

Hypothesis Generating Agent

Example queries and response

	Query	Response
1.	"What diseases is LRRK2 gene associated with, what pathways are affected by mutations in gene LRRK2?"	<p>I now know the final answer.</p> <p>Final Answer: The LRRK2 gene is primarily associated with Parkinson's disease. Mutations in the LRRK2 gene affect various biological pathways, including those related to kinase activity, protein interactions, and cellular signaling processes.</p>
2.	"How does gene LRRK2 interact with other genes in the pathway?"	<p>I have gathered extensive information about the gene LRRK2 and its interactions with various genes through the Gene Ontology (GO) annotations. The annotations indicate that LRRK2 is involved in numerous biological processes, molecular functions, and cellular components, which suggest its interactions with a wide range of other genes in various pathways.</p> <p>Final Answer: Gene LRRK2 interacts with numerous other genes in various pathways, as indicated by its extensive GO annotations related to processes such as signal transduction, protein phosphorylation, and cytoskeletal organization, among others.</p>
3.	"What are the potential downstream effects of overexpressing/downregulating gene LRRK2?"	<p>The observations from the knowledge base indicate that the LRRK2 gene is associated with numerous Gene Ontology (GO) terms, which suggest various biological processes and molecular functions that could be affected by its overexpression or downregulation.</p> <p>Some of the key GO annotations related to LRRK2 include:</p> <ol style="list-style-type: none">**Kinase Activity**: LRRK2 is involved in kinase activity (GO:0004672, GO:0004674), which suggests that changes in its expression could affect phosphorylation processes in cells.**Cellular Processes**: It is associated with processes such as cell signaling (GO:0007165), regulation of cellular processes (GO:0045860), and response to stress (GO:0006979).**Neurodegenerative Implications**: Given its association

		<p>with Parkinson's disease, overexpression or downregulation of LRRK2 may influence neurodegenerative pathways, potentially leading to altered neuronal function or survival.</p> <p>4. **Protein Interactions**: The gene is annotated with terms related to protein binding (GO:0005515), which indicates that it may interact with various proteins, affecting their functions and pathways.</p> <p>5. **Cellular Localization**: LRRK2 is involved in processes related to cellular localization (GO:0005783, GO:0005794), which could impact how proteins are distributed within the cell.</p> <p>6. **Regulation of Gene Expression**: It is also linked to the regulation of transcription (GO:0045944), suggesting that changes in LRRK2 levels could affect the expression of other genes.</p> <p>Overall, the potential downstream effects of overexpressing or downregulating LRRK2 could include altered kinase activity, disrupted cellular signaling, changes in protein interactions, and impacts on neuronal health, all of which are relevant to the pathology of Parkinson's disease.</p> <p>Final Answer: The potential downstream effects of overexpressing or downregulating the LRRK2 gene include altered kinase activity, disrupted cellular signaling, changes in protein interactions, impacts on neuronal health, and potential contributions to neurodegenerative processes, particularly in the context of Parkinson's disease.</p>
4.	"What are the predicted downstream gene interactions if gene PSEN1 is activated?"	<p>I have gathered information regarding the predicted downstream gene interactions when gene PSEN1 is activated. The interactions include both activation and inhibition relationships with various genes.</p> <p>Final Answer: When gene PSEN1 is activated, it is predicted to activate RYR3 and ITPR1, and inhibit EIF2AK3, ERN1, and ATF6. Additionally, it activates GSK3B.</p>
5.	"What are the predicted downstream gene interactions if gene PSEN1 is repressed?"	<p>The results indicate that when gene PSEN1 is repressed, it is predicted to interact with several downstream genes. Specifically, the interactions include:</p> <ol style="list-style-type: none"> 1. **RYR3** (activation) 2. **ITPR1** (activation) 3. **EIF2AK3** (inhibition) 4. **ERN1** (inhibition)

		<p>5. **ATF6** (inhibition) 6. **GSK3B** (activation)</p> <p>These interactions suggest that the repression of PSEN1 may lead to changes in the activity of these downstream genes, either activating or inhibiting their functions.</p> <p>Final Answer: The predicted downstream gene interactions if gene PSEN1 is repressed include RYR3 (activation), ITPR1 (activation), EIF2AK3 (inhibition), ERN1 (inhibition), ATF6 (inhibition), and GSK3B (activation).</p>
6.	<p>"What is the combined impact on biological processes and diseases if gene PSEN1 is activated and RYR3 is repressed?"</p>	<p>To summarize the findings regarding the combined impact of activating PSEN1 and repressing RYR3 on biological processes and diseases:</p> <p>1. **PSEN1 Activation**:</p> <ul style="list-style-type: none"> - PSEN1 is associated with numerous biological processes, particularly in the context of Alzheimer's disease. It interacts with various proteins and is involved in pathways related to cell signaling, apoptosis, and synaptic function. The activation of PSEN1 is linked to the modulation of several pathways that can influence neurodegeneration and cognitive functions. <p>2. **RYR3 Repression**:</p> <ul style="list-style-type: none"> - RYR3 is involved in calcium signaling and muscle contraction, and its repression may affect calcium homeostasis and signaling pathways. The gene is also implicated in various biological processes, including neurotransmitter release and muscle function. Repressing RYR3 could potentially lead to altered calcium signaling, which may have downstream effects on neuronal excitability and synaptic transmission. <p>3. **Combined Impact**:</p> <ul style="list-style-type: none"> - The activation of PSEN1 alongside the repression of RYR3 could lead to a complex interplay of signaling pathways. While PSEN1 activation may promote certain neuroprotective pathways, the repression of RYR3 could disrupt calcium signaling, potentially leading to adverse effects on neuronal health and function. This combination may exacerbate the pathological processes associated with Alzheimer's disease, as both genes are implicated in neurodegeneration. <p>In conclusion, the combined impact of activating PSEN1 and repressing RYR3 may lead to a nuanced effect on biological processes and diseases, particularly in the context of Alzheimer's disease, where both genes play significant roles.</p> <p>Final Answer: The combined impact of activating PSEN1 and</p>

		repressing RYR3 may exacerbate neurodegenerative processes associated with Alzheimer's disease, as PSEN1 activation promotes neuroprotective pathways while RYR3 repression disrupts calcium signaling, potentially leading to adverse effects on neuronal health and function.
--	--	---

Limitations

1. I have not included all the columns from the GAF file into the GO term nodes in the knowledge graph. I think adding additional relevant columns as attributes to the nodes could lead to richer responses from the agent
2. I am filtering the GO terms by the genes which I have found in the KGML files for each disease
3. The names of the gene, pathway and compound nodes are picked from the <graphics> tag inside the <entry> tag. Since it's usually a list of symbols, I found out that the first symbol is supposed to be the primary symbol for that gene, hence I selected that as the <name> for the gene. This could possibly be problematic as I don't have adequate domain knowledge to know what to do here
4. In the agent design I wasn't able to add output parsing properly using the AgentOutputParser class which was being used in a lot of examples I saw. With the ReAct prompting based agents, most examples used a standard prompt template and output template which in my case proved difficult as I built my custom prompt template for the queries above.

Possible extensions

1. Evidence prioritization is something that could be incorporated, so that certain edges in the graph are given a higher "score" compared to others, in order to steer the logic in the right direction. This can be introduced as an extension to the code.
2. The gene symbol synonyms or aliases have not been leveraged in the code, which could potentially lead to additional connections in the knowledge graph and result in richer hypothesis by the agent