
BioMedReasoner: Towards Multi-Hop Reasoning using Path-based Relational Learning on Biomedical Knowledge Graphs

Ahmad Wisnu Mulyadi, Lilija Wehling, Ansh Kumar, Gurdeep Singh*

Virtual Patient Engine, BioMed X GmbH, Heidelberg, Germany

{mulyadi, wehling, kumar, gsingh}@bmedx.com

Nicolas Boucher¹, Firas Abdessalem¹, Sven Jager²,

Mohammed H. Mosa³, Thomas Klabunde⁴, Tommaso Andreani^{4*}

¹Digital R&D, Accelerator, Sanofi Digital, Paris, France

²Digital R&D, Data & Computational Science, Sanofi Digital, Frankfurt am Main, Germany

³Therapeutic Area Type 1/17 Immunology & Inflammation, Sanofi R&D, Frankfurt am Main, Germany

⁴Translational Medicine Unit (TMU) - Disease Modeling, Sanofi R&D, Frankfurt am Main, Germany

{Nicolas.Boucher, Firas.Abdessalem, Sven.Jager, Mohammed.Mosa, Thomas.Klabunde, Tommaso.Andreani}@sanofi.com

Abstract

Scientific discovery of novel drug assets requires rigorous reasoning over the complex interplay of biomedical entities. By integrating multi-modal entities, biomedical knowledge graphs (KGs) have emerged as powerful resources for tasks such as identifying targets and uncovering mechanisms of drug action. However, effective reasoning over those KGs is often limited by their incompleteness and noisy information, hindering reliable downstream tasks. To address this challenge, we propose BioMedReasoner, a modular reasoning framework that performs KG completion as a precursor to interpretable multi-hop reasoning. Our approach builds on Neural Bellman-Ford Networks (NBFNet) that formulates link prediction as a path-based relational learning problem, enabling interpretable multi-hop predictions for biomedical applications. We evaluated BioMedReasoner on PrimeKG, a publicly available large-scale multi-modal biomedical KG, and demonstrated its effectiveness in KG completion and downstream reasoning. As a proof of concept, we focus on genes associated with inflammatory bowel disease (IBD), including the JAK family, TNF- α , IL17A, and IL17RA. By extracting paths connecting the genes to IBD and to their biological processes and molecular functions, and supplying these paths to a large language model (LLM) while withholding gene names, we show that interpretable graph paths enable biological reasoning and assessment of gene suitability as therapeutic targets. These results highlight the potential of path-based, interpretable biomedical KG reasoning to guide early-stage drug discovery.

*Correspondence: gsingh@bmedx.com; Tommaso.Andreani@sanofi.com

1 Introduction

Rapid advances in artificial intelligence have impacted various scientific fields, including the discovery of novel drug assets that require rigorous reasoning over the complex interplay of biomedical entities. To capture these interconnections across multi-modal data, biomedical knowledge graphs (KGs) have emerged as powerful resources. Constructed from large-scale biomedical corpora as tuples of factual information [5], KGs enable the modelling of complex biological processes and support a wide range of downstream applications [32], including drug discovery [43], drug combination effects [19], and precision therapeutics [14]. Despite their promise, extracting meaningful and interpretable insights from biomedical KGs remains a challenge, as it requires both completeness and robustness to noise. In practice, these criteria are barely satisfied, particularly in evolving KGs, which are prone to conflicting or imprecise knowledge [31, 40]. Consequently, KG completion has become an indispensable step for enhancing KG representations and enabling reliable reasoning.

KG completion is a fundamental task in relational learning aimed at predicting links between entities, thereby recovering latent information and uncovering novel facts [3, 37, 42]. A broad range of approaches have been proposed, leveraging geometric, algebraic, and analytical perspectives [4]. Inspired by algebraic formulations, numerous methods operate in predefined vector spaces to enable structural learning over large-scale relational data, often instantiated through variants of graph neural networks (GNNs) [30, 47, 13]. Such methods have achieved promising results in pharmacological tasks, including drug repurposing [17, 18] and synthetic lethality prediction [46, 6].

Recent efforts have explored enhancing KG reasoning by unifying relational learning with large language models (LLMs). For example, BioBridge [39] introduced a bridge module that aligns heterogeneous node modalities to support the search for cross-modal similarity, which can then serve as a contextual input for LLM prompts. Although effective for multi-modal integration, this approach offers limited interpretability with respect to graph elements (*e.g.*, paths, subgraphs), diminishing its utility for in-depth analysis. G-Retriever [16] further advanced this direction by contextualizing graphs for LLMs through retrieval-augmented generation (RAG), extracting subgraphs relevant to a user prompt. However, its applicability has not yet been demonstrated in biomedical domains, where deep reasoning is required to address complex and interdependent entities.

In this preliminary work, we propose BioMedReasoner (Fig. 1), a modular framework for interpretable multi-hop biomedical reasoning. We formulate multi-hop reasoning as path-driven relational learning using Neural Bellman-Ford Networks (NBFNet) [47]. This design enables (i) KG completion on biomedical KGs, (ii) extraction of path-level interpretations, and (iii) enhanced reasoning by integrating large language models with contextualized paths via retrieval-augmented generation (RAG). We evaluated our framework on PrimeKG [5], a publicly available large-scale biomedical KG encompassing diverse biomedical entities and relationships. As a proof of concept, we focus on inflammatory bowel disease (IBD), analyzing key genes such as the JAK family, TNF- α , IL17A, and IL17RA. Our framework identifies association paths linking these genes to IBD and their biological processes and molecular functions, which are further provided as auxiliary context to an LLM with the gene name withheld, enabling reasoning on questions such as gene suitability as therapeutic targets. By providing interpretable paths, BioMedReasoner allows pharmaceutical researchers to understand the mechanistic rationale behind predicted associations, prioritize targets with higher confidence, and generate biologically grounded hypotheses that can guide early-stage drug discovery.

2 Proposed Method

KG Completion We represent a biomedical KG as $\mathcal{G} = (\mathcal{V}, \mathcal{E}, \mathcal{R})$, where nodes $u \in \mathcal{V}$ are connected through edges $(u, r, v) \in \mathcal{E}$, each labeled with a relation type $r \in \mathcal{R}$. Following NBFNet [47], we address link prediction by estimating the probability $p(u, q, v)$, which denotes the likelihood of a triple connecting a source entity u and a target entity v under a query relation q . We further denote the corresponding representation as $\mathbf{h}_q(u, v)$, initialized as

$$\mathbf{h}_q^{\ell=0}(u, v) = f_{\mathcal{I}}(u, q, v), \quad \forall v \in \mathcal{V}, \quad (1)$$

where ℓ denotes the current layer index, and $f_{\mathcal{I}}$ is an indicator function $\mathbb{1}(u = v) * \mathbf{q}$ that initializes the node features of u with the relation embedding \mathbf{q} . These features are then iteratively updated

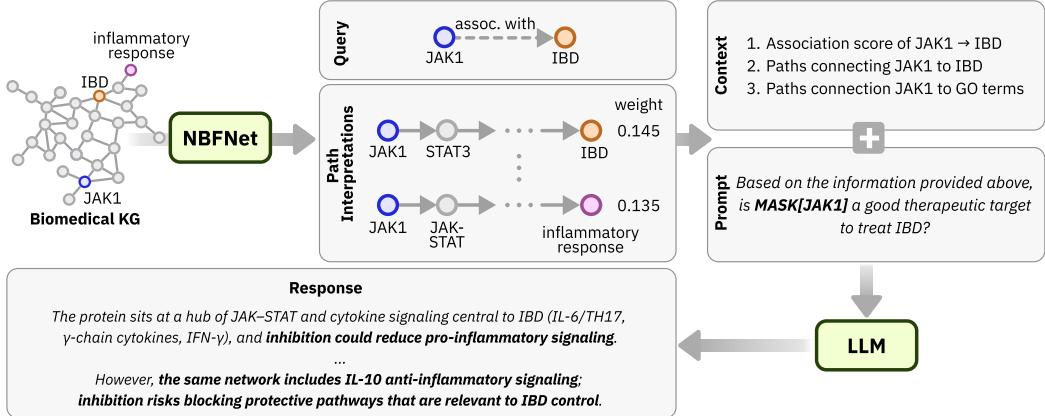


Figure 1: Overview of BioMedReasoner in addressing multi-hop reasoning over biomedical KG.

across successive layers through message-passing operations, as defined in Eqs. (2)–(3).

$$\mathbf{m}_q^\ell(x, r, v) = f_{\mathcal{M}}(\mathbf{h}_q^{\ell-1}(x, v); \mathbf{e}_q(x, r, v)), \quad \forall (x, r, v) \in \mathcal{E}(v) \quad (2)$$

$$\mathbf{h}_q^\ell(u, v) = f_{\mathcal{A}}(\{\mathbf{m}_q^\ell(x, r, v)\} \cup \{\mathbf{h}_q^{\ell=0}(u, v)\}) \quad (3)$$

Here, $f_{\mathcal{M}}$ denotes the message function, which maps two inputs—*i.e.*, the intermediate pair features $\mathbf{h}_q^{\ell-1}(x, v)$ and the associated edge features $\mathbf{e}_q(x, r, v)$ —into an embedding space using a chosen relation operator [3, 42, 34], producing the message $\mathbf{m}_q^\ell(x, r, v)$. The final representations are then further aggregated via a permutation-invariant aggregation function $f_{\mathcal{A}}$ [7].

