A181 · Causal Path Inference on a Literature-Derived Knowledge Graph for Variant Effect Interpretation

Presenter: Jici Jiang — Northeastern University

Authors: Jici Jiang

Session: Day 1 · Thu Oct 1, 2026 · 2:15–4:15 PM

Current variant effect predictors integrate genomic annotation and protein-level features, however, are limited to modeling direct variant-to-gene or gene-to-phenotype mappings without explicitly reasoning over the mechanistic chain. We present a framework to generate interpretable, multi-hop causal paths that link variants to phenotypes through intermediate signaling, protein-interaction and regulatory events, corresponding to curated variant-phenotype evidence. Our INDRA system builds the knowledge graph, integrating causal mechanism extraction from literature with pathway databases. Variants are encoded as reference and alternate embedding pairs from sequence models. Training and evaluation use 120,329 variants across 11,786 genes.

We formulate variant-effect prediction as stepwise path generation over a variant-relevant subgraph from constrained pathfinding. A variant-conditioned message-passing network initializes traversal from the associated gene, and a sequential decoder scores candidate next nodes and edge polarities at each hop. On held-out variants, paths reach the correct phenotype in 79% of cases with Hit@10 of 0.68, producing paths for mechanistic interpretation and downstream functional investigation.

Evaluation of paths is itself a challenge since no existing metric captures partial correctness of biomedical causal paths. General path comparison methods disregard biological semantics and regulatory direction which undervalues predictions that are mechanistically correct but structurally divergent. To address this problem, we introduce a dynamic-programming algorithm to score ontology-grounded semantic similarity and regulatory-direction consistency for biomedical causal paths. To evaluate this metric, we generated a corpus of ground truth causal paths paired with partially correct paths of graded divergence and compared our new approach with those currently used in the field, including LLM-as-a-judge scenarios.