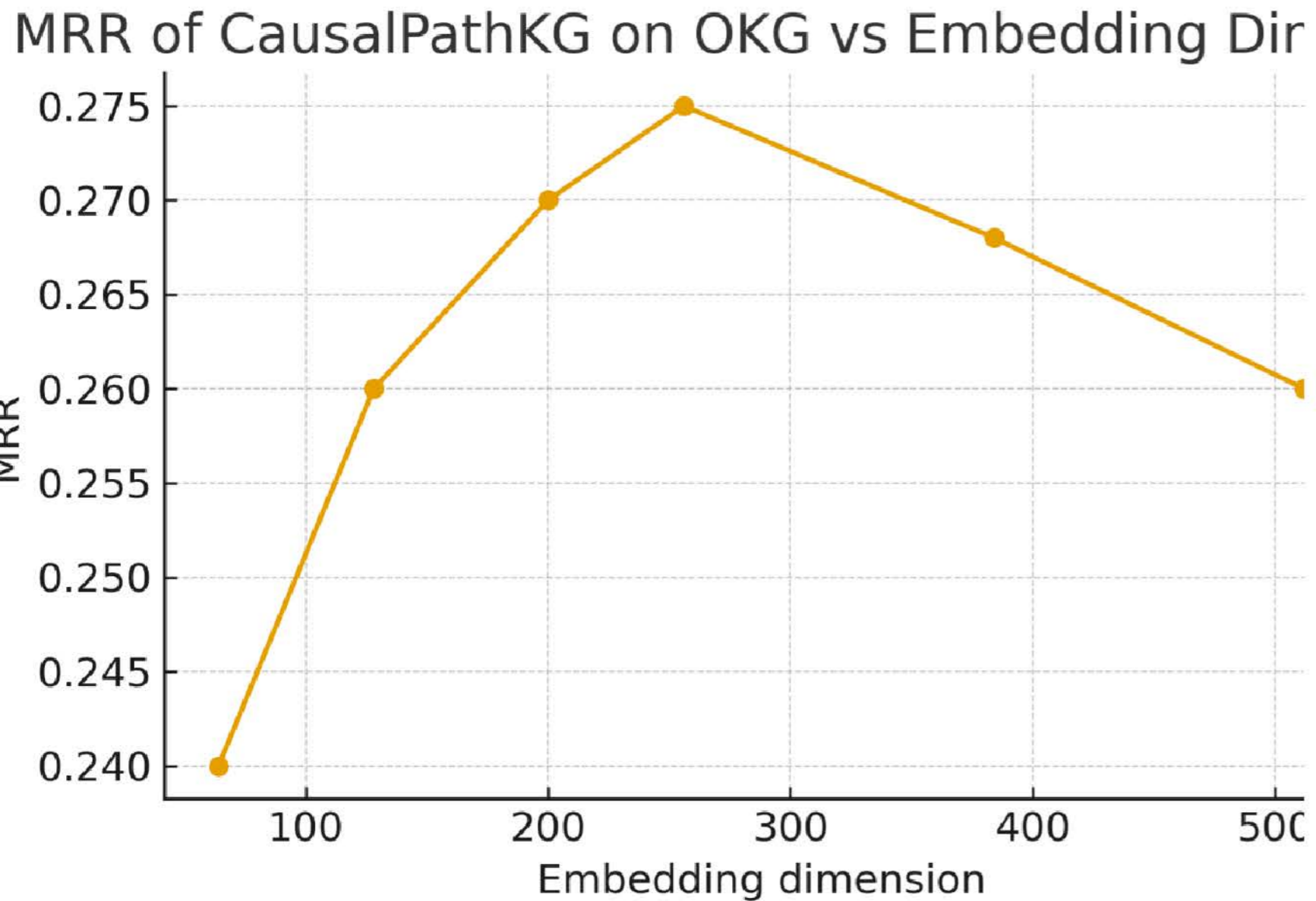
**CausalPathKG**

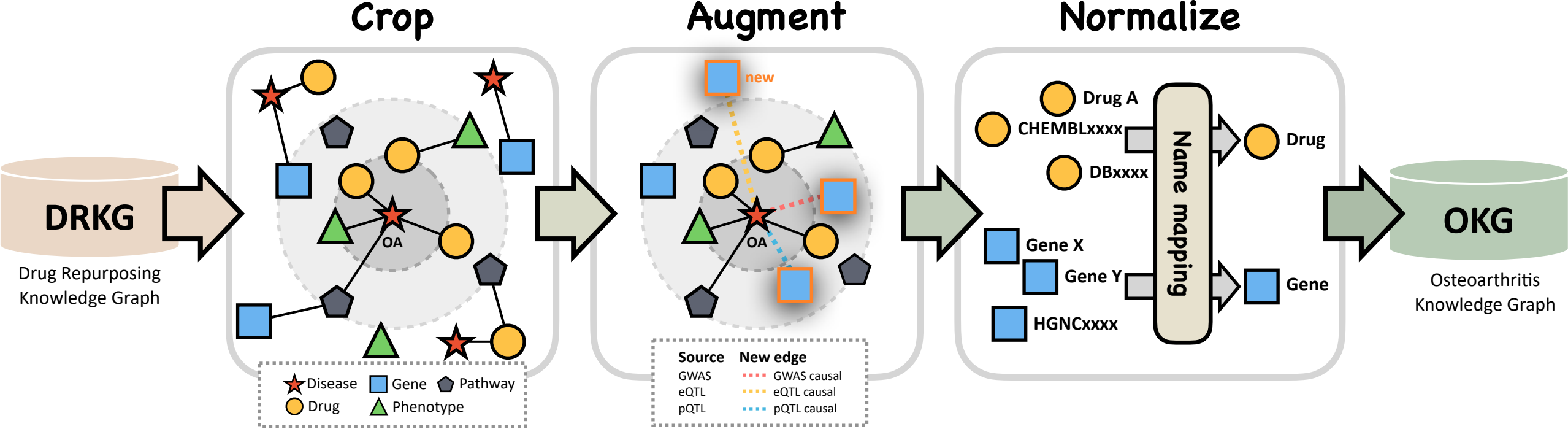
- investigational for
- inhibits
- participates in
- associated with
- .....

PR-AUC performance (random split) — grouped

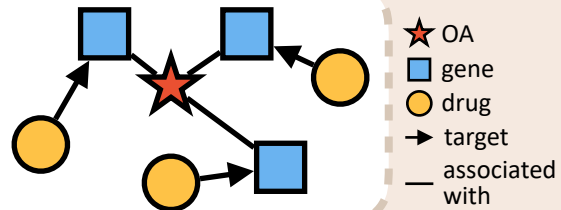
Ours & Ablations



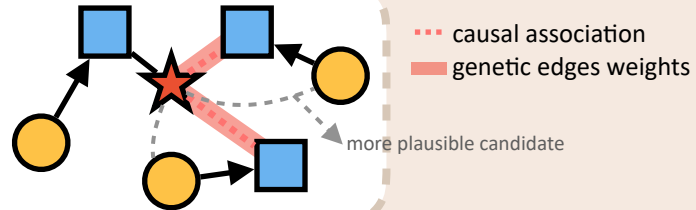




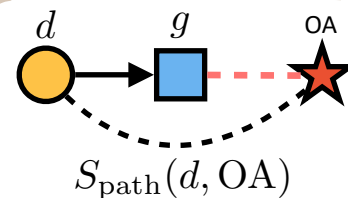