To address the link prediction task, we estimate the conditional probability $p(v|u, q)$ using Eq.(4), where f_{θ} is a feed-forward network followed by a sigmoid activation σ . All model parameters are then optimized by minimizing the negative log-likelihood over positive and negative samples, as formalized in Eq.(5).

$$p(v|u, q) = \sigma(f_{\theta}(\mathbf{h}_q^\ell(u, v))) \quad (4)$$

$$\mathcal{L} = -\log p(u, q, v) - \sum_{i=1}^n \frac{1}{n} \log(1 - p(u'_i, q, v'_i)) \quad (5)$$

Multi-hop Path-based RAG To enhance our model’s reasoning capabilities and generate interpretable insights that span multiple hops from the source node, we adopt a path-based RAG approach. Following NBFNet [47], we define $f_{\mathcal{P}}$ as a path visualization function to identify the paths contributing to the prediction of a triple $p(u, q, v)$. This is achieved by estimating the importance of each path—computed via automatic differentiation—and subsequently applying beam search to select the top- K most relevant paths. Formally, the multi-hop path formulation is defined as:

$$\{\Phi_k^{u \rightarrow v}, \omega_k\}_{k=1}^K = f_{\mathcal{P}}(u, q, v), \quad (6)$$

where $\Phi_k^{u \rightarrow v}$ represents the k -th path and ω_k as its corresponding weight. We removed the paths containing a reverse triple of any of their constituent triples. In addition, for each triple, only the path with the highest weight was retained, while all other paths were excluded from subsequent analyses. Next, we used a frozen LLM as the downstream reasoning module within our framework, adopting RAG to infer responses y , as shown in Eq. (7). In particular, along with a prompt, we provided a set of biomedical auxiliary context, including: (i) an association score $\alpha^{G \rightarrow D}$ between a gene and a disease, computed from its predicted reciprocal rank (*i.e.*, the inverse of the rank position assigned to the correct target entity, giving higher scores when the true entity is ranked closer to the top); (ii) paths $\Phi^{G \rightarrow D}$ connecting the gene to the disease; and (iii) paths $\Phi^{G \rightarrow BP}$, $\Phi^{G \rightarrow MF}$ connecting the gene to biological processes and molecular functions, respectively.

$$y = \text{LLM}(\text{prompt}; \alpha^{G \rightarrow D}, \Phi^{G \rightarrow D}, \Phi^{G \rightarrow BP}, \Phi^{G \rightarrow MF}). \quad (7)$$

3 Experimental Results

Dataset and Data Preprocessing To evaluate the effectiveness of our proposed framework, we utilized PrimeKG [5] as the publicly available heterogeneous biomedical KG commonly used in the literature. Following the dataset filtering procedure of BioBridge [39], we selected heterogeneous nodes including genes/proteins, drugs, diseases, biological processes, molecular functions, and cellular components. This results in 84,981 nodes and 3,904,610 triples. The dataset is further split into training, validation, and test sets with 80:10:10 ratio. For specific biomedical case studies on IBD, we hold out triples connecting selected well-studied genes (*e.g.*, JAK1, JAK2, JAK3, TYK2, TNF- α , IL17A, IL17RA) to IBD and its subtypes according to the MONDO ontology². We also hold out the drugs that target these genes and indicated in IBD and its subtypes to avoid any information leakage.

Implementation Details We followed overall NBFNet [47] configuration as our backbone for our proposed BioMedReasoner in performing KG completion. Specifically, we employed 6 layers of message-passing operations, each with 8 hidden dimensions. The message function f_M was implemented using DistMult [42], and the aggregation function f_A used principal neighborhood aggregation (PNA) [7]. The feed-forward network f_θ was fixed at 16 hidden dimensions with ReLU activations. All model parameters were optimized using the Adam optimizer with a learning rate of 0.005 for 5 epochs. Training and inference were conducted on 8 NVIDIA HGX H100 GPUs, each with 80GB memory. For reasoning with LLMs, we employed OpenAI GPT-5 [44] and provide the detailed system prompt used for the reasoning case studies in Appendix B.

KG Completion with BioMedReasoner We evaluated BioMedReasoner on PrimeKG to assess its ability to complete missing links. The model demonstrates strong performance across standard evaluation metrics (*e.g.*, mean rank (MR), mean reciprocal rank (MRR), and Hits@K with $K \in \{1, 3, 10\}$), as shown in Fig. 2. These results underscore the effectiveness of path-based relational reasoning in addressing KG incompleteness, surpassing traditional knowledge graph embedding models such as TransE [3] and DistMult [42], and further validate our choice of NBFNet as the backbone model, which has been shown to outperform a range of conventional KG embedding approaches in prior work [47]. We also provide a granular breakdown of BioMedReasoner’s performance in predicting links between genes and other node types existed in the dataset, including diseases, molecular functions, biological processes, and cellular components (see Table 2 in Appendix A).

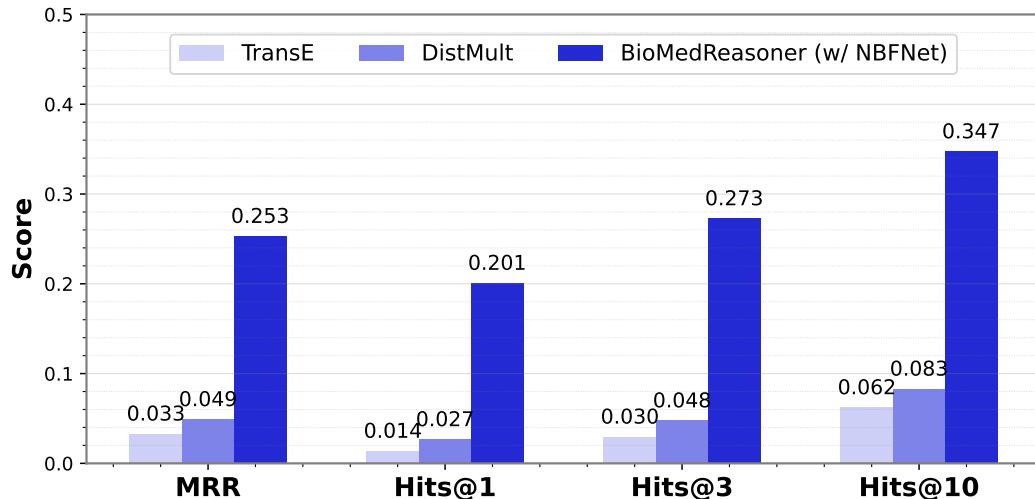


Figure 2: KG completion performances of BioMedReasoner that utilized NBFNet as backbone on testing set of PrimeKG compared againsts TransE and DistMult.

²http://purl.obolibrary.org/obo/MONDO_0005265

Case Study: Genes Associated with IBD As a proof of concept, we applied BioMedReasoner to predict association of selected genes with IBD and present an association heatmap, as presented in Fig. 3 in Appendix A. TNF- α exhibits the strongest predicted association with IBD, consistent with extensive clinical evidence [45, 36]. Members of the JAK family rank next, aligning with their well-established therapeutic relevance [29, 10, 8]. Compared to TNF- α and the JAK family, IL17A and IL17RA show weaker associations, in line with clinical findings regarding their involvement in IBD [20, 35]. Finally, we also include LRRK1, a gene primarily associated with rare bone disorders rather than IBD [28]. These predictions support the biological plausibility of BioMedReasoner’s outputs and highlight its capability to prioritize therapeutically relevant targets to some extent.

Table 1: Case studies of path-based reasoning of BioMedReasoner related to JAK1 and IL17A genes.

| Case Study: JAK1 | |
|---------------------------|--|
| Context | <ul style="list-style-type: none"> • Association Assessment: Strong <p>• Paths JAK1 → IBD: $\langle \text{PROTEIN, ppi, PTPN2} \rangle \rightarrow \langle \text{PTPN2, associated with, IBD} \rangle$ $\langle \text{PROTEIN, ppi, JAK3} \rangle \rightarrow \langle \text{JAK3, ppi, PTPN2} \rangle \rightarrow \langle \text{PTPN2, associated with, IBD} \rangle$</p> |
| Context Reference | Ferrao <i>et al.</i> , [11]; Lees <i>et al.</i> , [24]; Sugamura <i>et al.</i> , [33]; Habib <i>et al.</i> , [15] |
| Response | <p>[...] Advantages: Dense network proximity to an IBD-associated node Inhibition could dampen overactive cytokine/JAK-STAT signaling, Connections to protein phosphorylation and kinase activity suggest it sits at a signaling control point where inhibition may yield broad anti-inflammatory effects. [...]</p> <p>[...] Disadvantages: Participation in interleukin-10-mediated signaling indicates potential to disrupt regulatory cytokine signaling, risking immune dysregulation if inhibited. [...]</p> |
| Response Reference | Pérez-Jeldres <i>et al.</i> , [27]; Pattinson <i>et al.</i> , [26] |
| Case Study: IL17A | |
| Context | <ul style="list-style-type: none"> • Association Assessment: Weak <p>• Paths IL17A → IBD: $\langle \text{PROTEIN, ppi}^{-1}, \text{TRAF6} \rangle \rightarrow \langle \text{TRAF6, ppi, RIPK2} \rangle \rightarrow \langle \text{RIPK2, associated with}^{-1}, \text{IBD} \rangle$</p> |
| Context Reference | Yan <i>et al.</i> , [41]; Lee <i>et al.</i> , [23]; Datta <i>et al.</i> , [9]; Wang <i>et al.</i> , [38] |
| Response | <p>[...] Advantages: The protein connects to IL-17/IL-17A-mediated and broader cytokine-mediated signaling [...]</p> <p>[...] Disadvantages: Links to epithelial/keratinocyte proliferation suggest a role in epithelial biology; inhibition may hinder mucosal repair. [...]</p> |
| Response Reference | Kuwabara <i>et al.</i> , [22]; Gaffen [12]; Lee <i>et al.</i> , [23] |

Case Study: Path Interpretability and Biological Context Beyond link predictions, BioMedReasoner extracted interpretable multi-hop paths linking genes (JAK1 and IL17A) to IBD, and to their known biological processes and molecular functions (Table 1) as defined in the Gene Ontology (GO) resource [2, 1]. Incorporating the GO terms provides biological and functional context, enabling mechanistic interpretation of the predicted associations and supporting reasoning about why a gene may influence disease. Moreover, by supplying these paths to a large language model (GPT5) while withholding gene names, we facilitated biologically informed reasoning regarding gene target suitability. For detailed LLM responses along with the input context, please refer to the dedicated case study subsections in Appendices C and D.

