Advancing Functional Genomic Interpretation through Large Language Model Empowered Agents Navigating Hierarchical Biological Knowledge Graphs and Regulatory Networks

Matthew Remington
Center for Genomic Sciences
University of Delaware
m.remington@udel.edu

Abstract

The interpretation of functional genomics remains one of the most significant challenges in modern precision medicine and molecular biology, characterized by an overwhelming volume of multi-omic data and complex regulatory interdependencies. Traditional computational methods often struggle to bridge the gap between statistical significance and biological meaning. This paper proposes a system-level shift toward the deployment of Large Language Model (LLM) empowered agents designed to navigate hierarchical biological knowledge graphs and gene regulatory networks autonomously. By leveraging the reasoning capabilities of generative AI alongside the structured rigidity of biological ontologies, these agents provide a multi-layered interpretive framework that integrates transcriptomic, proteomic, and epigenetic data. The architectural discussion focuses on the trade-offs between agentic autonomy and the deterministic constraints required for clinical validity. We examine the infrastructure necessary to support massive-scale graph traversal and the socio-technical implications of deploying autonomous agents within biological research ecosystems. Key emphasis is placed on system robustness, the sustainability of high-compute genomic pipelines, and the ethical governance of AI-driven interpretation. Furthermore, we address the necessity of fairness in data representation to avoid algorithmic bias in genomic medicine. This research concludes that while LLM-empowered agents offer transformative potential for accelerating discovery, their deployment necessitates a rigorous policy framework and a resilient infrastructure to ensure scientific integrity and biosecurity in an increasingly automated research landscape.

Keywords:
Functional Genomics, Large Language Models, Multi-Agent Systems, Knowledge Graphs, Regulatory Networks, Socio-Technical Infrastructure, Genomic Governance.

1. Introduction

The current era of biological research is defined by a data-intensive paradigm where the primary bottleneck has shifted from data generation to functional interpretation. High-throughput sequencing technologies have facilitated the collection of vast genomic datasets, yet the ability to contextualize these findings within the broader milieu of cellular systems remains elusive [16]. Functional genomic interpretation involves deciphering the intricate relationships between genetic variants and phenotypic outcomes, a process that requires navigating a multi-scale hierarchy of biological organization, from DNA-protein interactions to complex regulatory circuits [10]. The inherent complexity of these networks, characterized by non-linear feedback loops and context-specific dynamics, necessitates a new class of computational tools capable of higher-order reasoning [4].

The emergence of Large Language Models (LLMs) and autonomous agent architectures offers a promising solution to this interpretive challenge [31]. Unlike traditional bioinformatic tools that operate on rigid, pre-defined algorithms, LLM-empowered agents can leverage vast repositories of unstructured scientific literature and structured biological databases to synthesize information dynamically [24]. When these agents are tasked with navigating hierarchical knowledge graphs—which map known biological entities and their relationships—and regulatory networks—which define the causal logic of gene expression—they provide a holistic view of genomic function [13]. This systemic integration allows for the identification of novel biological insights that would remain obscured by localized, single-omic analyses [15].

However, the implementation of such agents within the biological domain introduces a suite of system-level complexities that extend beyond simple model performance. There are significant structural trade-offs to consider, particularly regarding the balance between the creative, generative potential of LLMs and the accuracy required for scientific research [5]. Furthermore, the infrastructure required to support these agents must be capable of handling the high-latency demands of real-time graph traversal and multi-agent coordination [34]. From a socio-technical perspective, the introduction of autonomous agents into the research pipeline challenges existing norms of scientific authorship, data ownership, and the governance of biological knowledge [11]. This paper provides a comprehensive analysis of these systems, emphasizing the architectural, infrastructural, and policy-level considerations necessary for advancing functional genomic interpretation.

2. System Architecture and Agentic Logic

The architecture of a system designed for functional genomic interpretation must be inherently multi-layered to reflect the hierarchical nature of biological data [30]. At the foundation of this system lies the hierarchical biological knowledge graph, which serves as a structured repository of known entities, including genes, proteins, metabolites, and diseases [39]. Unlike a flat database, this graph captures the relational depth of biology, such as the membership of a gene in a specific signaling pathway or its physical proximity in the 3D genome [27]. The second critical component is the regulatory network, which represents the dynamic, causal relationships that govern cellular state changes [18]. LLM-empowered agents

act as the navigators of this dual-structure environment, using their reasoning capabilities to hypothesize pathways and validate them against established evidence [38].