### (a) Base RotatE



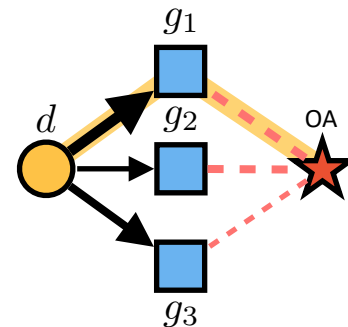
### (b) Weighted Genetic Edge



### (c) Effector Gene Path Regularization



How to score drug-OA pairs?



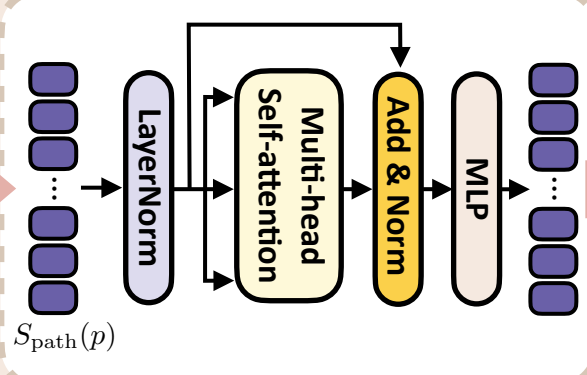
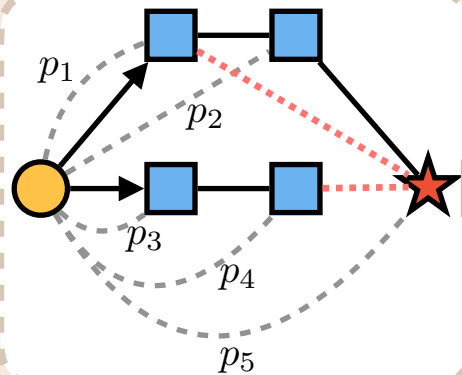
Aggregate all 2-hop paths

$d \rightarrow g \rightarrow \text{OA}$

and take the **strongest**

max

### (d) Multi-hop Path Attention



Aggregation

### (e) Training Strategy