4 Limitations and future Works

While our results highlight the potential of BioMedReasoner for interpretable biomedical reasoning, several limitations remain. First, the current knowledge graph lacks important node types that could enhance interpretability and biological relevance. For example, the clinical failure of anti-IL17A therapies in inflammatory bowel disease is closely linked to Th17 cells and $\gamma\delta$ T cells [23], and incorporating such cell-specific knowledge would provide a richer mechanistic basis for reasoning.

Second, our current implementation of NBFNet relies exclusively on graph structure and does not leverage rich node or edge features. Integrating representations from biological foundation models, such as AlphaFold2 [21] or ESM [25] for protein structure and function, together with drug–target binding information, could enable more mechanistic reasoning. Such integration may help explain clinically observed differences, such as why Etanercept (targeting TNF- α) is ineffective in IBD, whereas Adalimumab and Infliximab (also targeting TNF- α) demonstrate efficacy.

Third, while LLMs can reason over contextualized paths, their outputs remain sensitive to prompt design and underlying model biases, raising challenges for robustness and reproducibility. Addressing these limitations may require improved prompting strategies, alignment techniques, and systematic evaluation of LLM-based reasoning.

Finally, our evaluation was limited to single-disease case studies. Broader validation across multiple disease domains will be necessary to establish generalization. Future directions include enhanced path-ranking strategies, multi-modal data integration, and rigorous benchmarking to strengthen both the interpretability and predictive reliability of biomedical reasoning frameworks.

5 Conclusion

In this preliminary work, we introduced BioMedReasoner, a modular framework for interpretable multi-hop reasoning over biomedical knowledge graphs. By integrating Neural Bellman-Ford Networks for path-driven relational learning with retrieval-augmented generation using large language models, our approach enables both knowledge graph completion and extraction of path-level explanations. We applied our framework to PrimeKG through case studies on IBD, where BioMedReasoner predicted strong associations for TNF- α , followed by JAK family members, whereas IL17A and IL17RA exhibited weaker links, consistent with evidence reported in the literature. We further extracted interpretable paths connecting these genes to IBD and to their Gene Ontology biological processes, and provided these paths as context to an LLM to evaluate target suitability. These results demonstrate how path-based reasoning can generate explanatory narratives that bridge structured biomedical knowledge with natural language reasoning. Beyond IBD, the modular design of BioMedReasoner makes it readily adaptable to diverse biomedical applications, including disease comorbidity prediction, biomarker discovery, and guiding precision medicine strategies, while offering interpretable insights that can increase confidence and translational value in pharmaceutical research.

Acknowledgments

We acknowledge Henrik Cordes from Sanofi as well as Thomas Rückle from BioMed X GmbH for collaborative in-depth discussion in developing BioMedReasoner. We thank Vultr³ for graciously providing free credits for training BioMedReasoner.

³<https://www.vultr.com/>

Funding

This study was funded by Sanofi.

Conflict Interests

NB, FA, SJ, MHM, TK, and TA are Sanofi employees and may hold shares and/or stock options in the company.

References

- [1] Suzi A Aleksander, James Balhoff, Seth Carbon, J Michael Cherry, Harold J Drabkin, Dustin Ebert, Marc Feuermann, Pascale Gaudet, Nomi L Harris, et al. The gene ontology knowledgebase in 2023. *Genetics*, 224(1), 2023.
- [2] Michael Ashburner, Catherine A Ball, Judith A Blake, David Botstein, Heather Butler, J Michael Cherry, Allan P Davis, Kara Dolinski, Selina S Dwight, Janan T Eppig, et al. Gene ontology: tool for the unification of biology. *Nature genetics*, 25(1):25–29, 2000.
- [3] Antoine Bordes, Nicolas Usunier, Alberto Garcia-Duran, Jason Weston, and Oksana Yakhnenko. Translating embeddings for modeling multi-relational data. In C.J. Burges, L. Bottou, M. Welling, Z. Ghahramani, and K.Q. Weinberger, editors, *Advances in Neural Information Processing Systems*, volume 26. Curran Associates, Inc., 2013.
- [4] Jiahang Cao, Jinyuan Fang, Zaiqiao Meng, and Shangsong Liang. Knowledge graph embedding: A survey from the perspective of representation spaces. *ACM Computing Surveys*, 56(6):1–42, 2024.
- [5] Payal Chandak, Kexin Huang, and Marinka Zitnik. Building a knowledge graph to enable precision medicine. *Scientific Data*, 10(1):67, 2023.
- [6] Xuexin Chen, Ruichu Cai, Zhengting Huang, Zijian Li, Jie Zheng, and Min Wu. Interpretable high-order knowledge graph neural network for predicting synthetic lethality in human cancers. *Briefings in Bioinformatics*, 26(2):bbaf142, 04 2025.
- [7] Gabriele Corso, Luca Cavalleri, Dominique Beaini, Pietro Liò, and Petar Veličković. Principal Neighbourhood Aggregation for Graph Nets. *Advances in Neural Information Processing Systems*, 33:13260–13271, 2020.
- [8] Silvio Danese, Séverine Vermeire, Wen Zhou, Aileen L Pangan, Jesse Siffledeen, Susan Greenbloom, Xavier Hébuterne, Geert D’Haens, Hiroshi Nakase, Julian Panés, et al. Upadacitinib as induction and maintenance therapy for moderately to severely active ulcerative colitis: results from three phase 3, multicentre, double-blind, randomised trials. *The lancet*, 399(10341):2113–2128, 2022.
- [9] Shyamasree Datta, Michael Novotny, Paul G Pavicic, Chenyang Zhao, Tomasz Herjan, Justin Hartupee, and Thomas Hamilton. IL-17 regulates CXCL1 mRNA stability via an AUUUA/tristetraprolin-independent sequence. *The journal of immunology*, 184(3):1484–1491, 2010.
- [10] Brian G Feagan, Silvio Danese, Edward V Loftus, Séverine Vermeire, Stefan Schreiber, Timothy Ritter, Ronald Fogel, Rajiv Mehta, Sandeep Nijhawan, Radosław Kempinski, et al. Filgotinib as induction and maintenance therapy for ulcerative colitis (SELECTION): a phase 2b/3 double-blind, randomised, placebo-controlled trial. *The Lancet*, 397(10292):2372–2384, 2021.
- [11] Ryan Ferrao, Heidi JA Wallweber, Hoangdung Ho, Christine Tam, Yvonne Franke, John Quinn, and Patrick J Lupardus. The structural basis for class II cytokine receptor recognition by JAK1. *Structure*, 24(6):897–905, 2016.
- [12] Sarah L Gaffen. Structure and signalling in the IL-17 receptor family. *Nature Reviews Immunology*, 9(8):556–567, 2009.
- [13] Mikhail Galkin, Xinyu Yuan, Hesham Mostafa, Jian Tang, and Zhaocheng Zhu. Towards Foundation Models for Knowledge Graph Reasoning. In *The Twelfth International Conference on Learning Representations*, 2024.
- [14] Shanghua Gao, Richard Zhu, Zhenglun Kong, Ayush Noori, Xiaorui Su, Curtis Ginder, Theodoros Tsilgkaridis, and Marinka Zitnik. TxAgent: An AI Agent for Therapeutic Reasoning Across a Universe of Tools. *arXiv preprint arXiv:2503.10970*, 2025.

[15] Tania Habib, Shantha Senadheera, Kenneth Weinberg, and Kenneth Kaushansky. The common γ chain (γc) is a required signaling component of the IL-21 receptor and supports IL-21-induced cell proliferation via JAK3. *Biochemistry*, 41(27):8725–8731, 2002.

[16] Xiaoxin He, Yijun Tian, Yifei Sun, Nitesh Chawla, Thomas Laurent, Yann LeCun, Xavier Bresson, and Bryan Hooi. G-Retriever: Retrieval-augmented generation for textual graph understanding and question answering. *Advances in Neural Information Processing Systems*, 37:132876–132907, 2024.

[17] Yue Hu, Svitlana Oleshko, Samuele Firmani, Zhaocheng Zhu, Hui Cheng, Maria Ulmer, Matthias Arnold, Maria Colomé-Tatché, Jian Tang, Sophie Xhonneux, et al. BioPathNet: Enhancing Link Prediction in Biomedical Knowledge Graphs through Path Representation Learning. *Research Square*, pages rs–3, 2024.

[18] K Huang, P Chandak, Q Wang, S Havaldar, A Vaid, J Leskovec, G Nadkarni, BS Glicksberg, N Gehlenborg, and M Zitnik. Zero-shot prediction of therapeutic use with geometric deep learning and clinician centered design. *Health Informatics (March 2023)*, 10(2023.03):19–23287458.

