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

Running Title: Multi-Agent Gene Regulatory Network Analysis

Authors

Jose A. Bird, PhD

Bird AI Solutions

jbird@birdaisolutions.com

ABSTRACT

Gene regulatory network analysis is essential for understanding disease mechanisms, identifying biomarkers, and prioritizing therapeutic targets. However, traditional workflows require labor-intensive manual effort across multiple databases with limited scalability. We present RegNetAgents, an LLM-powered multi-agent framework that automates gene regulatory analysis through intelligent workflow orchestration. The system integrates network modeling, perturbation simulation, pathway enrichment, and multi-domain interpretation into a conversational interface accessible via Claude Desktop without programming expertise.

Built on ARACNe networks from 500,000+ single cells across 10 cell types, the framework deploys four specialized agents executing in parallel with rule-based fallback for reliability. Automated perturbation analysis ranks candidate therapeutic targets using network centrality metrics (PageRank, degree centrality).

In a demonstration study of five colorectal cancer biomarkers (MYC, CTNNB1, CCND1, TP53, KRAS), framework classifications showed complete concordance with published literature across this limited sample. Perturbation analysis identified literature-supported TP53 interactors (WWTR1, YAP1, CHD4) among the top-ranked network neighbors based on topology. Additional high-ranking regulators (RBPMS, PRRX2, THRA, IKZF2) represent novel hypotheses prioritized for experimental validation.

Complete analysis of 99 regulators across 5 genes completed in 15-62 seconds, representing orders of magnitude speedup over manual workflows. RegNetAgents

transforms labor-intensive analysis into second-scale automated hypothesis generation accessible to experimental biologists.

Keywords: gene regulatory networks, single-cell RNA-seq, multi-agent systems, workflow orchestration, biomarker discovery, therapeutic target identification, network centrality, PageRank, LangGraph, Model Context Protocol

INTRODUCTION

Background

Gene regulatory networks (GRNs) represent the complex interactions between transcription factors and their target genes, forming the fundamental control systems that determine cellular identity, function, and disease states (1,2). Understanding these networks is critical for identifying disease biomarkers, elucidating disease mechanisms, and discovering therapeutic targets (3,4). Single-cell RNA sequencing (scRNA-seq) technologies have enabled reconstruction of cell-type-specific regulatory networks at unprecedented resolution, revealing how regulatory architecture varies across tissues and cell types (5,6).

Traditional gene regulatory analysis requires researchers to manually query multiple databases and tools: network databases such as STRING (7) and BioGRID (8) for interaction data, pathway enrichment tools such as Enrichr (9) and DAVID (10) for functional annotation, and extensive literature curation for domain-specific context. This fragmented workflow presents several critical limitations. First, the process is labor-intensive and time-consuming, requiring researchers to navigate different web interfaces, export data, and manually integrate findings across domains for each gene analyzed. Second, sequential processing limits scalability—analyzing multiple genes or comparing across cell types requires repeating the entire workflow. Third, domain expertise remains siloed, with cancer relevance, drug development potential, clinical actionability, and systems-level effects analyzed separately rather than in an integrated framework. Finally, the lack of conversational interfaces necessitates technical expertise, limiting accessibility to researchers without computational backgrounds.

Recent advances in artificial intelligence and workflow orchestration present opportunities to address these limitations. Multi-agent systems enable parallel execution of specialized analysis components (11), while graph-based workflow engines such as LangGraph provide intelligent routing and state management for complex analytical pipelines (12). The Model Context Protocol (MCP) enables integration of computational tools with conversational AI interfaces (13), potentially bridging the accessibility gap for non-computational researchers.

Regulatory Network Data Sources

Cell-type-specific regulatory networks were obtained as pre-computed ARACNe networks prepared and released by the GREmLN team (14). GREmLN (Gene Regulatory Embedding-based Large Neural model) is a foundation model that embeds gene regulatory network

structure directly within its attention mechanism, trained on 11 million scRNA-seq profiles across 162 cell types. The GREmLN team processed single-cell RNA-seq data from the CELLxGENE Data Portal (15) using the ARACNe (Algorithm for the Reconstruction of Accurate Cellular Networks) algorithm (16,17) and publicly released these pre-computed networks. Our RegNetAgents framework leverages the GREmLN team's pre-computed ARACNe networks as input data for multi-agent analysis; we do not use the GREmLN foundation model itself.

ARACNe Network Inference: ARACNe-AP (Adaptive Partitioning) uses mutual information to identify direct regulatory interactions while eliminating indirect associations through data processing inequality. The algorithm computes pairwise mutual information scores between genes, applies statistical significance testing (p-value threshold: 1e-8), and removes indirect edges using DPI to produce high-confidence transcription factor-target relationships. Networks were generated from metacell-aggregated expression matrices (5 cells per metacell) using the top 1,024 highly variable genes per cell type. The resulting networks are provided as tab-separated value files with columns for regulator gene ID, target gene ID, mutual information score, Spearman correlation coefficient, bootstrap count, and log-transformed p-values.

Network Processing: We obtained pre-computed ARACNe networks (network.tsv format) from the GREmLN Quickstart Tutorial and converted them to optimized NetworkX-compatible pickle caches (network_index.pkl) for subsecond loading. We utilized 10 cell types spanning immune cells (CD14 monocytes, CD16 monocytes, CD20 B cells, CD4 T cells, CD8 T cells, NK cells, NKT cells, monocyte-derived dendritic cells), blood cells (erythrocytes), and epithelial tissue, representing over 500,000 single cells with networks containing up to 183,247 regulatory edges for epithelial cells.

Our Contribution

We present RegNetAgents, a multi-agent AI framework that streamlines gene regulatory network analysis and makes sophisticated computational methods accessible to experimental biologists. The system integrates network analysis, pathway enrichment, perturbation simulation, and multi-domain interpretation into a unified, conversational interface. Key contributions include: (1) **Workflow automation**: parallel multi-agent execution enabling second-scale analysis that replaces multi-hour manual workflows, (2) **Hypothesis generation**: automated perturbation analysis that simulates regulator inhibition to identify and rank candidate therapeutic targets for experimental validation, (3) **Multi-domain integration**: four specialized domain agents providing cancer, drug development, clinical, and systems biology perspectives in a single query, (4) **Accessibility**: natural language interface via Model Context Protocol enabling conversational access without programming expertise, and (5) **Demonstration**: case study on colorectal cancer biomarkers showing the framework can recapitulate literature-confirmed regulatory relationships and generate testable hypotheses. The framework is designed as a hypothesis generation and experimental prioritization tool, not a replacement for experimental validation.

Validation Strategy

To demonstrate framework capabilities and validate analytical accuracy, we analyzed a well-characterized colorectal cancer biomarker panel (MYC, CTNNB1, CCND1, TP53, KRAS) where regulatory relationships are established in literature. This approach allows comparison of framework predictions against known biology—validating the tool produces sensible results while generating testable hypotheses for experimental follow-up. We use this case study not to claim novel biological discoveries, but to demonstrate the framework can recapitulate established regulatory patterns and prioritize candidates for experimental validation.

METHODS

Architecture Overview

RegNetAgents implements a multi-agent architecture using LangGraph for workflow orchestration (Figure 1). The system consists of five agent types organized in a directed acyclic graph: initialization/validation, network modeling, pathway enrichment, domain analysis (four parallel agents), and integration/reporting. Each agent operates on a shared state object containing analysis parameters, intermediate results, and domain-specific insights.

The workflow begins with query parsing and gene identifier validation, converting gene symbols to Ensembl IDs for consistent network lookups. Network modeling identifies upstream regulators and downstream targets from pre-computed ARACNe networks, classifying genes into regulatory roles (hub regulator, heavily regulated, intermediate node, regulator, weakly regulated) based on in-degree and out-degree. Genes with more than five upstream regulators trigger automated perturbation analysis. Pathway enrichment queries the Reactome database API for statistical enrichment of affected pathways. Four domain agents—cancer biology, drug discovery, clinical relevance, and systems biology—execute in parallel, each providing specialized interpretation of network position, regulatory patterns, and biological significance. Finally, an integration agent synthesizes findings across domains into a comprehensive report with ranked insights and experimental recommendations.

Data Sources and Network Construction

Regulatory Networks

Cell-type-specific regulatory networks were obtained as pre-computed ARACNe networks prepared by the GREmLN team (Zhang et al. 2025), which processed single-cell RNA-seq datasets from the CELLxGENE Data Portal (Census release: 2024-07-01) using the ARACNe-AP algorithm. For each cell type, the top 1,024 highly variable genes were selected and metacell aggregation (5 cells per metacell) was performed to reduce technical noise. ARACNe-AP parameters: mutual information estimation via adaptive partitioning, data processing inequality (DPI) tolerance of 1.0 to eliminate indirect edges, p-value threshold of

1e-8 for statistical significance, and 100 bootstrap iterations for edge robustness assessment.

