

# SUPPLEMENTARY MATERIAL

---

## RegNetAgents: A Validated Multi-Agent AI Framework for Automated Gene Regulatory Network Analysis and Therapeutic Target Prioritization

**Authors:** Jose A. Bird, PhD

**Affiliation:** Bird AI Solutions

### Table of Contents

- [Table S1: Domain Agent LLM Prompt Templates](#table-s1-domain-agent-llm-prompt-templates)
- [S1.1 Cancer Biology Agent](#s11-cancer-biology-agent)
- [S1.2 Drug Development Agent](#s12-drug-development-agent)
- [S1.3 Clinical Medicine Agent](#s13-clinical-medicine-agent)
- [S1.4 Systems Biology Agent](#s14-systems-biology-agent)
- [Notes on Prompt Engineering](#notes-on-prompt-engineering)
- [Reproducibility Notes](#reproducibility-notes)

### Table S1: Domain Agent LLM Prompt Templates

RegNetAgents employs four specialized domain agents that use structured LLM prompts to generate scientific insights. Each agent receives gene-specific context (network topology, functional annotations, pathway enrichment) and returns structured JSON responses with domain-specific interpretations and rationales.

#### LLM Configuration:

- Model: llama3.1:8b (via Ollama local inference)
- Temperature: 0.3 (for scientific accuracy)
- Timeout: 90 seconds (for parallel multi-gene analysis)
- Retry logic: 2 attempts with graceful fallback to rule-based heuristics

#### Context Variables (populated dynamically for each gene):

- {gene}: Gene symbol (e.g., "TP53", "MYC")

- {function\_context}: NCBI/UniProt functional description
- {gene\_info}: Network topology (regulatory role, degree centrality)
- {enriched\_pathways}: Reactome pathway enrichment results (when available)
- {regulator\_count}, {target\_count}: Cascade analysis results
- {tissue\_context}: Cross-cell-type activity
- {pagerank}: PageRank centrality score

### S1.1 Cancer Biology Agent

#### System Prompt:

*You are an expert cancer biologist analyzing gene regulatory networks. Provide scientifically accurate, evidence-based analysis in structured JSON format only.*

#### User Prompt Template:

*Analyze the gene {gene} from a cancer biology perspective. {function\_context}*

*Gene Network Context:*

- *Regulatory Role: {regulatory\_role}*
- *Upstream Regulators: {num\_regulators}*
- *Downstream Targets: {num\_targets}*
- *Network Position: in-degree={num\_regulators}, out-degree={num\_targets}*

*Enriched Pathways: {enriched\_pathways}*

*Provide a cancer biology analysis in this EXACT JSON format:*

```
{ "oncogenic_potential": "high/moderate/low", "oncogenic_rationale": "brief scientific
explanation based on network topology and cancer biology", "tumor_suppressor_likelihood":
"high/moderate/low", "tumor_suppressor_rationale": "brief scientific explanation",
"therapeutic_target_score": 0.0-1.0, "therapeutic_rationale": "explanation of druggability and
therapeutic potential", "cancer_pathways": ["pathway1", "pathway2"], "biomarker_potential":
"high/moderate/low", "biomarker_utility": "diagnostic/prognostic/predictive",
"biomarker_rationale": "explanation", "research_priority": "high/moderate/low", "summary":
"1-2 sentence synthesis of cancer relevance" }
```

*Base your analysis on:*

- *Network centrality (hub regulators are critical for cancer)*

- Regulatory control (highly regulated genes often tumor suppressors)
- Pathway involvement (cancer-related pathways)
- Known cancer biology principles

Provide only the JSON, no additional text.

#### Example Output Structure:

```
{ "oncogenic_potential": "high", "oncogenic_rationale": "TP53 acts as a hub regulator with 163 downstream targets, indicating strong regulatory influence characteristic of oncogenes when dysregulated.", "tumor_suppressor_likelihood": "high", "tumor_suppressor_rationale": "Heavily regulated by 7 upstream factors, consistent with checkpoint function.", "therapeutic_target_score": 0.8, "llm_powered": true }
```

## S1.2 Drug Development Agent

### System Prompt:

You are an expert in drug discovery and development analyzing gene regulatory networks. Provide scientifically accurate, evidence-based analysis in structured JSON format only.

### User Prompt Template:

Analyze the gene {gene} from a drug development perspective. {function\_context}

Gene Network Context:

- Regulatory Role: {regulatory\_role}
- Upstream Regulators: {num\_regulators}
- Downstream Targets: {num\_targets}
- Total Regulators Found: {regulator\_count}
- Total Targets Found: {target\_count}

Provide drug development analysis in this EXACT JSON format:

```
{ "druggability_score": 0.0-1.0, "druggability_rationale": "explanation of druggability based on structure and network", "target_class": "kinase/GPCR/transcription_factor/nuclear_receptor/other", "intervention_strategy": "inhibition/activation/modulation/allosteric", "intervention_rationale": "why this strategy is appropriate", "development_complexity": "high/moderate/low", "cascade_effects": ["effect1", "effect2"], "clinical_trial_readiness": "ready/needs_preclinical/needs_research/not_suitable", "development_timeline": "estimated years", "summary": "1-2 sentence synthesis of drug development potential" }
```

*Base analysis on:*

- *Network topology (hub genes may have broad effects)*
- *Regulatory control (heavily regulated may be indirect targets)*
- *Cascade effects (downstream impact)*
- *Known drug target classes*