[19] Yepeng Huang, Xiaorui Su, Varun Ullanat, Ivy Liang, Lindsay Clegg, Damilola Olabode, Nicholas Ho, Bino John, Megan Gibbs, and Marinka Zitnik. Multimodal AI predicts clinical outcomes of drug combinations from preclinical data. *arXiv preprint arXiv:2503.02781*, 2025.

[20] Wolfgang Hueber, Bruce E Sands, Steve Lewitzky, Marc Vandemeulebroecke, Walter Reinisch, Peter DR Higgins, Jan Wehkamp, Brian G Feagan, Michael D Yao, Marek Karczewski, et al. Secukinumab, a human anti-IL-17A monoclonal antibody, for moderate to severe Crohn’s disease: unexpected results of a randomised, double-blind placebo-controlled trial. *Gut*, 61(12):1693–1700, 2012.

[21] John Jumper, Richard Evans, Alexander Pritzel, Tim Green, Michael Figurnov, Olaf Ronneberger, Kathryn Tunyasuvunakool, Russ Bates, Augustin Žídek, Anna Potapenko, et al. Highly accurate protein structure prediction with AlphaFold. *nature*, 596(7873):583–589, 2021.

[22] Taku Kuwabara, Fumio Ishikawa, Motonari Kondo, and Terutaka Kakiuchi. The role of IL-17 and related cytokines in inflammatory autoimmune diseases. *Mediators of inflammation*, 2017(1):3908061, 2017.

[23] Jacob S Lee, Cristina M Tato, Barbara Joyce-Shaikh, Muhammet F Gulen, Corinne Cayatte, Yi Chen, Wendy M Blumenschein, Michael Judo, Gulezi Ayanoglu, Terrill K McClanahan, et al. Interleukin-23-independent IL-17 production regulates intestinal epithelial permeability. *Immunity*, 43(4):727–738, 2015.

[24] Charlie W Lees, JC Barrett, M Parkes, and J Satsangi. New IBD genetics: common pathways with other diseases. *Gut*, 60(12):1739–1753, 2011.

[25] Zeming Lin, Halil Akin, Roshan Rao, Brian Hie, Zhongkai Zhu, Wenting Lu, Nikita Smetanin, Robert Verkuil, Ori Kabeli, Yaniv Shmueli, et al. Evolutionary-scale prediction of atomic-level protein structure with a language model. *Science*, 379(6637):1123–1130, 2023.

[26] Michael J Pattison, Kirsty F MacKenzie, and J Simon C Arthur. Inhibition of JAKs in macrophages increases lipopolysaccharide-induced cytokine production by blocking IL-10-mediated feedback. *The Journal of Immunology*, 189(6):2784–2792, 2012.

[27] Tamara Pérez-Jeldres, Christopher J Tyler, Joshua D Boyer, Thangaraj Karuppuchamy, Andrés Yarur, Daniel A Giles, Shaila Yeasmin, Luke Lundborg, William J Sandborn, Derek R Patel, et al. Targeting cytokine signaling and lymphocyte traffic via small molecules in inflammatory bowel disease: JAK inhibitors and S1PR agonists. *Frontiers in pharmacology*, 10:212, 2019.

[28] Janice M Reimer, Andrea M Dickey, Yu Xuan Lin, Robert G Abrisch, Sebastian Mathea, Deep Chatterjee, Elizabeth J Fay, Stefan Knapp, Matthew D Daugherty, Samara L Reck-Peterson, et al. Structure of lrrk1 and mechanisms of autoinhibition and activation. *Nature structural & molecular biology*, 30(11):1735–1745, 2023.

[29] William J Sandborn, Chinyu Su, Bruce E Sands, Geert R D’Haens, Séverine Vermeire, Stefan Schreiber, Silvio Danese, Brian G Feagan, Walter Reinisch, Wojciech Niezychowski, et al. Tofacitinib as induction and maintenance therapy for ulcerative colitis. *New England Journal of Medicine*, 376(18):1723–1736, 2017.

[30] Michael Schlichtkrull, Thomas N. Kipf, Peter Bloem, Rianne van den Berg, Ivan Titov, and Max Welling. Modeling relational data with graph convolutional networks. In Aldo Gangemi, Roberto Navigli, Maria-Ester Vidal, Pascal Hitzler, Raphaël Troncy, Laura Hollink, Anna Tordai, and Mehwish Alam, editors, *The Semantic Web*, pages 593–607, Cham, 2018. Springer International Publishing.

[31] Baoxu Shi and Tim Weninger. Open-world knowledge graph completion. In *Proceedings of the AAAI conference on artificial intelligence*, volume 32, 2018.

[32] Gurdeep Singh, Lilija Wehling, Ahmad Wisnu Mulyadi, Rakesh Hadne Sreenath, Thomas Klabunde, Tommaso Andreani, and Douglas McCloskey. Talk2Biomodels and Talk2KnowledgeGraph: AI agent-based application for prediction of patient biomarkers and reasoning over biomedical knowledge graphs. In *ICLR 2025 Workshop on Machine Learning for Genomics Explorations*.

[33] K Sugamura, T Takeshita, M Kondo, N Ishii, M Nakamura, J Van Snick, and Y Kimura. Sharing of the IL-2 receptor γ chain with the functional IL-9 receptor complex. *Cytokine*, 6(5):555, 1994.

[34] Zhiqing Sun, Zhi-Hong Deng, Jian-Yun Nie, and Jian Tang. Rotate: Knowledge graph embedding by relational rotation in complex space. In *International Conference on Learning Representations*, 2019.

[35] Stephan R Targan, Brian Feagan, Severine Vermeire, Remo Panaccione, Gil Y Melmed, Carol Landers, Dalin Li, Chris Russell, Richard Newmark, Nan Zhang, et al. A randomized, double-blind, placebo-controlled phase 2 study of brodalumab in patients with moderate-to-severe Crohn’s disease. *Official journal of the American College of Gastroenterology ACG*, 111(11):1599–1607, 2016.

[36] Tom Thomas, Matthias Friedrich, Charlotte Rich-Griffin, Mathilde Pohin, Devika Agarwal, Julia Pakpoor, Carl Lee, Ruchi Tandon, Aniko Rendek, Dominik Aschenbrenner, et al. A longitudinal single-cell atlas of anti-tumour necrosis factor treatment in inflammatory bowel disease. *Nature immunology*, 25(11):2152–2165, 2024.

[37] Théo Trouillon, Johannes Welbl, Sebastian Riedel, Éric Gaussier, and Guillaume Bouchard. Complex embeddings for simple link prediction. In *International conference on machine learning*, pages 2071–2080. PMLR, 2016.

[38] Xinmin Wang, Hang Bao, Yuya Wang, Yalu Wang, Cheng Guo, Yanning Wu, Yongbin Xu, and Yali Li. Innovative peptide therapeutics targeting IL17RA to regulate inflammatory responses. *Scientific Reports*, 15(1):8542, 2025.

[39] Zifeng Wang, Zichen Wang, Balasubramaniam Srinivasan, Vassilis N. Ioannidis, Huzefa Rangwala, and RISHITA ANUBHAI. Biobridge: Bridging biomedical foundation models via knowledge graphs. In *The Twelfth International Conference on Learning Representations*, 2024.

[40] Ran Xu, Patrick Jiang, Linhao Luo, Cao Xiao, Adam Cross, Shirui Pan, Jimeng Sun, and Carl Yang. A survey on unifying large language models and knowledge graphs for biomedicine and healthcare. In *Proceedings of the 31st ACM SIGKDD Conference on Knowledge Discovery and Data Mining V.2, KDD ’25*, page 6195–6205, New York, NY, USA, 2025. Association for Computing Machinery.

[41] Jie Yan, Matija Hedl, Clara Abraham, et al. An inflammatory bowel disease–risk variant in INAVA decreases pattern recognition receptor–induced outcomes. *The Journal of clinical investigation*, 127(6):2192–2205, 2017.

[42] Bishan Yang, Scott Wen-tau Yih, Xiaodong He, Jianfeng Gao, and Li Deng. Embedding entities and relations for learning and inference in knowledge bases. In *Proceedings of the International Conference on Learning Representations (ICLR) 2015*, May 2015.

[43] Qing Ye, Yundian Zeng, Linlong Jiang, Yu Kang, Peichen Pan, Jiming Chen, Yafeng Deng, Haitao Zhao, Shibo He, Tingjun Hou, and Chang-Yu Hsieh. A knowledge-guided graph learning approach bridging phenotype- and target-based drug discovery. *Advanced Science*, 12(16):2412402, 2025.

[44] Yuan Yuan, Tina Sriskandarajah, Anna-Luisa Brakman, Alec Helyar, Alex Beutel, Andrea Vallone, and Saachi Jain. From Hard Refusals to Safe-Completions: Toward Output-Centric Safety Training. *arXiv preprint arXiv:2508.09224*, 2025.

[45] Zhen Zeng, Hao Lin, Mingshan Jiang, Jing Yuan, Xi Li, Yongbin Jia, Li Yang, and Hu Zhang. Anti-tnf α in inflammatory bowel disease: from originators to biosimilars. *Frontiers in Pharmacology*, 15:1424606, 2024.

[46] Ke Zhang, Min Wu, Yong Liu, Yimiao Feng, and Jie Zheng. KR4SL: knowledge graph reasoning for explainable prediction of synthetic lethality. *Bioinformatics*, 39:i158–i167, 06 2023.

[47] Zhaocheng Zhu, Zuobai Zhang, Louis Pascal Xhonneux, and Jian Tang. Neural bellman ford networks: A general graph neural network framework for link prediction. *Advances in Neural Information Processing Systems*, 34:29476–29490, 2021.