Networks are provided in tab-separated value (TSV) format with the following structure: regulator gene ID (Ensembl), target gene ID (Ensembl), mutual information score (MI), Spearman correlation coefficient (SCC), bootstrap count, and log-transformed p-value. We downloaded pre-computed networks from the GREmLN Quickstart Tutorial (https://virtualcellmodels.cziscience.com/quickstart/gremln-quickstart) and converted them to NetworkX-compatible pickle caches for rapid querying. Network statistics: epithelial cells (183,247 edges), CD8 T cells (3,154 edges), monocyte-derived dendritic cells (5,317 edges), erythrocytes (19,398 edges), NKT cells (2,509 edges), CD14 monocytes (2,009 edges), CD16 monocytes (1,236 edges), CD4 T cells (1,371 edges), CD20 B cells (1,128 edges), and NK cells (404 edges). Gene identifiers use Ensembl IDs (GRCh38) with bidirectional mapping to HGNC gene symbols for user queries.

Pathway Annotation

Pathway enrichment analysis uses the Reactome Pathway Database (https://reactome.org), a manually curated, peer-reviewed database of human biological pathways (18). Given a set of genes (query gene plus regulators and targets), we query the Reactome Analysis Service API (https://reactome.org/AnalysisService/) with POST requests containing gene lists. The API returns enriched pathways with statistical validation: p-values from hypergeometric tests and false discovery rate (FDR) corrections via Benjamini-Hochberg method. We report pathways with FDR < 0.05 as significantly enriched.

For pathway enrichment, gene lists are constructed from the query gene plus its top 10 upstream regulators and top 10 downstream targets (ranked by network centrality). This focused approach balances biological signal with statistical specificity: limiting to the immediate regulatory neighborhood captures direct mechanistic relationships while preventing pathway over-enrichment that occurs with large gene sets. Reactome enrichment analysis performs optimally with focused gene sets of 10-50 genes (as recommended in Reactome documentation); larger gene lists dilute specificity by matching too many broad pathways, while smaller lists lack statistical power. For hub regulators like MYC (427 targets), using all targets would produce overly general pathway enrichment; the top 10 approach focuses on the strongest regulatory relationships (highest PageRank and degree centrality) most likely to drive functional effects.

Gene Annotation

Gene-level annotations including full names and functional descriptions are retrieved from local NCBI and UniProt databases. Gene identifier conversion between gene symbols and Ensembl IDs uses the Ensembl REST API (https://rest.ensembl.org) (19) with local caching for performance optimization.

Multi-Agent Workflow

Network Analysis Agent

The network modeling agent performs five core functions: (1) gene identifier resolution (symbol to Ensembl ID conversion with synonym matching), (2) network lookup to retrieve all regulatory edges involving the query gene, (3) regulator identification (genes with edges pointing to the query gene), (4) target identification (genes the query gene regulates), and (5) regulatory role classification based on network topology.

Regulatory roles are assigned algorithmically based on network topology:

- **Hub regulator**: >20 downstream targets (high regulatory influence)

- **Heavily regulated**: >15 upstream regulators and ≤20 targets (complex regulatory control)

- **Intermediate node**: >5 targets AND >5 regulators (balanced regulatory role)

- **Regulator**: >0 targets but ≤5 regulators (modest regulatory activity)

- **Weakly regulated**: 0 targets and ≤15 regulators (simple endpoint)

These exploratory thresholds prioritize hub status over heavily-regulated status when genes meet both criteria, ensuring genes with extensive downstream influence are classified as hubs regardless of their input complexity. Network position metrics include in-degree (number of regulators), out-degree (number of targets), and regulatory role classification.

Perturbation Analysis Agent

For genes with five or more upstream regulators, we perform automated perturbation analysis to identify potential therapeutic targets. This analysis simulates inhibiting each upstream regulator individually and evaluates therapeutic potential using standard network centrality metrics.

Network Centrality Metrics for Therapeutic Target Ranking:

We calculate three core centrality measures for each regulator R using NetworkX (20) implementations:

Degree Centrality:

$C_D(R) = deg(R) / (N - 1)$

Where deg(R) is the total number of connections (in-degree + out-degree) and N is the total number of nodes in the network. Measures overall network connectivity.

Out-Degree Centrality (Primary Hub Metric):

$C_{out}(R) = deg_{out}(R) / (N - 1)$

Where deg_out(R) is the number of downstream targets regulated by R. This metric identifies hub regulators with broad downstream effects and estimates potential off-target impacts of therapeutic intervention.

PageRank (Primary Ranking Metric):

$PR(R) = (1-\alpha)/N + \alpha \times \Sigma[PR(v) / L(v)]$ for all v in M(R)

Where M(R) is the set of nodes with edges pointing to R, L(v) is the out-degree of node v (number of outbound edges), $\alpha = 0.85$ is the damping factor, and N is the total number of nodes. For large networks (≥1,000 nodes), PageRank is normalized by dividing by the maximum PageRank value to ensure interpretability. This is Google's algorithm adapted for biological networks, measuring connection quality rather than quantity (21).

Ranking and Interpretation:

Regulators are ranked by PageRank (primary), as this metric was identified as the best predictor of successful drug targets in protein interaction networks (22). We also provide alternative rankings by out-degree centrality for comparison. PageRank differentiates therapeutic potential even when regulators contribute equally to target gene regulation. According to Mora & Donaldson (2021), approved drug targets show significantly higher PageRank and degree centrality compared to non-targets (22).

For each regulator, we report:

- Network centrality metrics (PageRank, degree centrality, out-degree centrality)

- Number of downstream targets (off-target effect estimate)

- Cascade overlap count (shared regulatory targets with target gene)

- Top 5 affected cascades (genes regulated by both the inhibited regulator and the target gene)

Therapeutic Interpretation:

High PageRank (>0.30) combined with hub regulator status (>200 targets) indicates strong therapeutic potential but requires consideration of potential off-target effects. All regulators of a given target gene contribute equally to direct regulatory input (1/num_regulators), so centrality metrics differentiate therapeutic potential based on network position and influence.

Perturbation Analysis Limitations:

Our perturbation analysis makes several simplifying assumptions that warrant consideration:

1. **Topology-based ranking**: Analysis ranks regulators based on network topology (connectivity, cascade overlap) rather than predicting quantitative gene expression changes. This approach identifies regulators for experimental validation but does not model dynamic regulatory responses or predict expression fold-changes.

2. **Additive regulatory effects**: The analysis assumes regulators contribute independently to target gene regulation (equal contribution = 1/num_regulators for each regulator). In reality, regulatory effects may be synergistic, antagonistic, or context-dependent. Combinatorial regulatory logic (e.g., AND/OR gates, feed-forward loops) is not explicitly modeled.

3. **Chromatin context**: ARACNe networks capture steady-state correlation patterns but do not incorporate chromatin accessibility, histone modifications, or DNA methylation states. Regulatory potential predicted from network topology may differ from regulatory activity in specific chromatin contexts or differentiation states.

4. **Static network assumption**: Networks represent time-averaged regulatory relationships and do not capture dynamic rewiring during development, differentiation, or disease progression. Regulator importance may vary across cellular states not represented in the averaged network.

Despite these limitations, topology-based perturbation analysis can identify literature-supported functional interactors (as demonstrated with TP53 Hippo pathway effectors) and provides a principled ranking for experimental prioritization. Results should be interpreted as hypotheses for validation, not predictions of molecular outcomes. Experimental validation remains essential to confirm regulatory relationships and quantify expression changes.

Pathway Enrichment Agent

The pathway agent constructs a gene set consisting of the query gene, all upstream regulators, and a sample of downstream targets (up to 50 genes to optimize API performance). This gene set is submitted to Reactome's over-representation analysis endpoint, which performs hypergeometric tests against all curated pathways in the database.

Reactome returns enriched pathways ranked by statistical significance, including:

- Pathway ID and name

- P-value (uncorrected)

- False discovery rate (FDR, Benjamini-Hochberg correction)

- Number of query genes in pathway

- Total pathway size

- Species (Homo sapiens)

We report all pathways with FDR < 0.05 as significantly enriched. For queries returning fewer than 3 enriched pathways at FDR < 0.05, we relax to p < 0.05 uncorrected to provide exploratory pathway associations.

Domain Analysis Agents

Four domain-specific agents execute in parallel, each providing specialized biological interpretation. **Agents are LLM-powered by default**, using local language model inference (Ollama with llama3.1:8b) to generate scientific insights with rationales. The system implements graceful fallback to rule-based heuristics if LLM is unavailable, ensuring reliability. Each analysis result includes an `llm_powered: true/false` flag for transparency.

LLM-Powered Mode (Primary):

Each domain agent receives structured prompts containing:

- Gene functional descriptions from NCBI/UniProt databases (providing validated biological context)

- Gene regulatory context (role, upstream regulators, downstream targets, PageRank)

- Pathway enrichment results (when available)

- Cross-cell-type expression patterns (for clinical agent)

Complete prompt templates for all four domain agents are provided in **Supplementary Table S1**, enabling full reproducibility of LLM analyses.