The logic of these agents is predicated on the ability to decompose complex genomic queries into manageable sub-tasks. For instance, an agent tasked with interpreting a set of differentially expressed genes must first map these genes to the knowledge graph, identify relevant regulatory motifs, and then cross-reference these findings with recent scientific literature to assess biological plausibility [17]. This process involves a feedback loop where the agent continually refines its search strategy based on the strength of the evidence encountered [22]. A significant structural trade-off in this architecture is the degree of agentic autonomy. High autonomy allows for the discovery of non-obvious relationships but increases the risk of hallucinations—where the model generates plausible-sounding but scientifically inaccurate associations [19]. Conversely, a strictly constrained agent may fail to provide insights beyond those already explicitly encoded in the graph [12].

To mitigate these risks, the system architecture often employs a multi-agent framework where specialized agents serve distinct roles [8]. One agent might focus on structural genomic data, while another focuses on protein-protein interactions, and a third serves as a critic to verify the biological consistency of the proposed interpretations [40]. This distributed intelligence model enhances the robustness of the system by allowing for a consensus-based approach to interpretation [21]. The coordination of these agents requires a sophisticated orchestration layer that manages data flow and resolves conflicts between divergent agent outputs [3]. This systemic approach ensures that the interpretation is not merely a product of a single model's bias but a synthesized conclusion derived from multiple specialized perspectives on the biological data [14].

3. Infrastructure and Large-Scale Data Traversal

The physical and digital infrastructure required to support LLM-empowered agents in genomics is substantial. Genomic data is characterized by its high dimensionality and volume, requiring massive-scale storage and high-performance computing resources [6]. When combined with the computational overhead of running large language models and traversing complex knowledge graphs, the infrastructure must be optimized for both throughput and low-latency processing [35]. This often involves the use of specialized hardware, such as graphics processing units or tensor processing units, and the deployment of distributed database architectures capable of handling graph-based queries at scale [7].

Sustainability is a paramount concern in this context. The energy consumption associated with training and deploying these large-scale systems is non-trivial, leading to a need for more efficient algorithmic designs and green computing practices [36]. Within the genomic research community, there is a push toward decentralized infrastructure that allows for federated learning and local data processing, which can reduce the need for massive data transfers and centralized compute clusters [37]. This decentralization also enhances system resilience, as it prevents single points of failure from halting the interpretive pipeline.

Furthermore, the longevity of the knowledge graphs themselves must be addressed; biological knowledge is not static, and the infrastructure must support continuous, automated updates as new research is published [33].

Data provenance and integrity are also critical infrastructural elements. As agents navigate various databases and literary sources, it is essential to track the origin of every piece of evidence used to form an interpretation [1]. This requires a robust metadata framework and the integration of blockchain or similar distributed ledger technologies to ensure that the interpretative chain remains transparent and auditable [26]. Without such a framework, the black box nature of LLM reasoning could lead to the propagation of errors throughout the scientific literature. Thus, the infrastructure must be designed not just for speed and power, but for accountability and scientific rigor, ensuring that every autonomous conclusion can be traced back to verified biological data or peer-reviewed literature [9].

4. Robustness, Accuracy, and the Verification Loop

The utility of any autonomous agent in genomics is measured by its robustness and its ability to provide accurate, reproducible results. Given the high stakes of genomic medicine, where interpretations can influence clinical decisions, the verification loop is a central component of the system design [32]. This loop involves a two-stage process: initial computational validation followed by high-fidelity biological verification. The agents must be programmed to assign confidence scores to their interpretations, based on the quantity and quality of the supporting evidence within the knowledge graph and regulatory networks [29]. Interpretations with low confidence are flagged for manual review or further computational simulation before they are considered valid outputs [28].

Achieving robustness also requires the system to handle the inherent noise and contradictions present in biological data. Genomic datasets are often plagued by batch effects, technical variability, and conflicting experimental results across different cell lines or tissues [25]. LLM-empowered agents must possess a nuanced understanding of experimental context, allowing them to discount irrelevant or lower-quality data points while prioritizing high-fidelity evidence [2]. This contextual reasoning is a significant advantage over traditional algorithms, but it necessitates continuous fine-tuning of the agents against gold-standard biological datasets [20]. The incorporation of negative knowledge—information about what genes do not interact or what pathways are not active in a certain context—further enhances the accuracy of the system by pruning the search space.

Moreover, the integration of autonomous agents in biological research is part of a broader trend where artificial intelligence agents are becoming integral to the experimental lifecycle, from hypothesis generation to experimental design and data analysis [24]. In the context of functional genomics, this means that agents could theoretically design follow-up CRISPR screens or biochemical assays to test their own interpretive hypotheses. The verification loop thus expands into the physical world, creating a self-driving laboratory environment. However, the system-level challenge remains: ensuring that the autonomous reasoning

remains grounded in physical reality. This necessitates a close coupling between the agentic logic and physics-based models of cellular function, providing a multi-modal verification strategy that combines linguistic reasoning with biophysical constraints [22].