A Additional Analysis

We further analyzed the KG completion performance by grouping results according to unique triple types, as presented in Table 2. In addition, we performed an auxiliary analysis of association scores between the preselected genes and IBD subtypes, as shown in the heatmap in Fig. 3.

Table 2: Performance breakdown on the test set by unique triple type (*i.e.*, head type, edge type, tail type) with respect to genes. Specifically, Gene → Disease triples involve the `association_with` relation, while all other triples use the `interacts_with` relation.

| Metric | Gene – Disease | Gene – Bio. Proc. | Gene – Mol. Func. | Gene – Cell Comp. |
|---------|----------------|-------------------|-------------------|-------------------|
| MRR | 0,247 | 0,047 | 0,195 | 0,174 |
| Hits@1 | 0,199 | 0,025 | 0,178 | 0,053 |
| Hits@3 | 0,262 | 0,050 | 0,179 | 0,244 |
| Hits@10 | 0,343 | 0,087 | 0,229 | 0,405 |

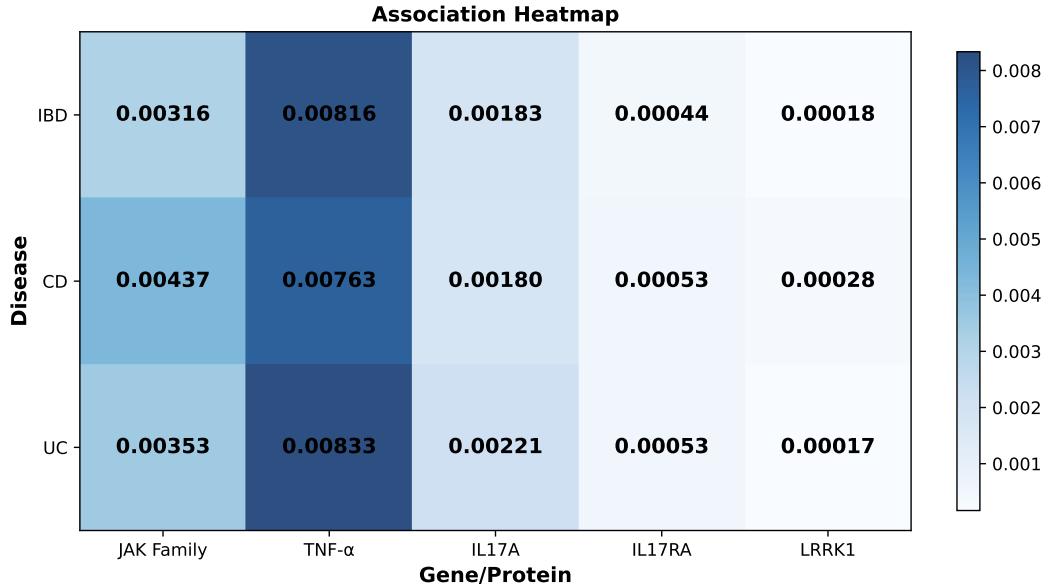


Figure 3: Association scores heatmap of preselected genes and IBD sub-types estimated by reciprocal rank with both directions of link predictions. IBD: Inflammatory Bowel Disease, CD: Crohn’s Disease, UC: Ulcerative Colitis.

B System Prompt

The prompt for prompting language models to answer whether the given target is suitable for IBD.

prompt = """

You are a biomedical research assistant with expertise in genetics, disease mechanisms, and drug target discovery. Your task is to evaluate whether inhibiting a given protein is a promising therapeutic target for treating Inflammatory Bowel Disease (IBD).

You will be provided with:

1. Predicted gene-disease association strength.
2. Knowledge graph nodes/paths connecting the gene/protein to IBD.
3. Knowledge graph nodes/paths connecting the gene/protein to its known Gene Ontology (GO) terms describing the biological processes (BP), and molecular functions (MF).

Your task:

1. Interpret the evidence (association strength, KG paths, GO terms).
2. Assess whether the gene has a therapeutic role in IBD that could be modulated by inhibition based on the evidence. Note that inhibition means reducing or blocking the gene/protein's activity.
3. Provide a structured answer with:

Advantages: Arguments that support inhibiting the protein as a therapeutic strategy for IBD.

Disadvantages: Arguments against inhibiting the protein as a therapeutic strategy for IBD.

Overall assessment: Based on the advantages and disadvantages, provide 3-5 bullet points summarizing whether inhibiting the protein is a promising therapeutic target for IBD. Keep them short and concise.

Please note strong association does not guarantee therapeutic relevance, and weak association does not rule it out. Ground your judgement completely based on the information provided. Do not make up any information or use prior knowledge.

"""

C Case Study: JAK1

We report both LLM response and the corresponding paths connecting JAK1 with IBD, and with its biological processes and molecular functions as the input context.

C.1 LLM response

Advantages:

- Very strong predicted association with IBD.
- Dense network proximity to an IBD-associated node: direct protein-protein interaction with PTPN2 (which is associated with inflammatory bowel disease), plus multiple converging paths via JAK3/KPNA1 to PTPN2.
- Extensive involvement in cytokine-mediated signaling (direct links to STAT1/STAT2/STAT3, JAK3, IL27RA, IFNLR1, OSMR; GO: cytokine-mediated signaling pathway; receptor signaling via JAK-STAT; multiple interleukin-mediated pathways including IL-2/4/6/7/9/10/15; interferon-gamma-mediated signaling). Inhibition could dampen overactive cytokine/JAK-STAT signaling.
- Connections to protein phosphorylation and kinase activity suggest it sits at a signaling control point where inhibition may yield broad anti-inflammatory effects.

Disadvantages:

- Strong linkage to antiviral pathways (GO: response to virus; type I and type III interferon signaling), implying inhibition may impair antiviral defense and increase infection risk.
- Participation in interleukin-10-mediated signaling indicates potential to disrupt regulatory cytokine signaling, risking immune dysregulation if inhibited.
- Evidence to IBD is indirect in the KG (through PTPN2 association rather than a direct association for the protein itself), leaving causality uncertain.
- GO links to cell adhesion and protein localization to cell-cell junctions suggest possible effects on epithelial barrier functions if inhibited.
- Broad roles in receptor signaling via JAK-STAT and protein phosphorylation imply pleiotropic on-target effects and safety concerns.

Overall assessment:

- Biologically plausible target: inhibition could suppress cytokine/JAK-STAT-driven immune signaling connected to IBD pathways.
- Network evidence is strong but indirect; the link to IBD relies on proximity to PTPN2 rather than a direct association.
- High risk of broad immunosuppression, including dampening antiviral interferon pathways and multiple interleukin pathways.
- Potential epithelial barrier and systemic signaling liabilities from wide pathway involvement.
- Overall, inhibition could be promising only with careful strategy to avoid broad pathway suppression; current evidence suggests notable safety trade-offs and the need for more direct IBD-specific validation.

C.2 Paths connecting JAK1 → IBD

Path #0: $\langle \text{PROTEIN IN QUESTION, ppi, PTPN2} \rangle \rightarrow \langle \text{PTPN2, associated with, inflammatory bowel disease} \rangle$

Path #1: $\langle \text{PROTEIN IN QUESTION, ppi}^{-1}, \text{PTPN2} \rangle \rightarrow \langle \text{PTPN2, associated with, inflammatory bowel disease} \rangle$

Path #2: $\langle \text{PROTEIN IN QUESTION, ppi, PTPN2} \rangle \rightarrow \langle \text{PTPN2, associated with}^{-1}, \text{inflammatory bowel disease} \rangle$

Path #3: $\langle \text{PROTEIN IN QUESTION, ppi}^{-1}, \text{PTPN2} \rangle \rightarrow \langle \text{PTPN2, associated with}^{-1}, \text{inflammatory bowel disease} \rangle$

Path #4: $\langle \text{PROTEIN IN QUESTION, ppi}^{-1}, \text{JAK3} \rangle \rightarrow \langle \text{JAK3, ppi, PTPN2} \rangle \rightarrow \langle \text{PTPN2, associated with, inflammatory bowel disease} \rangle$

Path #5: $\langle \text{PROTEIN IN QUESTION, ppi}^{-1}, \text{JAK3} \rangle \rightarrow \langle \text{JAK3, ppi}^{-1}, \text{PTPN2} \rangle \rightarrow \langle \text{PTPN2, associated with, inflammatory bowel disease} \rangle$

Path #6: $\langle \text{PROTEIN IN QUESTION, ppi}^{-1}, \text{JAK3} \rangle \rightarrow \langle \text{JAK3, ppi, PTPN2} \rangle \rightarrow \langle \text{PTPN2, associated with}^{-1}, \text{inflammatory bowel disease} \rangle$

Path #7: $\langle \text{PROTEIN IN QUESTION, ppi}^{-1}, \text{JAK3} \rangle \rightarrow \langle \text{JAK3, ppi}^{-1}, \text{PTPN2} \rangle \rightarrow \langle \text{PTPN2, associated with}^{-1}, \text{inflammatory bowel disease} \rangle$

Path #8: $\langle \text{PROTEIN IN QUESTION, ppi, JAK3} \rangle \rightarrow \langle \text{JAK3, ppi, PTPN2} \rangle \rightarrow \langle \text{PTPN2, associated with, inflammatory bowel disease} \rangle$

Path #9: $\langle \text{PROTEIN IN QUESTION, ppi}^{-1}, \text{JAK3} \rangle \rightarrow \langle \text{JAK3, ppi}^{-1}, \text{KPNA1} \rangle \rightarrow \langle \text{KPNA1, ppi}^{-1}, \text{PTPN2} \rangle \rightarrow \langle \text{PTPN2, associated with, inflammatory bowel disease} \rangle$