Agents return JSON-structured insights with:

- Domain-specific classifications (e.g., oncogenic potential: high/moderate/low, intervention strategy: inhibition/activation)

- Scientific rationales explaining each classification

- Network topology metrics (PageRank, degree centrality)

- 1-2 sentence summaries for integration

LLM prompts request specific JSON formats with predefined keys. Response parsing includes validation, missing key detection with placeholder insertion, and classification verification. Retry logic (2 attempts) handles transient failures.

Rule-Based Fallback (Reliability):

If LLM is unavailable or fails, agents use fast heuristic algorithms that classify genes based on network topology:

Cancer Biology Agent: Classifies oncogenic potential based on network centrality (>50 targets = high, >20 = moderate, <20 = low), tumor suppressor likelihood (heavily regulated genes with >10 regulators = high), and biomarker potential (diagnostic for heavily regulated genes, prognostic for hub regulators, predictive based on regulatory architecture). Classifications are heuristics derived from network position and regulatory role.

Drug Development Agent: Identifies intervention strategies based on regulatory architecture (hub regulators = inhibition candidates, heavily regulated genes = activation candidates), estimates development complexity (high for hubs due to potential off-target effects, moderate for intermediate nodes, low for heavily regulated genes), and provides qualitative assessments of downstream cascade effects based on target counts.

Clinical Relevance Agent: Classifies biomarker utility (diagnostic for heavily regulated genes with pathway involvement, prognostic for regulators with cancer pathway enrichment, predictive for regulators in druggable pathways), assesses tissue specificity from cross-cell-type regulatory patterns, and provides qualitative disease association estimates based on regulatory architecture.

Systems Biology Agent: Quantifies network centrality metrics (PageRank, out-degree, in-degree), classifies regulatory hierarchy position (hub/intermediate/heavily-regulated), and estimates perturbation impact scope (system-wide for hubs, localized for heavily regulated genes) based on network topology.

Classification Rationale: Agent classifications are derived from network topology thresholds empirically chosen to reflect regulatory roles: (1) Oncogenic potential: >50 targets = high (strong regulatory influence), 20-50 = moderate, <20 = low; (2) Tumor suppressor likelihood: heavily regulated genes with >10 regulators = high (multiple regulatory controls suggest gatekeeper function); (3) Intervention complexity: hub regulators = high (many downstream targets create off-target risks), intermediates = moderate, heavily regulated = low; (4) Biomarker utility: heavily regulated genes = diagnostic candidates (pathway outputs), hub regulators = prognostic candidates (pathway controllers). These thresholds are exploratory heuristics chosen for demonstration purposes and have not been systematically validated against benchmark datasets (e.g., COSMIC Cancer Gene Census, DGIdb). They provide qualitative guidance for experimental prioritization but should not be interpreted as validated predictive scores. Future work should optimize these thresholds through systematic comparison with known cancer genes and druggable targets. Network centrality metrics (PageRank, degree centrality) remain the primary validated quantitative outputs, with perturbation rankings confirmed against experimental literature.

Performance: LLM mode adds ~10-12 seconds per gene (4 agents in parallel, ~2.5-3s each) compared to rule-based mode. For 5-gene analysis: ~6 seconds (rule-based) vs ~45-62 seconds (LLM). Both modes remain orders of magnitude faster than manual literature review (hours to days).

LLM Reproducibility: LLM-generated outputs exhibit variability across runs due to temperature-based sampling (temperature = 0.3 for llama3.1:8b). Core classifications (oncogenic potential, intervention strategies) remain consistent across runs, but specific wording in rationales may vary. For deterministic analysis, rule-based mode provides identical results on repeated execution. JSON structure validation ensures all required fields are present regardless of LLM output variability. For the analyses reported in this manuscript, we used rule-based mode for quantitative network metrics (perturbation rankings, PageRank scores, degree centrality) and LLM mode for qualitative biological interpretations (domain-specific rationales). All quantitative network metrics are deterministic and fully reproducible.

Domain agents operate independently and can be extended or replaced without affecting other workflow components.

Integration and Reporting Agent

The integration agent synthesizes results from all upstream agents into a comprehensive analysis report. Output format is structured JSON containing:

- Gene analysis summary (gene, cell type, regulatory role)

- Network analysis (regulators, targets, network position)

- Perturbation analysis results (if applicable, with ranked therapeutic targets)

- Pathway enrichment (significant pathways with FDR values)

- Cross-cell-type comparison (regulatory role across all 10 cell types)

- Domain-specific insights (cancer, drug, clinical, systems perspectives)

- Key insights summary (aggregated metrics and classifications)

- Workflow metadata (execution time, completed steps)

For multi-gene queries, the integration agent additionally generates comparative analyses and identifies common regulators, shared pathways, and complementary biomarker panels.

Model Context Protocol Integration

The RegNetAgents workflow is exposed via a Model Context Protocol (MCP) server, enabling conversational access through Claude Desktop. The MCP server implements eight tools:

1. **comprehensive_gene_analysis**: Full workflow execution (network + regulators + targets + perturbation + pathways + domains)

2. **multi_gene_analysis**: Parallel processing of multiple genes with comparative analysis

3. **pathway_focused_analysis**: Pathway enrichment for user-specified gene lists

4. **analyze_regulators**: Detailed upstream regulator analysis with perturbation

5. **analyze_targets**: Detailed downstream target analysis with cascade effects

6. **cross_cell_comparison**: Regulatory role comparison across all 10 cell types

7. **workflow_status**: Real-time execution monitoring

8. **workflow_insights**: Performance analytics and routing decisions

Each MCP tool accepts structured parameters (gene symbols, cell types, analysis depth) and returns JSON-formatted results. The LangGraph workflow engine operates independently of the MCP layer, enabling the same analytical pipeline to be deployed via command-line interfaces, REST APIs, or computational notebooks.

Implementation Details

The system is implemented in Python 3.8+ using LangGraph (v0.1+) for workflow orchestration. Network data structures use dictionaries with Ensembl IDs as keys for O(1) lookup performance. Gene ID mapping is cached in memory (pickle serialization) for instant symbol-to-Ensembl conversion. Network indices are pre-computed and loaded at server startup to minimize query latency.

External API calls (Reactome, Ensembl REST API) use asynchronous requests with 10-second timeouts and exponential backoff retry logic. Parallel agent execution uses Python's concurrent.futures ThreadPoolExecutor with a maximum of 4 concurrent threads.

Performance Benchmarking

We measured end-to-end execution time for three analysis scenarios on a standard laptop (Intel i7, 16GB RAM, Windows 11):

1. **Single gene comprehensive analysis**: TP53 in epithelial cells (all workflow steps)

2. **Multi-gene analysis**: 5 genes (MYC, CTNNB1, CCND1, TP53, KRAS) in epithelial cells

3. **Cross-cell-type comparison**: TP53 across all 10 cell types

Each scenario was executed 10 times with cold starts (server restarted between runs). Execution time includes all workflow steps from query initiation to final JSON output, including Reactome API calls.

Data Availability

Regulatory network data were obtained from the GREmLN foundation model (Zhang et al. 2025, bioRxiv 2025.07.03.663009). Preprocessed ARACNe networks for 10 cell types (CD14 monocytes, CD16 monocytes, CD20 B cells, CD4 T cells, CD8 T cells, NK cells, NKT cells, monocyte-derived dendritic cells, erythrocytes, and epithelial cells) are publicly available

through the GREmLN Quickstart Tutorial at https://virtualcellmodels.cziscience.com/quickstart/gremln-quickstart. Networks are provided as TSV files downloaded via Google Drive in the tutorial materials. The underlying scRNA-seq data (11 million profiles across 162 cell types) were sourced from the CZ CELLxGENE Data Portal (Census release: 2024-07-01). Pathway annotations use the publicly accessible Reactome API (https://reactome.org/AnalysisService/). Gene-level annotations use local NCBI and UniProt databases. Gene identifier conversion uses the Ensembl REST API (https://rest.ensembl.org).

RESULTS

Validation Summary

Framework validation against established colorectal cancer biology demonstrated:

- **Complete concordance** with published biomarker classifications across limited sample (5/5 genes classified consistently with literature)

- **High-ranking identification** of literature-supported TP53 interactors from Hippo pathway (WWTR1, YAP1, CHD4) based on network topology

- **Pathway coherence** between network topology and biological pathways (Hippo signaling enrichment, FDR = 0.020)

- **Novel hypothesis generation** for experimental validation (RBPMS, PRRX2, THRA, IKZF2)

These results demonstrate the framework produces biologically meaningful outputs suitable for hypothesis generation and experimental prioritization.

System Performance

RegNetAgents achieves second-scale execution times for gene regulatory network analysis (Table 1). The system supports two execution modes: (1) rule-based mode for fast, deterministic analysis, and (2) LLM-powered mode with AI-generated scientific insights via local language model inference (Ollama/llama3.1:8b).