*Provide only the JSON, no additional text.*

### **S1.3 Clinical Medicine Agent**

#### **System Prompt:**

*You are an expert clinician and translational researcher analyzing gene networks for precision medicine applications.*

#### **User Prompt Template:**

*Analyze the gene {gene} from a clinical medicine and personalized healthcare perspective.  
{function\_context}*

*Gene Network Context:*

- *Regulatory Role: {regulatory\_role}*
- *Upstream Regulators: {num\_regulators}*
- *Downstream Targets: {num\_targets}*
- *Tissue Distribution: {tissue\_context}*

*Provide a clinical analysis in this EXACT JSON format:*

```
{ "disease_association_likelihood": "high/moderate/low", "disease_rationale": "brief  
explanation of disease relevance", "biomarker_utility":  
"diagnostic/prognostic/predictive/therapeutic", "biomarker_rationale": "brief explanation of  
biomarker potential", "clinical_actionability": "high/moderate/low", "actionability_rationale":  
"brief explanation of clinical utility", "tissue_specificity": "tissue-  
specific/broadly_expressed/ubiquitous", "diagnostic_potential": "high/moderate/low",  
"summary": "1-2 sentence clinical significance summary" }
```

*Focus on:*

- *Disease association potential based on network position*
- *Biomarker utility for diagnosis, prognosis, or therapeutic monitoring*
- *Clinical actionability and translational potential*

- Tissue specificity implications for personalized medicine

*Provide only the JSON, no additional text.*

## **S1.4 Systems Biology Agent**

### **System Prompt:**

*You are an expert systems biologist analyzing gene regulatory networks. Provide scientifically accurate, evidence-based analysis in structured JSON format only.*

### **User Prompt Template:**

*Analyze the gene {gene} from a systems biology and network theory perspective.  
{function\_context}*

*Gene Network Context:*

- Regulatory Role: {regulatory\_role}
- Upstream Regulators: {num\_regulators} (total in cascade: {reg\_count})
- Downstream Targets: {num\_targets} (total in cascade: {target\_count})
- PageRank Centrality: {pagerank}
- Total Network Degree: {num\_regulators + num\_targets}

*Provide a systems biology analysis in this EXACT JSON format:*

```
{ "network_centrality": 0.0-1.0, "centrality_rationale": "brief explanation of network position",  
  "regulatory_hierarchy": "master|hub|intermediate|peripheral", "hierarchy_rationale": "brief  
explanation of hierarchical position", "information_flow": "high|moderate|low",  
  "flow_rationale": "brief explanation of information processing", "network_vulnerability":  
  "critical|important|moderate|minimal", "vulnerability_rationale": "brief explanation of  
network impact", "perturbation_impact": "system-wide|modular|localized|minimal",  
  "perturbation_rationale": "brief explanation of knockout/perturbation effects",  
  "evolutionary_conservation": "high|moderate|low", "conservation_rationale": "brief inference  
about evolutionary importance", "summary": "1-2 sentence systems biology summary" }
```

*Focus on:*

- Network topology and centrality (degree, betweenness, PageRank implications)
- Hierarchical position and regulatory control
- Information flow and signal transduction
- Network robustness and vulnerability to perturbation

- Evolutionary conservation inferred from network position

*Provide only the JSON, no additional text.*

## Notes on Prompt Engineering

### Structured Output Constraints:

All prompts explicitly request JSON-only responses with predefined keys and value constraints (e.g., "high|moderate|low", "0.0-1.0"). This ensures:

1. Parseable structured output for downstream integration
2. Consistent classification schemes across genes
3. Minimal hallucination through constrained response space
4. Graceful error handling via JSON validation

### Context Integration:

Each prompt incorporates multiple information sources:

- Gene functional annotations (UniProt/NCBI) provide biological grounding
- Network topology metrics constrain interpretations to data-supported claims
- Pathway enrichment results (when available) add mechanistic context
- Cross-cell-type analysis provides tissue specificity

### Fallback Mechanism:

If LLM calls fail (timeout, parsing error, unavailable service), agents automatically fall back to rule-based heuristics using network topology thresholds:

- Oncogenic potential: >50 targets = high, 20-50 = moderate, <20 = low
- Tumor suppressor likelihood: >10 regulators = high (checkpoint function)
- Druggability: Hubs = complex (off-target risks), heavily regulated = lower priority

Each analysis result includes an `llm_powered: true/false` flag for transparency.

### Temperature Selection:

Temperature = 0.3 balances determinism (reproducibility) with linguistic variety (natural rationale text). Lower temperatures (0.0-0.2) produced overly repetitive text; higher temperatures (0.5-1.0) increased inconsistency across runs.

## Reproducibility Notes

### To reproduce LLM analyses:

1. Install Ollama: <https://ollama.com/download>
2. Pull model: `ollama pull llama3.1:8b`
3. Configure environment variables:
  - `OLLAMA_HOST=http://localhost:11434`
  - `OLLAMA_MODEL=llama3.1:8b`
  - `OLLAMA_TEMPERATURE=0.3`
  - `OLLAMA_TIMEOUT=90`
  - `USE_LLM_AGENTS=true`
4. Run analysis via MCP server or Python API

### LLM output variability:

Due to temperature-based sampling, exact wording of rationales will vary across runs. However, core classifications (high/moderate/low, diagnostic/prognostic) remain consistent. For deterministic analysis, use rule-based mode (`USE_LLM_AGENTS=false`).

**Network metrics (PageRank, degree centrality) are deterministic** and fully reproducible across all runs.

### Code Availability:

Complete implementation available at: <https://github.com/jab57/RegNetAgents>