Path #10: $\langle \text{PROTEIN IN QUESTION, ppi}^{-1}, \text{JAK3} \rangle \rightarrow \langle \text{JAK3, ppi, KPNA1} \rangle \rightarrow \langle \text{KPNA1, ppi}^{-1}, \text{PTPN2} \rangle \rightarrow \langle \text{PTPN2, associated with, inflammatory bowel disease} \rangle$

C.3 Paths connecting JAK1 → GO terms

Path #0: $\langle \text{PROTEIN IN QUESTION, ppi, TGFBR2} \rangle \rightarrow \langle \text{TGFBR2, interacts with, ATP binding} \rangle$

Path #1: $\langle \text{PROTEIN IN QUESTION, ppi, MDK} \rangle \langle \text{MDK, ppi, STAT1} \rangle \rightarrow \langle \text{STAT1, interacts with, CCR5 chemokine receptor binding} \rangle$

Path #2: $\langle \text{PROTEIN IN QUESTION, ppi, MDK} \rangle \langle \text{MDK, ppi, JOSD2} \rangle \rightarrow \langle \text{JOSD2, ppi, LRTOMT} \rangle \rightarrow \langle \text{LRTOMT, interacts with, developmental process} \rangle \rightarrow \langle \text{developmental process, parent-child, cellular developmental process} \rangle \rightarrow \langle \text{cellular developmental process, parent-child, cell differentiation} \rangle$

Path #3: $\langle \text{PROTEIN IN QUESTION, ppi, MDK} \rangle \langle \text{MDK, ppi, JOSD2} \rangle \rightarrow \langle \text{JOSD2, ppi, LRTOMT} \rangle \rightarrow \langle \text{LRTOMT, interacts with, developmental process} \rangle \rightarrow \langle \text{developmental process, parent-child, cellular developmental process} \rangle \rightarrow \langle \text{cellular developmental process, parent-child, cell differentiation} \rangle$

Path #4: $\langle \text{PROTEIN IN QUESTION, ppi, EIF2AK2} \rangle \rightarrow \langle \text{EIF2AK2, interacts with, response to virus} \rangle \rightarrow \langle \text{response to virus, parent-child, cellular response to virus} \rangle$

Path #5: $\langle \text{PROTEIN IN QUESTION, ppi, EIF2AK2} \rangle \rightarrow \langle \text{EIF2AK2, ppi, RPLP0} \rangle \rightarrow \langle \text{RPLP0, ppi, RPS14} \rangle \rightarrow \langle \text{RPS14, ppi, RPS15A} \rangle \rightarrow \langle \text{RPS15A, interacts with, response to virus} \rangle \rightarrow \langle \text{response to virus, parent-child, cellular response to virus} \rangle$

Path #6: $\langle \text{PROTEIN IN QUESTION, ppi, EIF2AK2} \rangle \rightarrow \langle \text{EIF2AK2, ppi, RPLP0} \rangle \rightarrow \langle \text{RPLP0, ppi, TBP} \rangle \langle \text{TBP, ppi, TNF} \rangle \rightarrow \langle \text{TNF, interacts with, response to virus} \rangle \rightarrow \langle \text{response to virus, parent-child, cellular response to virus} \rangle$

Path #7: $\langle \text{PROTEIN IN QUESTION, ppi, IL27RA} \rangle \rightarrow \langle \text{IL27RA, interacts with, cell surface receptor signaling pathway} \rangle \rightarrow \langle \text{cell surface receptor signaling pathway, parent-child, cytokine-mediated signaling pathway} \rangle$

Path #8: $\langle \text{PROTEIN IN QUESTION, ppi, IFNLR1} \rangle \rightarrow \langle \text{IFNLR1, interacts with, cytokine-mediated signaling pathway} \rangle$

Path #9: $\langle \text{PROTEIN IN QUESTION, ppi, STAT1} \rangle \rightarrow \langle \text{STAT1, interacts with, cytokine-mediated signaling pathway} \rangle$

Path #10: <PROTEIN IN QUESTION, ppi, OSMR> → <OSMR, interacts with, cytokine-mediated signaling pathway>

Path #11: <PROTEIN IN QUESTION, ppi, IL27RA> → <IL27RA, interacts with, cytokine-mediated signaling pathway>

Path #12: <PROTEIN IN QUESTION, ppi, STAT3> → <STAT3, interacts with, growth hormone receptor signaling pathway> → <growth hormone receptor signaling pathway, parent-child, growth hormone receptor signaling pathway via JAK-STAT>

Path #13: <PROTEIN IN QUESTION, ppi, JAK3> → <JAK3, interacts with, enzyme linked receptor protein signaling pathway> → <enzyme linked receptor protein signaling pathway, parent-child, transmembrane receptor protein tyrosine kinase signaling pathway> → <transmembrane receptor protein tyrosine kinase signaling pathway, parent-child, growth hormone receptor signaling pathway> → <growth hormone receptor signaling pathway, parent-child, growth hormone receptor signaling pathway via JAK-STAT>

Path #14: <PROTEIN IN QUESTION, interacts with, cytokine-mediated signaling pathway> → <cytokine-mediated signaling pathway, parent-child, interferon-gamma-mediated signaling pathway>

Path #15: <PROTEIN IN QUESTION, ppi, IFNLR1> → <IFNLR1, interacts with, cytokine-mediated signaling pathway> → <cytokine-mediated signaling pathway, parent-child, interleukin-10-mediated signaling pathway>

Path #16: <PROTEIN IN QUESTION, ppi, UBC> → <UBC, interacts with, cytokine-mediated signaling pathway> → <cytokine-mediated signaling pathway, parent-child, interleukin-10-mediated signaling pathway>

Path #17: <PROTEIN IN QUESTION, ppi, TEC> → <TEC, interacts with, cytokine-mediated signaling pathway> → <cytokine-mediated signaling pathway, parent-child, interleukin-10-mediated signaling pathway>

Path #18: <PROTEIN IN QUESTION, interacts with, cytokine-mediated signaling pathway> → <cytokine-mediated signaling pathway, parent-child, interleukin-10-mediated signaling pathway>

Path #19: <PROTEIN IN QUESTION, ppi, TSHR> → <TSHR, interacts with, cell surface receptor signaling pathway> → <cell surface receptor signaling pathway, parent-child, cytokine-mediated signaling pathway> → <cytokine-mediated signaling pathway, parent-child, interleukin-10-mediated signaling pathway>

Path #20: <PROTEIN IN QUESTION, interacts with, cytokine-mediated signaling pathway> → <cytokine-mediated signaling pathway, parent-child, interleukin-15-mediated signaling pathway>

Path #21: <PROTEIN IN QUESTION, ppi, STAT1> → <STAT1, ppi, IL27RA> → <IL27RA, interacts with, cell surface receptor signaling pathway> → <cell surface receptor signaling pathway, parent-child, cytokine-mediated signaling pathway> → <cytokine-mediated signaling pathway, parent-child, interleukin-15-mediated signaling pathway>

Path #22: <PROTEIN IN QUESTION, ppi, EIF2AK2> → <EIF2AK2, ppi, TIRAP> → <TIRAP, interacts with, cell surface receptor signaling pathway> → <cell surface receptor signaling pathway, parent-child, cytokine-mediated signaling pathway> → <cytokine-mediated signaling pathway, parent-child, interleukin-15-mediated signaling pathway>

Path #23: <PROTEIN IN QUESTION, ppi, SOCS1> → <SOCS1, ppi, TIRAP> → <TIRAP, interacts with, cell surface receptor signaling pathway> → <cell surface receptor signaling pathway, parent-child, cytokine-mediated signaling pathway> → <cytokine-mediated signaling pathway, parent-child, interleukin-15-mediated signaling pathway>

Path #24: <PROTEIN IN QUESTION, interacts with, cytokine-mediated signaling pathway> → <cytokine-mediated signaling pathway, parent-child, interleukin-2-mediated signaling pathway>

Path #25: <PROTEIN IN QUESTION, ppi, STAT1> → <STAT1, ppi, IL27RA> → <IL27RA, interacts with, cell surface receptor signaling pathway> → <cell surface

receptor signaling pathway, parent-child, cytokine-mediated signaling pathway) →
(cytokine-mediated signaling pathway, parent-child, interleukin-2-mediated signaling pathway)

Path #26: <PROTEIN IN QUESTION, ppi, JAK3> → <JAK3, interacts with, interleukin-4-mediated signaling pathway>

Path #27: <PROTEIN IN QUESTION, interacts with, cytokine-mediated signaling pathway> → (cytokine-mediated signaling pathway, parent-child, interleukin-6-mediated signaling pathway)

Path #28: <PROTEIN IN QUESTION, ppi, STAT1> → <STAT1, ppi, IL27RA> → <IL27RA, interacts with, cell surface receptor signaling pathway> → (cell surface receptor signaling pathway, parent-child, cytokine-mediated signaling pathway) → (cytokine-mediated signaling pathway, parent-child, interleukin-6-mediated signaling pathway)

Path #29: <PROTEIN IN QUESTION, ppi, EIF2AK2> → <EIF2AK2, ppi, TIRAP> → <TIRAP, interacts with, cell surface receptor signaling pathway> → (cell surface receptor signaling pathway, parent-child, cytokine-mediated signaling pathway) → (cytokine-mediated signaling pathway, parent-child, interleukin-6-mediated signaling pathway)

Path #30: <PROTEIN IN QUESTION, ppi, SOCS3> → <SOCS3, ppi, MAPK8IP1> → <MAPK8IP1, ppi, SPAG6> → <SPAG6, ppi, SH2B2> → <SH2B2, interacts with, cytokine-mediated signaling pathway> → (cytokine-mediated signaling pathway, parent-child, interleukin-6-mediated signaling pathway)

