A100 · From human genetics evidence to therapeutic insights at scale: a calibrated language-model specialist for target discovery in immunology

Presenter: Mahasweta Bhattacharya — Sanofi

Authors: Mahasweta Bhattacharya

Session: Day 2 · Fri Oct 2, 2026 · 10:00–10:45 AM

Genetically validated targets are twice as likely to succeed in clinical trials, yet systematically decoding target mechanism of action in each genetic locus to translate GWAS associations into actionable insights requires multi-omic integration—a complex, time-intensive process. Our AI-ready data ecosystem integrating human genetics, proteomics, and transcriptomics across 500 indications enables scalable, mechanism-driven target discovery through multi-specialist agentic AI frameworks. Here, we demonstrate a genetics specialist LLM agent, which synthesizes genetic evidences into interpretable biological rationale. We built the genetics specialist using Claude 4.5 Sonnet via LangChain, operating as a defined scientific persona. The specialist integrates Mendelian randomization, molecular QTL actionability, tissue specificity, and variant pathogenicity to identify the most likely effector gene for a disease. Hallucinations were mitigated via: temperature = 0.1, max tokens = 2048, and prompt constraints requiring cited evidence and variant annotations. Across 800 gene-indication pairs spanning 100 immune indications, the genetics specialist enriched for approved targets ~20% more effectively than Open Targets genetic association, with consistent results across five independent runs. The agent provided mechanistic rationale complementing algorithmic gene ranking, with confidence scores reflecting evidence strength. For approved IL12B in Crohn's disease (confidence: 0.65), it identified the shared p40 subunit of IL-12/IL-23 with colocalization in immune tissues. Notably, the agent's top call was ITGA4 (confidence: 0.85), prioritized via 17 colocalization events predominantly in Th17 cells in two cohorts (largest: 20,873 cases). From eQTL directionality (risk allele elevates ITGA4 expression), the agent inferred an antagonist strategy—concordant with approved anti-α4 therapies. Biological rationales were reviewed by expert geneticists, confirming robust reasoning by the genetics specialist. The genetics specialist achieved robust enrichment for immune indications while providing interpretable mechanistic rationale, demonstrating that LLMs can accelerate expert-level evidence synthesis while maintaining rigor. This genetics-first approach establishes a validated foundation for a multi-specialist framework, extending to transcriptomics and proteomics specialists enabling systematic multi-omic target discovery—delivering auditable biological arguments at scale.

From human genetics evidence to therapeutic insights at scale: a calibrated language-model specialist for target discovery in immunology

Mahasweta Bhattacharya[1], Samuel Lessard[1], Tejaswi Iyyanki[1], Michael Chao[1], Mahesh Raundhal[1], Crystal Rawlings[1], Cora Ames[2], Dongyu Liu[2], Clement Chatelain[1], Franck Rapaport[2], Shameer Khader[2], Hamid Mattoo[1], Emanuele de Rinaldis[1]

[1] *Target, Disease and Systems Biology, Sanofi Research*

[2] *Computational and AI Strategy*

Genetically validated targets are twice as likely to succeed in clinical trials, yet systematically decoding target mechanism of action in each genetic locus to translate GWAS associations into actionable insights requires multi-omic integration—a complex, time-intensive process. Our AI-ready data ecosystem integrating human genetics, proteomics, and transcriptomics across 500 indications enables scalable, mechanism-driven target discovery through multi-specialist agentic AI frameworks. Here, we demonstrate a genetics specialist LLM agent, which synthesizes genetic evidences into interpretable biological rationale.

We built the genetics specialist using Claude 4.5 Sonnet via LangChain, operating as a defined scientific persona. The specialist integrates Mendelian randomization, molecular QTL actionability, tissue specificity, and variant pathogenicity to identify the most likely effector gene for a disease. Hallucinations were mitigated via: temperature = 0.1, max tokens = 2048, and prompt constraints requiring cited evidence and variant annotations.

Across 800 gene-indication pairs spanning 100 immune indications, the genetics specialist enriched for approved targets ~20% more effectively than Open Targets genetic association, with consistent results across five independent runs. The agent provided mechanistic rationale complementing algorithmic gene ranking, with confidence scores reflecting evidence strength. For approved IL12B in Crohn's disease (confidence: 0.65), it identified the shared p40 subunit of IL-12/IL-23 with colocalization in immune tissues. Notably, the agent's top call was ITGA4 (confidence: 0.85), prioritized via 17 colocalization events predominantly in Th17 cells in two cohorts (largest: 20,873 cases). From eQTL directionality (risk allele elevates ITGA4 expression), the agent inferred an antagonist strategy—concordant with approved anti-α4 therapies. Biological rationales were reviewed by expert geneticists, confirming robust reasoning by the genetics specialist.

The genetics specialist achieved robust enrichment for immune indications while providing interpretable mechanistic rationale, demonstrating that LLMs can accelerate expert-level evidence synthesis while maintaining rigor. This genetics-first approach establishes a validated foundation for a multi-specialist framework, extending to transcriptomics and proteomics specialists enabling systematic multi-omic target discovery—delivering auditable biological arguments at scale.