Rule-based mode provides rapid analysis without LLM dependencies. Single gene comprehensive analysis of TP53 in epithelial cells (network modeling, complete perturbation analysis of all 7 regulators, Reactome pathway enrichment) completed in 0.60 seconds. Multi-gene analysis of 5 genes (MYC, CTNNB1, CCND1, TP53, KRAS) including complete perturbation analysis of all 99 regulators completed in 15.49 seconds with parallel execution.

LLM-powered mode adds specialized domain analysis with scientific rationales from four parallel agents (cancer biology, drug discovery, clinical relevance, systems biology). Single gene comprehensive analysis with LLM insights averages ~15 seconds (4× slower than rule-based due to local LLM inference). Five-gene comprehensive analysis with parallel LLM

execution completes in ~62 seconds, providing AI-generated rationales for all genes simultaneously. Cross-cell-type comparison of TP53 across all 10 cell types remains instant (<0.01 seconds) in both modes, leveraging pre-computed network indices.

Performance breakdown: Network lookups from pre-computed indices are near-instantaneous (<1 ms), perturbation analysis of all regulators scales linearly with regulator count, Reactome API calls take 0.3-1.5 seconds per query, and LLM inference adds ~3-4 seconds per domain agent call. PageRank pre-computation in network cache enables instant perturbation analysis. Despite comprehensive regulator analysis and optional LLM insights, execution remains orders of magnitude faster than manual workflows, which typically involve sequential querying of network databases, pathway enrichment tools, and extensive literature curation.

Table 1. Performance Benchmarks

Analysis Type	Genes	Execution Time	Components
Focused (rule-based)	TP53	0.68 sec	Network lookup, regulators, targets, complete perturbation (all 7 regulators)
Comprehensive (rule-based)	TP53	0.60 sec	Network analysis + Reactome pathway enrichment (16 pathways, FDR correction)
Comprehensive (LLM-powered)	TP53	~15 sec	Network + perturbation + pathways + 4 LLM agents with scientific rationales
Multi-gene (rule-based)	5 genes	15.49 sec	Network analysis, complete perturbation (all 99 regulators, parallel execution)
Multi-gene (LLM-powered)	5 genes	~62 sec	Network + perturbation (99 regulators) + pathways + 4 parallel LLM agents
Cross-cell comparison	TP53, 10 types	<0.01 sec	Regulatory role comparison across all cell types (pre-computed indices)

Rule-based mode provides deterministic analysis without LLM dependencies. LLM-powered mode adds domain-specific insights via local Ollama inference (llama3.1:8b) with 4 parallel agents. PageRank pre-computation in network cache enables instant perturbation analysis. All measurements via Claude Desktop MCP integration on standard laptop (Intel i7, 16GB RAM) with stable internet connection and Ollama server running locally.

Workflow Automation and Performance

Traditional gene regulatory analysis requires manual querying of multiple databases and tools across fragmented workflows. For a single gene (e.g., TP53 in epithelial cells), researchers must: (1) query network databases (STRING, BioGRID) for regulators and targets (5-10 minutes), (2) perform pathway enrichment via web interfaces (Enrichr, Reactome; 10-15 minutes), (3) curate literature for domain-specific context (PubMed searches, paper review; 60-120 minutes), and (4) manually interpret network position and biological significance (30-60 minutes). This fragmented workflow requires 2-4 hours per gene and extensive copy-paste operations across tools.

RegNetAgents automates this entire workflow into a single natural language query ("Analyze TP53 in epithelial cells") via Claude Desktop. The multi-agent framework executes all steps in parallel: network modeling retrieves regulators/targets from pre-computed indices (<1ms), pathway enrichment queries Reactome API (0.3-0.5s), and four specialized LLM agents generate domain insights concurrently (cancer, drug, clinical, systems biology; 3-4s per agent in parallel). Total execution time: 0.6 seconds (rule-based mode) or ~15 seconds (LLM-powered mode), representing orders of magnitude speedup compared to manual workflows (Figure 4C).

This acceleration enables exploratory analyses previously impractical due to time constraints. For example, analyzing a 5-gene biomarker panel with complete perturbation analysis (99 total regulators) requires 15.49 seconds (rule-based) or ~62 seconds (LLM-powered), versus an estimated 10-20 hours of manual effort (2-4 hours per gene × 5 genes). The framework transforms gene regulatory analysis from a multi-hour undertaking into an interactive, conversational experience accessible to experimental biologists without computational expertise.

LLM-Powered Domain Analysis Adds Scientific Context

Beyond workflow automation, RegNetAgents demonstrates the value of integrating local language models into computational biology pipelines. The system operates in two modes: (1) rule-based mode using fast heuristic algorithms, and (2) LLM-powered mode using local Ollama inference (llama3.1:8b) with four specialized domain agents.

Rule-based mode provides qualitative classifications (oncogenic potential: high/moderate/low, intervention strategies, biomarker utility) based on network topology and regulatory patterns, completing in 0.6 seconds per gene. While fast and deterministic, this mode lacks scientific context explaining WHY classifications were assigned.

LLM-powered mode integrates gene functional descriptions from NCBI/UniProt databases with network context to generate domain-specific rationales. For example, for TP53, the cancer biology agent provides: "TP53 functions as a hub regulator (163 targets) consistent with its role as master tumor suppressor. High regulatory input (7 regulators) suggests multiple regulatory checkpoints controlling p53 activity, aligning with its critical gatekeeper function in genomic stability." This scientific interpretation adds ~14 seconds per gene (4 agents × 3-4s each, parallel execution) but provides experimentalists with actionable biological context beyond network metrics alone (Figure 4D).

Each analysis result includes an `llm_powered: true/false` flag for transparency, and the system gracefully falls back to rule-based mode if Ollama is unavailable, ensuring reliability. The LLM mode demonstrates how AI can augment computational biology workflows by providing interpretable, context-rich outputs that bridge the gap between numerical analysis and biological understanding.

Demonstration Case Study: Colorectal Cancer Biomarker Analysis

To demonstrate framework capabilities and validate analytical outputs against established biology, we analyzed a colorectal cancer (CRC) biomarker panel consisting of five genes with well-characterized CRC roles: MYC (transcriptional amplification oncogene), CTNNB1 (β-catenin, Wnt pathway effector), CCND1 (cyclin D1, cell cycle regulator), TP53 (tumor suppressor), and KRAS (oncogenic signaling). All genes were analyzed in the epithelial cell network, representing the tissue of origin for colorectal adenocarcinomas. This case study serves to illustrate how the framework processes gene queries, generates hypotheses, and enables comparison with published literature—not to claim novel biological discoveries.

Multi-Gene Regulatory Network Analysis

The five-gene panel exhibited distinct regulatory architectures (Table 2, Figure 2). Three genes emerged as hub regulators with extensive downstream connectivity: TP53 (163 targets, 7 regulators), MYC (427 targets, 25 regulators), and CTNNB1 (310 targets, 18 regulators), indicating central roles in signal amplification and oncogenic pathway activation. Two genes showed heavily regulated profiles with no identified downstream regulatory relationships in the epithelial network: CCND1 (0 targets, 42 regulators - extensively controlled) and KRAS (0 targets, 7 regulators), indicating they function as end-point effectors in signaling cascades.

Table 2. Colorectal Cancer Biomarker Panel Analysis

Gene	Regulatory Role	Targets	Regulators	Biomarker Type	Top Candidate Regulator (PageRank)
MYC	Hub Regulator	427	25	Diagnostic	ID4 (0.622)
CTNNB1	Hub Regulator	310	18	Diagnostic	CHD2 (0.530)
CCND1	Heavily Regulated	0	42	Diagnostic	ZBTB20 (0.600)
TP53	Hub Regulator	163	7	Prognostic	**WWTR1 (0.473)**
KRAS	Heavily Regulated	0	7	Predictive	GPBP1 (0.609)

Perturbation analysis performed for all five genes - all upstream regulators analyzed (25, 18, 42, 7, and 7 regulators respectively, total of 99 regulators). The system simulates inhibiting each upstream regulator and ranks candidate targets using PageRank (primary ranking, best predictor of drug target success per Mora & Donaldson 2021) and out-degree centrality (secondary ranking). Top candidate shown for each gene. These rankings serve as hypotheses for experimental validation. Detailed TP53 perturbation results presented below (Table 3) as representative example with complete rankings of all 7 regulators.

Biomarker Classification and Validation

RegNetAgents automatically classified biomarker types based on regulatory architecture and domain agent analysis:

Diagnostic Biomarkers (MYC, CTNNB1, CCND1): Genes with high regulatory input and pathway enrichment in proliferation/Wnt signaling pathways. MYC and CTNNB1 function as hub regulators that amplify oncogenic signals, while CCND1 acts as a terminal effector. These genes serve as indicators of disease presence, with expression levels reflecting oncogenic pathway activation. Literature validation: MYC amplification occurs in 15-20% of CRCs and correlates with poor prognosis (23,24); CTNNB1 mutations/dysregulation occur in 40-80% of CRCs via APC loss and Wnt activation (25,26); CCND1 overexpression occurs in 30-60% of CRCs and drives G1/S transition (27).