Path #31: <PROTEIN IN QUESTION, interacts with, cytokine-mediated signaling pathway> → (cytokine-mediated signaling pathway, parent-child, interleukin-7-mediated signaling pathway)

Path #32: <PROTEIN IN QUESTION, interacts with, cytokine-mediated signaling pathway> → (cytokine-mediated signaling pathway, parent-child, interleukin-9-mediated signaling pathway)

Path #33: <PROTEIN IN QUESTION, ppi, STAT1> → <STAT1, ppi, IL27RA> → <IL27RA, interacts with, cell surface receptor signaling pathway> → (cell surface receptor signaling pathway, parent-child, cytokine-mediated signaling pathway) → (cytokine-mediated signaling pathway, parent-child, interleukin-9-mediated signaling pathway)

Path #34: <PROTEIN IN QUESTION, ppi, PTPN2> → <PTPN2, ppi, STAT3> → <STAT3, ppi, TSHR> → <TSHR, interacts with, cell surface receptor signaling pathway> → (cell surface receptor signaling pathway, parent-child, cytokine-mediated signaling pathway) → (cytokine-mediated signaling pathway, parent-child, interleukin-9-mediated signaling pathway)

Path #35: <PROTEIN IN QUESTION, ppi, TEC> <TEC, ppi, VAV1> → <VAV1, ppi, TUBB> → <TUBB, interacts with, cellular process> → (cellular process, parent-child, signal transduction) → (signal transduction, parent-child, intracellular signal transduction)

Path #36: <PROTEIN IN QUESTION, ppi, TEC> → <TEC, interacts with, intracellular signal transduction>

Path #37: <PROTEIN IN QUESTION, ppi, PTPN6> → <PTPN6, interacts with, intracellular signal transduction>

Path #38: <PROTEIN IN QUESTION, ppi, HAVCR2> → <HAVCR2, interacts with, metal ion binding>

Path #39: <PROTEIN IN QUESTION, ppi, TEC> → <TEC, interacts with, non-membrane spanning protein tyrosine kinase activity>

Path #40: <PROTEIN IN QUESTION, ppi, PTPN6> → <PTPN6, ppi, SYK> → <SYK, interacts with, non-membrane spanning protein tyrosine kinase activity>

Path #41: <PROTEIN IN QUESTION, ppi, EIF2AK2> → <EIF2AK2, interacts with, non-membrane spanning protein tyrosine kinase activity>

Path #42: <PROTEIN IN QUESTION, ppi, MDK> → <MDK, interacts with, positive regulation of cell adhesion> → <positive regulation of cell adhesion, parent-child, positive regulation of cell-cell adhesion> → <positive regulation of cell-cell adhesion, parent-child, positive regulation of homotypic cell-cell adhesion>

Path #43: <PROTEIN IN QUESTION, ppi, PRKAA1> → <PRKAA1, interacts with, positive regulation of cellular protein localization> → <positive regulation of cellular protein localization, parent-child, positive regulation of protein localization to nucleus>

Path #44: <PROTEIN IN QUESTION, ppi, MAP3K5> → <MAP3K5, ppi, PIK3R1> → <PIK3R1, ppi, PRKAA1> → <PRKAA1, interacts with, positive regulation of cellular protein localization> → <positive regulation of cellular protein localization, parent-child, positive regulation of protein localization to nucleus>

Path #45: <PROTEIN IN QUESTION, ppi, UBC> → <UBC, ppi, ARHGEF18> → <ARHGEF18, interacts with, protein localization to cell-cell junction>

Path #46: <PROTEIN IN QUESTION, ppi, RACK1> → <RACK1, interacts with, protein phosphatase binding>

Path #47: <PROTEIN IN QUESTION, ppi, SHOC2> → <SHOC2, interacts with, protein phosphatase binding>

Path #48: <PROTEIN IN QUESTION, ppi, MAP3K5> → <MAP3K5, interacts with, protein phosphatase binding>

Path #49: <PROTEIN IN QUESTION, ppi, TEC> → <TEC, interacts with, protein phosphorylation>

Path #50: <PROTEIN IN QUESTION, ppi, TEC> <TEC, ppi, WAS> → <WAS, ppi, NCK2> <NCK2, ppi, PKP4> → <PKP4, ppi, MKNK1> → <MKNK1, interacts with, protein phosphorylation>

Path #51: <PROTEIN IN QUESTION, ppi, TGFBR2> → <TGFBR2, interacts with, protein phosphorylation>

Path #52: <PROTEIN IN QUESTION, ppi, MAP3K5> → <MAP3K5, interacts with, protein kinase activity> → <protein kinase activity, parent-child, protein tyrosine kinase activity>

Path #53: <PROTEIN IN QUESTION, ppi, STAT1> → <STAT1, interacts with, receptor signaling pathway via JAK-STAT>

Path #54: <PROTEIN IN QUESTION, ppi, SOCS3> → <SOCS3, interacts with, receptor signaling pathway via JAK-STAT>

Path #55: <PROTEIN IN QUESTION, ppi, SLC15A1> → <SLC15A1, ppi, OPA3> → <OPA3, interacts with, response to stimulus> → <response to stimulus, parent-child, response to chemical> → <response to chemical, parent-child, response to antibiotic>

Path #56: <PROTEIN IN QUESTION, ppi, MDK> <MDK, ppi, STAT1> → <STAT1, interacts with, type I interferon signaling pathway>

Path #57: <PROTEIN IN QUESTION, ppi, UBC> → <UBC, interacts with, cytokine-mediated signaling pathway> → <cytokine-mediated signaling pathway, parent-child, type III interferon signaling pathway>

Path #58: <PROTEIN IN QUESTION, ppi, IL27RA> → <IL27RA, interacts with, cell surface receptor signaling pathway> → <cell surface receptor signaling pathway, parent-child, cytokine-mediated signaling pathway> → <cytokine-mediated signaling pathway, parent-child, type III interferon signaling pathway>

Path #59: <PROTEIN IN QUESTION, ppi, IFNLR1> → <IFNLR1, interacts with, cytokine-mediated signaling pathway> → <cytokine-mediated signaling pathway, parent-child, type III interferon signaling pathway>

Path #60: <PROTEIN IN QUESTION, ppi, TEC> → <TEC, interacts with, cytokine-mediated signaling pathway> → <cytokine-mediated signaling pathway, parent-child, type III interferon signaling pathway>

Path #61: <PROTEIN IN QUESTION, interacts with, cytokine-mediated signaling pathway> → <cytokine-mediated signaling pathway, parent-child, type III interferon signaling pathway>

Path #62: <PROTEIN IN QUESTION, ppi, STAT2> → <STAT2, interacts with, ubiquitin-like protein ligase binding> → <ubiquitin-like protein ligase binding, parent-child, ubiquitin protein ligase binding>

Path #63: <PROTEIN IN QUESTION, ppi, STAT1> → <STAT1, interacts with, ubiquitin-like protein ligase binding> → <ubiquitin-like protein ligase binding, parent-child, ubiquitin protein ligase binding>

Path #64: <PROTEIN IN QUESTION, ppi, MDM2> → <MDM2, interacts with, ubiquitin protein ligase binding>

D Case Study: IL17A

We report both LLM response and the corresponding paths connecting IL17A with IBD, and with its biological processes and molecular functions as the input context.

D.1 LLM response

Advantages:

- Biologic plausibility: The protein connects to IL-17/IL-17A-mediated and broader cytokine-mediated signaling (via IL17RA/IL17RC and TRAF3IP2), which drive pro-inflammatory cytokine outputs (links to positive regulation of IL-6, IL-12, IL-16 via PELI1/PARK7). Inhibition could dampen these inflammatory cascades.
- Network centrality to immune activation: Multiple PPIs with TRAF6 and downstream adapters place the protein within innate/adaptive immune response pathways relevant to mucosal inflammation.
- Indirect disease linkage: A path to IBD exists through TRAF6 → RIPK2 (RIPK2 associated with IBD), providing a mechanistic bridge between the protein and an established IBD signaling axis.

Disadvantages:

- Weak and indirect disease association: The only IBD link is via TRAF6 → RIPK2; there is no direct association of the protein itself with IBD.
- Potential loss of host defense: The protein is connected to defense responses to bacteria (including Gram-negative) and fungi (via NOD2/TRAF6 and MALT1 paths). Inhibition could impair mucosal antimicrobial defense.
- Possible barrier/repair effects: Links to epithelial/keratinocyte proliferation suggest a role in epithelial biology; inhibition may hinder mucosal repair.
- Broad signaling involvement: Extensive connections to cell surface receptor signaling (including IL-17 and Notch pathway edges) and protein dimerization suggest wide immunologic roles, raising safety concerns if inhibited.

Overall assessment:

- Evidence supports involvement in IL-17-driven inflammatory signaling, so inhibition could reduce pro-inflammatory cytokines.
- However, association with IBD is weak and only indirect; therapeutic relevance is uncertain.
- Strong ties to antibacterial/antifungal defense and epithelial biology raise safety risks if inhibited.
- Overall, this appears a low-priority IBD target for inhibition pending disease-specific functional validation in intestinal models and patient tissues.

