

SPARQL検索・LLM解釈による候補集合との照合に基づく
KGEマウスモデル候補の後段評価
— LLM統合候補評価レイヤーによる確認優先度付け —

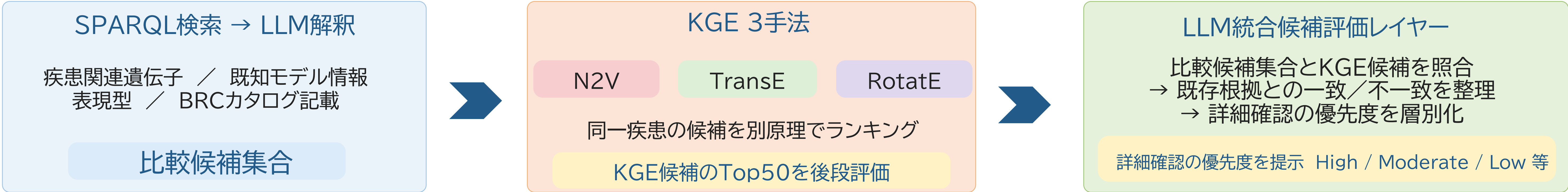
櫛田達矢・臼田大輝・高田豊行・柁屋啓志(理研バイオリソース研究センター)

要旨：疾患モデルマウス探索において、SPARQL検索結果をLLMで解釈・整理した比較候補集合と、Node2Vec・TransE・RotatEによるKGE候補を照合した。既知根拠との一致や候補順位の違いを後段評価し、確認優先度を付与するLLM統合候補評価レイヤーを検討した。2型糖尿病では、N2Vは既存候補の再発見、TransEは別軸の順位づけに有用であった。

研究の問い

知識グラフ埋め込み(KGE)は知識グラフのどの「つながり」を候補順位に反映し、どのようなマウスを上位化するのか？
→ SPARQL検索結果をLLMで解釈・整理して得た候補集合と照合し、根拠と確認優先度を後段評価する。

全体設計:SPARQL検索結果のLLM解釈とKGEを並列に作り、後段で照合



1. 3手法の違い

N2V	疾患ノードと候補マウスノードの cosine similarity。関係種別を明示せず、グラフ全体の近接性を反映。
TransE	query-relation-mouse の score_hrt.GENE/PHENO は rank fusion。
RotatE	同じく relation-specific score。2型糖尿病では score_hrt を確認できたのは 7/50。 score_hrt = query-relation-mouse 三つ組に対するKGEスコア

2. 複数疾患で候補集合は大きく異なる