Prognostic Biomarker (TP53): Hub regulator with high network centrality and enrichment in TP53-regulation pathways. TP53 status predicts patient outcomes and treatment response. Literature validation: TP53 mutations occur in 50-70% of CRCs and associate with advanced stage, metastasis, and poor survival (28,29).

Predictive Biomarker (KRAS): Target gene with moderate clinical actionability. KRAS status predicts response to specific therapies (anti-EGFR antibodies). Literature validation:

KRAS mutations occur in 40-45% of CRCs and confer resistance to cetuximab/panitumumab (30,31).

All five classifications aligned with published CRC biomarker literature, demonstrating 100% concordance with established clinical and research findings. Network analysis revealed distinct regulatory architectures: three hub regulators (TP53, MYC, CTNNB1) with extensive downstream connectivity (163, 427, and 310 targets respectively) indicating signal amplification roles, and two heavily regulated genes (CCND1 with 42 regulators, KRAS with 7 regulators) with no downstream regulation but multiple upstream inputs. These connectivity patterns align with known biological roles - TP53, MYC, and CTNNB1 as master regulatory hubs in tumor suppression and oncogenic signaling, while CCND1 and KRAS function as end-point effectors.

Perturbation Analysis: TP53 Candidate Regulator Prioritization

To illustrate framework capabilities for hypothesis generation, we performed automated perturbation analysis on TP53, which has 7 upstream regulators in epithelial cells. The analysis simulates inhibiting each regulator individually and ranks candidates using network centrality metrics (Table 3, Figure 3). These rankings serve as hypotheses for experimental validation, not predictions of therapeutic efficacy.

Table 3. TP53 Perturbation Analysis Results - Comparison with Known Biology

Rank	Regulator	PageRank	Out-Degree Centrality	Downstream Targets	Literature Status
1	WWTR1	0.473	0.020	293	[check] Literature-supported
2	RBPMS	0.469	0.028	403	Novel hypothesis
3	PRRX2	0.454	0.006	93	Novel hypothesis
4	CHD4	0.443	0.017	243	[check] Literature-supported
5	THRA	0.408	0.006	81	Novel hypothesis
6	YAP1	0.402	0.014	207	[check] Literature-supported
7	IKZF2	0.399	0.008	112	Novel hypothesis

*Regulators ranked by PageRank, a metric associated with drug target success per Mora & Donaldson (2021). PageRank scores >0.30 suggest high-quality network connections. Out-degree centrality measures downstream regulatory influence (hub identification). All 7

regulators contribute equally to TP53 direct regulation (14.3% each = 1/7 regulators). Literature status indicates whether published studies support functional interactions between the regulator and TP53. Three of seven high-ranking regulators (WWTR1, CHD4, YAP1) have literature support for TP53 interactions, demonstrating the framework identifies known functional interactors from network topology. Novel hypotheses (RBPMS, PRRX2, THRA, IKZF2) represent candidates prioritized for experimental validation.*

All seven regulators showed PageRank scores >0.30, meeting the threshold associated with successful drug targets in network studies. Each regulator contributes 14.3% of TP53's total regulatory input (1/7 regulators), representing equal direct regulatory loss. However, regulators differ substantially in network centrality: WWTR1 ranked highest by PageRank (0.473), while RBPMS showed highest out-degree centrality (0.028) with 403 downstream targets, suggesting broader downstream effects. YAP1 (PageRank 0.402) and WWTR1 (PageRank 0.473), both Hippo pathway effectors, showed pathway enrichment for "YAP1- and WWTR1 (TAZ)-stimulated gene expression" (FDR = 0.020), demonstrating biological coherence between network topology and pathway-level regulation.

Literature Comparison for Top-Ranked Candidates

We compared the top 3 network-ranked candidates (by PageRank) against published literature to assess whether topology-based ranking recapitulates known biology:

WWTR1 (TAZ): WW domain-containing transcription regulator 1, also known as TAZ, is a Hippo pathway effector that functions as a transcriptional co-activator. WWTR1 and its paralog YAP1 are key downstream effectors of Hippo signaling involved in cell fate decisions, proliferation control, and DNA damage responses (32,33). The high PageRank ranking identifies WWTR1 as a central network node, consistent with the established role of Hippo pathway components in regulating cell growth and tumor suppression (34,35). Literature documents bidirectional crosstalk between TP53 and Hippo pathway effectors, where TP53 can regulate YAP1/WWTR1 and vice versa depending on cellular context. The network connectivity between WWTR1 and TP53 represents a testable hypothesis for functional interaction warranting experimental validation of the specific regulatory direction in epithelial cells.

RBPMS: RNA-binding protein with multiple splicing that shows the highest degree centrality among TP53 regulators (403 downstream targets). The related protein RBPMS2 has been implicated in smooth muscle plasticity and gene regulation (36), suggesting potential roles in tissue-specific transcriptional control. While limited literature exists on RBPMS itself in cancer contexts, its extensive network connectivity and high PageRank (0.469) position it as a high-priority candidate for experimental validation in TP53 regulatory mechanisms.

PRRX2: Paired-related homeobox 2, a transcription factor involved in mesenchymal development and epithelial-mesenchymal transition pathways. PRRX2's high PageRank (0.454) despite moderate degree centrality reflects quality over quantity in network

connections—suggesting influence through key regulatory hubs rather than direct broad connectivity. Its limited characterization in TP53 regulatory contexts makes it a particularly intriguing novel hypothesis, as homeobox factors often orchestrate complex developmental and disease-relevant gene expression programs.

This comparison demonstrates that topology-based ranking can: (1) recapitulate experimentally validated regulators (WWTR1 and YAP1 from Hippo pathway, CHD4 from chromatin remodeling complexes (37-39)), and (2) generate testable hypotheses (RBPMS, PRRX2, THRA, IKZF2) for experimental follow-up. The framework serves as a hypothesis generation tool to prioritize candidates for experimental validation.

Cross-Cell-Type Regulatory Analysis

TP53 exhibited dramatic cell-type specificity in regulatory architecture. In epithelial cells, TP53 functions as a hub regulator (163 targets, 7 regulators), consistent with its central role in epithelial tumor suppression. In erythrocytes, TP53 retains hub status (41 targets, 1 regulator), though with reduced network complexity. In NKT cells, TP53 shows intermediate hub properties (28 targets, 4 regulators). Notably, in CD4 T cells, CD8 T cells, NK cells, and monocyte-derived dendritic cells, TP53 was not detected in the regulatory network (0 targets, 0 regulators).

Why is TP53 absent from immune cell networks? This observation likely reflects two complementary mechanisms:

1. **Low expression variability**: ARACNe networks are reconstructed from the top 1,024 highly variable genes in each cell type. If TP53 exhibits low expression variability across cells within a given immune cell population (constitutively low or uniformly high), it would be excluded from network construction despite being present in the cell. Differentiated immune cells may maintain stable TP53 levels without the dynamic regulation observed in proliferative epithelial tissues.

2. **Alternative stress response pathways**: Mature immune cells (T cells, NK cells, dendritic cells) rely predominantly on immune-specific stress response mechanisms (e.g., NF-κB signaling, interferon responses, cytokine-mediated apoptosis) rather than TP53-dependent cell cycle checkpoint control. These cells are largely post-mitotic and do not require the proliferation-gatekeeper function that TP53 provides in actively dividing epithelial tissues.

This cell-type specificity aligns with biological expectations: epithelial tissues experience high oncogenic stress (environmental exposures, high proliferation rates, barrier function) necessitating robust TP53 tumor suppressor networks for genomic surveillance and cell fate decisions. In contrast, differentiated immune cells operate under different selective pressures—antigen response, immune homeostasis, and inflammatory regulation—where TP53's canonical tumor suppressor role is less central to cellular function. The presence of TP53 regulatory activity in erythrocytes (precursor cells undergoing differentiation and DNA damage from oxidative stress) and NKT cells (semi-activated state) supports this

interpretation, as these represent intermediate functional states between proliferative and fully differentiated cells.

These findings demonstrate how cross-cell-type analysis reveals tissue-specific regulatory mechanisms critical for understanding disease susceptibility and therapeutic responses. TP53-targeted therapies may be most effective in epithelial cancers where TP53 maintains active regulatory networks, while alternative targets may be necessary for hematological malignancies arising from cells where TP53 shows minimal regulatory activity.