D.2 Paths connecting IL17A → IBD

Path #0: ⟨PROTEIN IN QUESTION, ppi⁻¹, TRAF6⟩ → ⟨TRAF6, ppi, RIPK2⟩ → ⟨RIPK2, associated with⁻¹, inflammatory bowel disease⟩

D.3 Paths connecting IL17A → GO terms

Path #0: ⟨PROTEIN IN QUESTION, ppi, IL17RA⟩ → ⟨IL17RA, interacts with, cell surface receptor signaling pathway⟩ → ⟨cell surface receptor signaling pathway, parent-child, Notch signaling pathway⟩

Path #1: ⟨PROTEIN IN QUESTION, ppi, TRAF3IP2⟩ → ⟨TRAF3IP2, ppi, IL17RA⟩ → ⟨IL17RA, interacts with, cell surface receptor signaling pathway⟩ → ⟨cell surface receptor signaling pathway, parent-child, Notch signaling pathway⟩

Path #2: ⟨PROTEIN IN QUESTION, interacts with, immune response⟩ → ⟨immune response, parent-child, adaptive immune response⟩

Path #3: ⟨PROTEIN IN QUESTION, ppi, DPP8⟩ → ⟨DPP8, interacts with, immune response⟩ → ⟨immune response, parent-child, adaptive immune response⟩

Path #4: ⟨PROTEIN IN QUESTION, ppi, TRAF6⟩ → ⟨TRAF6, interacts with, cellular response to cytokine stimulus⟩ → ⟨cellular response to cytokine stimulus, parent-child, cellular response to interleukin-1⟩

Path #5: ⟨PROTEIN IN QUESTION, ppi, TRAF6⟩ → ⟨TRAF6, ppi, NOD2⟩ → ⟨NOD2, interacts with, defense response to bacterium⟩ → ⟨defense response to bacterium, parent-child, defense response to Gram-negative bacterium⟩

Path #6: ⟨PROTEIN IN QUESTION, ppi, TRAF6⟩ → ⟨TRAF6, ppi, NOD2⟩ → ⟨NOD2, interacts with, defense response⟩ → ⟨defense response, parent-child, defense response to other organism⟩ → ⟨defense response to other organism, parent-child, defense response to bacterium⟩ → ⟨defense response to bacterium, parent-child, defense response to Gram-negative bacterium⟩

Path #7: ⟨PROTEIN IN QUESTION, ppi, TRAF6⟩ → ⟨TRAF6, ppi, MALT1⟩ → ⟨MALT1, interacts with, response to fungus⟩ → ⟨response to fungus, parent-child, defense response to fungus⟩

Path #8: ⟨PROTEIN IN QUESTION, interacts with, immune response⟩ → ⟨immune response, parent-child, defense response to other organism⟩ → ⟨defense response to other organism, parent-child, defense response to fungus⟩

Path #9: ⟨PROTEIN IN QUESTION, ppi, IL17RA⟩ → ⟨IL17RA, ppi, LYZL1⟩ → ⟨LYZL1, interacts with, metabolic process⟩ → ⟨metabolic process, parent-child, organic substance metabolic process⟩ → ⟨organic substance metabolic process, parent-child, macromolecule metabolic process⟩ → ⟨macromolecule metabolic process, parent-child, gene expression⟩

Path #10: ⟨PROTEIN IN QUESTION, interacts with, immune response⟩ → ⟨immune response, parent-child, defense response to other organism⟩ → ⟨defense response to other organism, parent-child, innate immune response⟩

Path #11: ⟨PROTEIN IN QUESTION, interacts with, cytokine-mediated signaling pathway⟩ → ⟨cytokine-mediated signaling pathway, parent-child, interleukin-17-mediated signaling pathway⟩

Path #12: ⟨PROTEIN IN QUESTION, ppi, IL17RA⟩ → ⟨IL17RA, interacts with, cell surface receptor signaling pathway⟩ → ⟨cell surface receptor signaling pathway, parent-child, cytokine-mediated signaling pathway⟩ → ⟨cytokine-mediated signaling pathway, parent-child, interleukin-17-mediated signaling pathway⟩

Path #13: ⟨PROTEIN IN QUESTION, interacts with, cytokine-mediated signaling pathway⟩ → ⟨cytokine-mediated signaling pathway, parent-child, interleukin-17A-mediated signaling pathway⟩

Path #14: ⟨PROTEIN IN QUESTION, ppi, IL17RA⟩ → ⟨IL17RA, interacts with, cell surface receptor signaling pathway⟩ → ⟨cell surface receptor signaling pathway, parent-child, cytokine-mediated signaling pathway⟩ → ⟨cytokine-mediated signaling pathway, parent-child, interleukin-17A-mediated signaling pathway⟩

Path #15: <PROTEIN IN QUESTION, ppi, TRAF6> <TRAF6, ppi, UBC> → <UBC, ppi, TNFRSF25> → <TNFRSF25, interacts with, cell surface receptor signaling pathway> → <cell surface receptor signaling pathway, parent-child, cytokine-mediated signaling pathway> → <cytokine-mediated signaling pathway, parent-child, interleukin-17A-mediated signaling pathway>

Path #16: <PROTEIN IN QUESTION, ppi, TRAF6> <TRAF6, ppi, UBC> → <UBC, ppi, TNFRSF10B> → <TNFRSF10B, interacts with, cell surface receptor signaling pathway> → <cell surface receptor signaling pathway, parent-child, cytokine-mediated signaling pathway> → <cytokine-mediated signaling pathway, parent-child, interleukin-17A-mediated signaling pathway>

Path #17: <PROTEIN IN QUESTION, ppi, TRAF6> <TRAF6, ppi, UBC> → <UBC, ppi, IFITM1> → <IFITM1, interacts with, cell surface receptor signaling pathway> → <cell surface receptor signaling pathway, parent-child, cytokine-mediated signaling pathway> → <cytokine-mediated signaling pathway, parent-child, interleukin-17A-mediated signaling pathway>

Path #18: <PROTEIN IN QUESTION, ppi, TRAF6> <TRAF6, ppi, UBC> → <UBC, ppi, PIP4K2B> → <PIP4K2B, interacts with, cell surface receptor signaling pathway> → <cell surface receptor signaling pathway, parent-child, cytokine-mediated signaling pathway> → <cytokine-mediated signaling pathway, parent-child, interleukin-17A-mediated signaling pathway>

Path #19: <PROTEIN IN QUESTION, ppi, TRAF6> <TRAF6, ppi, UBC> → <UBC, ppi, FADD> → <FADD, interacts with, cell surface receptor signaling pathway> → <cell surface receptor signaling pathway, parent-child, cytokine-mediated signaling pathway> → <cytokine-mediated signaling pathway, parent-child, interleukin-17A-mediated signaling pathway>

Path #20: <PROTEIN IN QUESTION, ppi, TRAF6> → <TRAF6, ppi, UCHL1> → <UCHL1, interacts with, cell population proliferation> → <cell population proliferation, parent-child, epithelial cell proliferation> → <epithelial cell proliferation, parent-child, keratinocyte proliferation>

Path #21: <PROTEIN IN QUESTION, ppi, TRAF6> → <TRAF6, ppi, PELI1> → <PELI1, interacts with, positive regulation of cytokine production> → <positive regulation of cytokine production, parent-child, positive regulation of interleukin-12 production>

Path #22: <PROTEIN IN QUESTION, ppi, TRAF6> → <TRAF6, ppi, PELI1> → <PELI1, interacts with, positive regulation of cytokine production> → <positive regulation of cytokine production, parent-child, positive regulation of interleukin-16 production>

Path #23: <PROTEIN IN QUESTION, ppi, TRAF6> → <TRAF6, ppi, PARK7> → <PARK7, interacts with, positive regulation of gene expression> → <positive regulation of gene expression, parent-child, positive regulation of cytokine production> → <positive regulation of cytokine production, parent-child, positive regulation of interleukin-16 production>

Path #24: <PROTEIN IN QUESTION, ppi, TRAF6> → <TRAF6, ppi, PELI1> → <PELI1, interacts with, positive regulation of cytokine production> → <positive regulation of cytokine production, parent-child, positive regulation of interleukin-6 production>

Path #25: <PROTEIN IN QUESTION, ppi, TRAF6> → <TRAF6, interacts with, protein binding> → <protein binding, parent-child, protein dimerization activity> → <protein dimerization activity, parent-child, protein heterodimerization activity>

Path #26: <PROTEIN IN QUESTION, ppi, LONP2> → <LONP2, interacts with, protein binding> → <protein binding, parent-child, protein dimerization activity> → <protein dimerization activity, parent-child, protein heterodimerization activity>

Path #27: <PROTEIN IN QUESTION, ppi, TRAF3IP2> → <TRAF3IP2, ppi, FBXW11> → <FBXW11, interacts with, protein binding> → <protein binding, parent-child, protein dimerization activity> → <protein dimerization activity, parent-child, protein heterodimerization activity>

Path #28: <PROTEIN IN QUESTION, interacts with, protein binding> → <protein binding, parent-child, protein dimerization activity> → <protein dimerization activity, parent-child, protein homodimerization activity>

Path #29: <PROTEIN IN QUESTION, ppi, IL17RC> → <IL17RC, interacts with, protein binding> → <protein binding, parent-child, protein dimerization activity> → <protein dimerization activity, parent-child, protein homodimerization activity>

Path #30: <PROTEIN IN QUESTION, ppi, IL17RC> → <IL17RC, ppi, DISP1> → <DISP1, interacts with, protein binding> → <protein binding, parent-child, protein dimerization activity> → <protein dimerization activity, parent-child, protein homodimerization activity>

Path #31: <PROTEIN IN QUESTION, ppi, IL17RC> → <IL17RC, ppi, DISP1> → <DISP1, interacts with, protein binding> → <protein binding, parent-child, protein dimerization activity> → <protein dimerization activity, parent-child, protein homodimerization activity>