Pathway Enrichment Analysis

Reactome pathway enrichment for the TP53 regulatory network (TP53 + 7 regulators + top 10 targets) identified 16 statistically significant enriched pathways (FDR < 0.05). Top pathways included:

1. **Regulation of TP53 Expression** (FDR = 0.0033, 2/4 genes)

2. **RUNX3 regulates YAP1-mediated transcription** (FDR = 0.0084, 2/9 genes)

3. **YAP1- and WWTR1 (TAZ)-stimulated gene expression** (FDR = 0.020, 2/18 genes)

4. **Transcriptional regulation by RUNX3** (FDR = 0.020, 3/108 genes)

5. **Signaling by Hippo** (FDR = 0.020, 2/22 genes)

The enrichment of Hippo signaling pathways (YAP1/WWTR1-mediated transcription) directly validates perturbation analysis results, which identified YAP1 and WWTR1 as top therapeutic targets. This pathway-network concordance demonstrates biological coherence of the automated analysis pipeline. Additionally, enrichment of TP53-specific regulatory pathways (Regulation of TP53 Expression, Regulation of TP53 Activity through Acetylation) confirms that network neighbors are functionally related to TP53 biology rather than spurious associations.

DISCUSSION

Principal Findings

We developed RegNetAgents, a multi-agent AI framework that streamlines gene regulatory network analysis and makes sophisticated computational methods accessible to experimental biologists. The system integrates network modeling, automated perturbation simulation for candidate regulator prioritization, pathway enrichment with statistical validation, and parallel domain-specific interpretation into a unified conversational interface accessible without programming expertise. A demonstration case study on colorectal cancer biomarkers showed the framework can recapitulate literature-confirmed regulatory patterns across five genes (99 total regulators analyzed) and generate testable hypotheses for experimental validation. Perturbation analysis of TP53 ranked highly literature-supported Hippo pathway interactors (WWTR1, CHD4, YAP1) alongside novel

hypotheses (RBPMS, PRRX2, THRA, IKZF2), demonstrating utility for hypothesis generation. Performance benchmarks show orders of magnitude speedup compared to manual multi-database workflows, with complete network analysis completing in 15.49 seconds for 5 genes (rule-based mode) or ~62 seconds (LLM-powered mode with domain agent insights), compared to labor-intensive manual alternatives. **The framework is designed as a hypothesis generation and experimental prioritization tool, not a replacement for experimental validation or expert biological interpretation.**

Framework Validation Against Known Biology

A critical question for any computational tool is whether it produces biologically meaningful results. Our colorectal cancer case study provides multiple lines of validation:

1. Biomarker classification accuracy: All five genes were correctly classified (diagnostic, prognostic, predictive) with complete agreement with published clinical literature across this limited sample (n=5), despite the framework having no prior knowledge of these classifications.

2. Known regulatory relationships: TP53 perturbation analysis ranked WWTR1 (TAZ) and YAP1 as top network neighbors (PageRank 0.473 and 0.402), both Hippo pathway effectors with documented functional interactions with TP53 in the literature. This ranking occurred purely from network topology analysis without pathway-specific knowledge. Note that ARACNe network edges represent statistical associations (mutual information) and do not necessarily indicate regulatory directionality; TP53 and Hippo pathway components exhibit bidirectional crosstalk in various contexts.

3. Pathway-network concordance: Pathway enrichment independently identified Hippo signaling (FDR = 0.020), confirming biological coherence between network structure and pathway-level regulation.

These validations demonstrate the framework recapitulates established biology while generating novel testable hypotheses (RBPMS, PRRX2, THRA, IKZF2). This positions RegNetAgents as a reliable hypothesis generation tool for experimental biologists.

Comparison to Existing Approaches

Traditional gene regulatory analysis tools operate in isolation: network databases (STRING, BioGRID) provide interaction data but lack pathway context; pathway enrichment tools (Enrichr, DAVID, Reactome web interface) require manual gene list preparation; and domain-specific interpretation (cancer relevance, drug development potential, clinical actionability) remains a manual literature curation task. Recent tools have begun addressing integration: NetworkAnalyst combines network visualization with enrichment analysis (40), while CARNIVAL infers causal networks from perturbation data (41). However, these tools lack conversational interfaces, require programming expertise, and do not provide automated therapeutic target prioritization via perturbation analysis.

RegNetAgents advances the field through four key innovations:

1. Multi-Agent Parallelization: Four domain-specific agents (cancer, drug, clinical, systems) execute simultaneously, providing integrated biological perspectives in a single query. This contrasts with sequential manual workflows where researchers must separately consider each domain.

2. Automated Perturbation Analysis: Network topology-based simulation identifies therapeutic targets by quantifying regulatory impact of inhibiting upstream regulators. While gene expression prediction tools exist (e.g., cell2cell for cell-cell communication modeling (42)), our approach prioritizes regulators for experimental validation based on network architecture rather than attempting to predict expression changes, which require dynamic models and experimental perturbation data currently unavailable at single-cell resolution across cell types.

3. Conversational Interface: Model Context Protocol integration enables natural language queries ("Analyze TP53 in epithelial cells for cancer pathways") rather than programming or web form interactions. This democratizes access to sophisticated network analysis for experimental biologists without computational training.

4. Pre-Computed Cell-Type Networks: Leveraging pre-computed ARACNe networks for 10 cell types (prepared by the GREmLN team) enables instant cross-cell-type comparisons, revealing tissue-specific regulatory mechanisms critical for understanding disease susceptibility and therapeutic responses.

Biological Insights from Case Studies

The colorectal cancer analysis revealed distinct regulatory architectures with therapeutic implications. Hub regulators (TP53, MYC, CTNNB1) with extensive downstream connectivity (163-427 targets) function as signal amplifiers but present off-target risks, while terminal effectors (CCND1, KRAS) offer more specific intervention points. TP53 perturbation analysis revealed distributed regulatory control (7 regulators, each ~14% contribution), suggesting combinatorial therapeutic strategies may be necessary. The identification of YAP1/WWTR1 (Hippo pathway effectors) as top regulators aligns with recent interest in Hippo pathway-targeted therapies (34,35), demonstrating the framework can recapitulate and extend experimentally derived therapeutic hypotheses.

Limitations and Considerations

Several limitations warrant consideration:

Validation Scope: Framework demonstration was conducted on a focused panel of five well-characterized colorectal cancer biomarkers (MYC, CTNNB1, CCND1, TP53, KRAS), showing complete concordance with published literature across this limited sample. While this provides evidence of biological validity for this specific gene set, broader systematic benchmarking across larger gene panels and diverse disease contexts is needed to establish generalizability. The five-gene demonstration serves as proof-of-concept rather than comprehensive validation across the full spectrum of human genes and regulatory relationships. Nevertheless, the framework's modular design enables application to any

gene set or cancer type, with validation scope expandable based on specific research questions.

Network Scope: The current implementation supports 10 cell types derived from publicly available scRNA-seq data. Expanding to additional cell types requires access to high-performance computing resources for ARACNe network reconstruction (12-14 hours per cell type, 64-128 GB RAM). We provide documentation for advanced users to add cell types, but this remains a computational barrier for typical research labs.

Topology-Based Perturbation Analysis: Our perturbation analysis ranks regulators based on network topology (connectivity, cascade overlap) rather than predicting gene expression changes. This design choice reflects data availability—dynamic gene expression prediction requires time-series perturbation experiments currently unavailable at single-cell resolution across cell types. Topology-based ranking identifies regulators for experimental validation rather than replacing experimental perturbation studies. Experimental validation remains essential to confirm predicted regulatory relationships and quantify expression changes.

Network Construction Methodology: ARACNe networks are reconstructed from scRNA-seq data using mutual information on the top 1,024 highly variable genes. Several important limitations apply: (1) Genes with low or uniform expression across cells will not appear in networks, potentially missing important regulators with constitutive low expression; (2) ARACNe identifies correlation-based statistical associations (mutual information) rather than directional regulatory relationships—network edges represent co-expression patterns that suggest potential regulation but do not confirm regulatory directionality or direct transcription factor binding; (3) False positive rates in inferred networks depend on sample size and data quality, meaning not all identified edges represent true biological regulation; (4) The method assumes steady-state gene expression and may not capture dynamic or transient regulatory relationships. Chromatin immunoprecipitation (ChIP-seq) data could complement networks with binding-site information to confirm regulatory directionality, though such data are not yet available at single-cell resolution across diverse cell types. Users should interpret network edges as hypotheses about regulatory relationships requiring experimental validation, not as established causal mechanisms.

Pathway Enrichment Dependence: Reactome pathway enrichment requires API calls, introducing latency (1-3 seconds) and external dependency. Offline pathway databases could reduce latency but would require periodic updates to maintain currency. Current implementation prioritizes up-to-date pathway annotations over speed.

Prototype Implementation: The system is currently deployed as a local Model Context Protocol server for Claude Desktop, requiring Python installation and network cache setup. While this provides a functional conversational interface, broader accessibility would benefit from web-based deployment. However, web deployment introduces challenges of API rate limiting, compute costs for network analysis, and data privacy considerations for proprietary gene lists.

Future Directions

Several extensions could enhance RegNetAgents' capabilities while maintaining simplicity and accessibility:

Expanded Cell Type Coverage: Integration of additional cell types from CellxGene as new datasets become available. The existing preprocessing pipeline can be applied to new cell types (cardiomyocytes, neurons, hepatocytes) to broaden disease applicability without requiring new methodologies.

Additional Gene Databases: The system currently integrates UniProt protein function annotations for comprehensive gene characterization. Future integration of DrugBank for existing therapeutic information and drug-target relationships would further enrich domain agent analyses with minimal architectural changes, as DrugBank provides structured APIs similar to currently integrated resources.

Performance Optimization: Caching strategies for frequently queried genes and pathway results could further reduce latency for common use cases. Pre-computation of centrality metrics for additional cell types would eliminate first-query overhead.

Web Deployment: While currently deployed locally via Model Context Protocol, a web-based interface would improve accessibility for researchers without Python expertise. This would require consideration of API rate limiting and compute resource management but follows established patterns from existing bioinformatics web tools.

Broader Impact and Intended Use

RegNetAgents addresses a critical accessibility barrier in computational biology: the expertise gap between sophisticated analytical methods and experimental researchers who could benefit from them. By providing a conversational interface to multi-hour analytical workflows, the framework enables biologists to rapidly generate hypotheses, prioritize experiments, and explore regulatory mechanisms without programming expertise.

Intended Use: The framework is designed for hypothesis generation and experimental prioritization, not for making clinical decisions or claiming novel biological discoveries. Users should:

1. **Validate hypotheses experimentally** - Network predictions require experimental confirmation

2. **Consult domain experts** - Biological interpretations should be reviewed by specialists

3. **Consider limitations** - Topology-based analysis does not predict dynamic gene expression changes

4. **Review literature** - Framework outputs should be compared against current biological knowledge

The modular architecture—separating workflow orchestration (LangGraph) from interface (MCP server)—enables deployment across multiple contexts: command-line tools for computational biologists, REST APIs for web applications, and computational notebooks for reproducible research. This flexibility facilitates integration into existing research workflows rather than requiring adoption of entirely new platforms.

Open data principles underpin the framework: regulatory networks derive from publicly available CellxGene data (via GREmLN team preprocessing), pathway annotations use the open Reactome database, and gene annotations come from NCBI-backed resources. This ensures reproducibility and transparency of analytical methods.

CONCLUSIONS

RegNetAgents transforms labor-intensive manual gene regulatory analysis into second-scale automated workflows through multi-agent AI orchestration, making sophisticated computational methods accessible to experimental biologists. Demonstration on colorectal cancer biomarkers showed the framework can recapitulate literature-confirmed regulatory patterns and generate testable hypotheses for experimental validation. The topology-based perturbation analysis ranked highly literature-supported TP53 interactors alongside novel candidates, demonstrating utility for hypothesis generation and experimental prioritization. The conversational interface via Model Context Protocol enables natural language queries without programming expertise. By integrating network analysis, perturbation simulation, pathway enrichment, and multi-domain interpretation into a unified accessible platform, RegNetAgents addresses critical bottlenecks in hypothesis generation and experimental prioritization. The modular architecture enables extension to new cell types, analysis agents, and deployment contexts. **This framework is intended as a hypothesis generation tool to assist researchers, not as a replacement for experimental validation or expert biological interpretation.**

ACKNOWLEDGMENTS

We thank the Chan Zuckerberg Initiative and the Califano Lab (CZ Biohub NY / Columbia University) for the GREmLN project and for preparing and publicly releasing the preprocessed ARACNe networks used in this work. We acknowledge the GREmLN development team (Zhang et al. 2025) for making these networks publicly available through the Virtual Cells Platform. We thank CellxGENE for curating and providing access to single-cell RNA-seq datasets and the Reactome team for maintaining the pathway database and API. We thank the Claude AI team at Anthropic for developing the Model Context Protocol enabling conversational interfaces.

Development Note: This project was developed with significant assistance from Claude Code (Anthropic's AI coding agent), which provided guidance on software architecture, Python implementation, workflow orchestration, documentation, and manuscript preparation. Microsoft Copilot provided valuable feedback on manuscript structure, clarity,

and technical rigor during the pre-submission review process. This work demonstrates how AI-assisted development tools can enable researchers from diverse backgrounds to implement sophisticated computational frameworks, broadening participation in computational biology research.

AUTHOR CONTRIBUTIONS

J.A.B. conceived the project, designed the analysis framework, performed validation studies, and wrote the manuscript with substantial assistance from Claude Code. Software implementation, architecture, and manuscript preparation were developed collaboratively with Claude Code (Anthropic AI coding agent). This work demonstrates the potential of AI-assisted development and writing tools to enable researchers from diverse backgrounds to tackle complex computational biology challenges.

COMPETING INTERESTS

The authors declare no competing interests.

REFERENCES

1. Babu MM, Luscombe NM, Aravind L, Gerstein M, Teichmann SA. Structure and evolution of transcriptional regulatory networks. Curr Opin Struct Biol. 2004;14(3):283-291.

2. Davidson EH, Levine MS. Properties of developmental gene regulatory networks. Proc Natl Acad Sci USA. 2008;105(51):20063-20066.

3. Califano A, Alvarez MJ. The recurrent architecture of tumour initiation, progression and drug sensitivity. Nat Rev Cancer. 2017;17(2):116-130.

4. Sonawane AR, Platig J, Fagny M, et al. Understanding tissue-specific gene regulation. Cell Rep. 2017;21(4):1077-1088.

5. Aibar S, González-Blas CB, Moerman T, et al. SCENIC: single-cell regulatory network inference and clustering. Nat Methods. 2017;14(11):1083-1086.

6. Kamimoto K, Stringa B, Hoffmann CM, et al. Dissecting cell identity via network inference and in silico gene perturbation. Nature. 2023;614(7949):742-751.

7. Szklarczyk D, Gable AL, Nastou KC, et al. The STRING database in 2021: customizable protein-protein networks, and functional characterization of user-uploaded gene/measurement sets. Nucleic Acids Res. 2021;49(D1):D605-D612.

8. Oughtred R, Rust J, Chang C, et al. The BioGRID database: A comprehensive biomedical resource of curated protein, genetic, and chemical interactions. Protein Sci. 2021;30(1):187-200.

9. Chen EY, Tan CM, Kou Y, et al. Enrichr: interactive and collaborative HTML5 gene list enrichment analysis tool. BMC Bioinformatics. 2013;14:128.

10. Sherman BT, Hao M, Qiu J, et al. DAVID: a web server for functional enrichment analysis and functional annotation of gene lists (2021 update). Nucleic Acids Res. 2022;50(W1):W216-W221.

11. Wooldridge M. An Introduction to MultiAgent Systems. 2nd ed. Wiley; 2009.

12. LangGraph Documentation. LangChain AI. https://langchain-ai.github.io/langgraph/

13. Anthropic. Model Context Protocol Documentation. https://modelcontextprotocol.io

14. Zhang M, Swamy V, Cassius R, Dupire L, Karaletsos T, Califano A. GREmLN: A Cellular Graph Structure Aware Transcriptomics Foundation Model. bioRxiv. 2025. doi:10.1101/2025.07.03.663009

15. Megill C, Martin B, Weaver C, et al. cellxgene: a performant, scalable exploration platform for high dimensional sparse matrices. bioRxiv. 2021. doi:10.1101/2021.04.05.438318

16. Margolin AA, Nemenman I, Basso K, et al. ARACNE: an algorithm for the reconstruction of gene regulatory networks in a mammalian cellular context. BMC Bioinformatics. 2006;7(Suppl 1):S7.

17. Lachmann A, Giorgi FM, Lopez G, Califano A. ARACNe-AP: gene network reverse engineering through adaptive partitioning inference of mutual information. Bioinformatics. 2016;32(14):2233-2235.

18. Gillespie M, Jassal B, Stephan R, et al. The reactome pathway knowledgebase 2022. Nucleic Acids Res. 2022;50(D1):D687-D692.

19. Yates AD, Achuthan P, Akanni W, et al. Ensembl 2020. Nucleic Acids Res. 2020;48(D1):D682-D688.

20. Hagberg AA, Schult DA, Swart PJ. Exploring network structure, dynamics, and function using NetworkX. Proceedings of the 7th Python in Science Conference (SciPy 2008); 2008. p. 11-15.

21. Koschützki D, Schreiber F. Centrality analysis methods for biological networks and their application to gene regulatory networks. Gene Regulation and Systems Biology. 2008;2:GRSB.S702.

22. Mora A, Donaldson IM. Effects of protein interaction data integration, representation and reliability on the use of network properties for drug target prediction. BMC Bioinformatics. 2021;22(1):1-29. doi:10.1186/s12859-021-04042-6

23. Erisman MD, Rothberg PG, Diehl RE, Morse CC, Spandorfer JM, Astrin SM. Deregulation of c-myc gene expression in human colon carcinoma is not accompanied by amplification or rearrangement of the gene. Mol Cell Biol. 1985;5(8):1969-1976.

24. Sears R, Nuckolls F, Haura E, Taya Y, Tamai K, Nevins JR. Multiple Ras-dependent phosphorylation pathways regulate Myc protein stability. Genes Dev. 2000;14(19):2501-2514.

25. Morin PJ, Sparks AB, Korinek V, et al. Activation of beta-catenin-Tcf signaling in colon cancer by mutations in beta-catenin or APC. Science. 1997;275(5307):1787-1790.

26. Segditsas S, Tomlinson I. Colorectal cancer and genetic alterations in the Wnt pathway. Oncogene. 2006;25(57):7531-7537.

27. Bartkova J, Lukas J, Strauss M, Bartek J. Cyclin D1 oncoprotein aberrantly accumulates in malignancies of diverse histogenesis. Oncogene. 1995;10(4):775-778.

28. Iacopetta B. TP53 mutation in colorectal cancer. Hum Mutat. 2003;21(3):271-276.

29. Olivier M, Hollstein M, Hainaut P. TP53 mutations in human cancers: origins, consequences, and clinical use. Cold Spring Harb Perspect Biol. 2010;2(1):a001008.

30. Lievre A, Bachet JB, Le Corre D, et al. KRAS mutation status is predictive of response to cetuximab therapy in colorectal cancer. Cancer Res. 2006;66(8):3992-3995.

31. Karapetis CS, Khambata-Ford S, Jonker DJ, et al. K-ras mutations and benefit from cetuximab in advanced colorectal cancer. N Engl J Med. 2008;359(17):1757-1765.

32. Strano S, Monti O, Pediconi N, et al. The transcriptional coactivator Yes-associated protein drives p73 gene-target specificity in response to DNA damage. Mol Cell. 2005;18(4):447-459.

33. Levy D, Adamovich Y, Reuven N, Shaul Y. Yap1 phosphorylation by c-Abl is a critical step in selective activation of proapoptotic genes in response to DNA damage. Mol Cell. 2008;29(3):350-361.

34. Zanconato F, Cordenonsi M, Piccolo S. YAP and TAZ: a signalling hub of the tumour microenvironment. Nat Rev Cancer. 2019;19(8):454-464.

35. Zhao B, Tumaneng K, Guan KL. The Hippo pathway in organ size control, tissue regeneration and stem cell self-renewal. Nat Cell Biol. 2011;13(8):877-883.

36. Sagnol S, Yang Y, Bessin Y, et al. Homodimerization of RBPMS2 through a new RRM-interaction motif is necessary to control smooth muscle plasticity. Nucleic Acids Res. 2014;42(15):10173-10184.

37. Polo SE, Kaidi A, Baskcomb L, Galanty Y, Jackson SP. Regulation of DNA-damage responses and cell-cycle progression by the chromatin remodelling factor CHD4. EMBO J. 2010;29(18):3130-3139.

38. Larsen DH, Poinsignon C, Gudjonsson T, et al. The chromatin-remodeling factor CHD4 coordinates signaling and repair after DNA damage. J Cell Biol. 2010;190(5):731-740.

39. Smeenk G, Wiegant WW, Vrolijk H, et al. The NuRD chromatin-remodeling complex regulates signaling and repair of DNA damage. J Cell Biol. 2010;190(5):741-749.

40. Zhou G, Soufan O, Ewald J, Hancock REW, Basu N, Xia J. NetworkAnalyst 3.0: a visual analytics platform for comprehensive gene expression profiling and meta-analysis. Nucleic Acids Res. 2019;47(W1):W234-W241.

41. Liu A, Trairatphisan P, Gjerga E, et al. From expression footprints to causal pathways: contextualizing large signaling networks with CARNIVAL. NPJ Syst Biol Appl. 2019;5:40.

42. Shao DD, Xue W, Krall EB, et al. KRAS and YAP1 converge to regulate EMT and tumor survival. Cell. 2014;158(1):171-184.

FIGURE LEGENDS

Figure 1. RegNetAgents Multi-Agent Architecture.

Workflow schematic showing the directed acyclic graph of agent execution. User queries enter through the MCP server interface and are routed to the initialization/validation agent for gene identifier resolution. The network modeling agent retrieves regulators and targets from pre-computed ARACNe networks and classifies regulatory roles. Intelligent routing determines whether perturbation analysis is triggered (genes with >5 regulators). Pathway enrichment queries Reactome API for statistical validation. Four domain-specific agents (cancer biology, drug discovery, clinical relevance, systems biology) execute in parallel, each providing specialized interpretation. The integration agent synthesizes findings into a comprehensive JSON report. Dashed lines indicate conditional execution; solid lines indicate required steps.

Figure 2. Colorectal Cancer Biomarker Panel Regulatory Architecture.

(A) Network diagram showing the five analyzed genes (MYC, CTNNB1, CCND1, TP53, KRAS) with regulators (upstream arrows) and targets (downstream arrows) in the epithelial cell network. Node size represents number of connections; node color indicates regulatory role classification (red = hub regulator, orange = heavily regulated, blue = weakly regulated). TP53, MYC, and CTNNB1 emerge as central hubs with extensive downstream connectivity. (B) Bar chart comparing number of regulators (left) and targets (right) for each gene. (C) Biomarker type classification showing diagnostic, prognostic, and predictive categories.

Figure 3. TP53 Perturbation Analysis and Therapeutic Target Ranking.

(A) Network diagram showing TP53 (center) with 7 upstream regulators. All regulators contribute equal direct regulatory loss (14.3% = 1/7 regulators). Node size represents downstream target count (larger = more targets = potential off-target effects). Top 3 regulators by PageRank are highlighted with color. (B) Horizontal bar chart of PageRank centrality scores for all 7 regulators, ranked from highest (WWTR1, 0.473) to lowest (IKZF2, 0.399). PageRank differentiates therapeutic potential when regulatory loss is equal. Color intensity indicates network influence. (C) Alternative ranking by degree centrality showing downstream target count, with RBPMS highest (403 targets). Annotations indicate regulators with literature validation (WWTR1, YAP1 - Hippo pathway) vs. novel hypotheses (RBPMS).

Figure 4. Framework Value Demonstration: Workflow Automation and LLM Intelligence.

(A) Traditional manual workflow for gene regulatory analysis requires sequential querying of multiple databases (STRING/BioGRID for network data, Reactome/Enrichr for pathway enrichment, PubMed for literature curation) followed by manual synthesis, totaling 2-4 hours per gene. Workflow shown as four sequential steps with time estimates. (B) RegNetAgents automates this workflow through natural language interaction with Claude Desktop, executing parallel multi-agent analysis (network modeling, perturbation analysis, pathway enrichment, four domain agents) in 0.6-15 seconds. (C) Performance comparison bar chart showing orders of magnitude speedup across single and multi-gene analyses in both rule-based and LLM-powered modes. Manual workflow baseline: 2.5 hours per gene (conservative estimate); 5-gene panel: 12.5 hours. Horizontal bars on logarithmic scale. (D) LLM-powered mode (right panel) adds scientific rationales and biological interpretations (+14 seconds) compared to rule-based network metrics and qualitative classifications only (left panel, 0.6 seconds), demonstrated with TP53 analysis. Example shows how LLM integrates gene function data (NCBI/UniProt) with network topology to provide biological context. System includes transparency flag (`llm_powered: true/false`) and graceful fallback to rule-based mode if Ollama unavailable.

DATA AVAILABILITY

All data and results needed to evaluate the conclusions in the paper are present in the paper and/or the figures.

Regulatory network data were obtained as pre-computed ARACNe networks prepared by the GREmLN team (Zhang et al. 2025, bioRxiv 2025.07.03.663009). These preprocessed ARACNe networks for 10 cell types are publicly available through the GREmLN Quickstart Tutorial at https://virtualcellmodels.cziscience.com/quickstart/gremln-quickstart.

Raw analysis outputs (JSON format) for the case study analyses presented in this manuscript are available from the corresponding author upon reasonable request.

Supplementary Material: Complete LLM prompt templates for all four domain agents, prompt engineering notes, and reproducibility instructions are provided in Supplementary Material (Table S1).

CODE AVAILABILITY

The RegNetAgents software implementation is publicly available at https://github.com/jab57/RegNetAgents under the MIT License.

Installation requires Python 3.8+ and takes approximately 5-10 minutes. Network data files (pre-computed ARACNe networks from GREmLN) are downloaded from the GREmLN Quickstart Tutorial as documented in the repository README. All analytical methods and workflow orchestration code are provided to facilitate independent use and extension by the research community.

Network data used in this study are publicly available through the GREmLN Quickstart Tutorial (https://virtualcellmodels.cziscience.com/quickstart/gremln-quickstart). Pathway enrichment was performed using the publicly accessible Reactome API (https://reactome.org/AnalysisService/).

END OF MANUSCRIPT