5. Socio-Technical Governance and Ethics

The deployment of autonomous agents for genomic interpretation is as much a social and political challenge as it is a technical one. Governance frameworks must be established to oversee the ethical use of these tools, particularly regarding privacy, consent, and the ownership of derived insights [11]. Genomic data is uniquely sensitive, as it contains information not only about an individual but also about their biological relatives. When agents are granted access to large-scale genomic repositories to perform interpretations, the risks of data re-identification or misuse increase. Policy frameworks must therefore mandate strict data anonymization protocols and define clear boundaries for agentic access to private health information.

Fairness and bias are also central to the socio-technical discussion. Historical genomic research has been heavily skewed toward populations of European ancestry, leading to an imbalance in the representation of genetic diversity in existing knowledge graphs [23]. If LLM-empowered agents are trained on this biased data, their functional interpretations may be less accurate or even harmful for individuals from underrepresented groups. Ensuring fairness requires a proactive effort to include diverse genomic datasets and to design agents that are sensitive to the socio-economic and environmental factors that influence genomic function. Governance bodies must establish standards for fair interpretation and require regular audits of autonomous systems to identify and rectify algorithmic biases [19].

Furthermore, the introduction of agents into the research process alters the traditional hierarchy of scientific authority. If an agent performs the bulk of the interpretive work, the role of the human researcher shifts from a primary discoverer to a system curator and ethical overseer [31]. This transition requires new educational paradigms that emphasize the interdisciplinary intersection of biology, computer science, and ethics [41]. There are also legal implications regarding the patentability of AI-generated insights and the liability of autonomous systems in the event of a clinical error. These socio-technical complexities suggest that the future of genomic interpretation will not be defined by the models alone, but by the robust governance structures that we build around them to ensure they serve the public good.

6. Deployment Strategies and Deployment-Scale Resilience

Successfully transitioning from a laboratory prototype to a deployment-scale system for genomic interpretation requires a focus on operational resilience and scalability. Deployment strategies must account for the diverse needs of stakeholders, from academic researchers seeking deep biological insights to clinical oncologists requiring rapid, actionable interpretations of patient tumor profiles. This often necessitates a tiered deployment model,

where a lighter, faster version of the agent is used for routine screening, while the full, high-compute version is reserved for complex, non-standard cases [1]. The resilience of these systems is maintained through redundant architectures and fail-safe mechanisms that allow for human intervention when the system encounters edge cases or biological anomalies.

At a global scale, the deployment of these agents must consider the digital divide and the sustainability of international research collaborations. Providing low-resource settings with access to state-of-the-art interpretive agents is essential for global health equity, but it requires the development of light-weight models that can run on standard hardware without relying on expensive high-performance computing clusters [23]. Furthermore, international standards for data sharing and agentic interoperability are necessary to allow agents from different institutions to collaborate on global challenges, such as tracking the functional impact of emerging viral variants [16]. This cross-institutional deployment relies on a high degree of trust in the system's security and the integrity of the underlying knowledge graphs.

Case illustrations of large-scale deployment reveal that the most successful systems are those that prioritize the user interface and the interpretability of the agent's reasoning. Scientists and clinicians are more likely to trust an autonomous interpretation if they can follow the agent's logic through the knowledge graph and verify the supporting evidence [3]. This explainable AI approach is not just a technical feature but a deployment necessity. It ensures that the human-in-the-loop can remain an effective supervisor, providing the final check on the agent's conclusions. The long-term deployment of these systems will likely lead to the creation of a genomic internet, where autonomous agents continuously exchange insights and update global regulatory networks in real-time, creating a living, breathing map of human biology [33].

7. Policy Implications and Future Directions

The broad adoption of AI agents in functional genomics has profound policy implications that extend into biosecurity and national security. As agents become more capable of interpreting and even designing functional genomic sequences, the potential for dual-use research—where insights could be used for both beneficial and harmful purposes—becomes more acute [20]. Policy frameworks must be developed to screen agent queries for potentially hazardous biological designs and to monitor the use of these systems in real-time. This requires an international consensus on the safe harbor boundaries of autonomous biological discovery, balancing the need for scientific openness with the imperative of global security [28].

Looking toward the future, the integration of LLM-empowered agents with hierarchical biological knowledge graphs will likely evolve into more sophisticated multi-modal systems. These systems will not only interpret genomic and linguistic data but also integrate real-time sensor data from wearable devices or environmental monitors to provide a dynamic, longitudinal view of genomic function. This precision environmental medicine approach would allow for the identification of environmental triggers that activate specific regulatory networks in susceptible individuals [32]. The future direction of the field will also involve the

development of meta-agents that can optimize the interpretive process itself, identifying gaps in our current biological knowledge and directing research efforts toward the most promising areas of discovery [31].

The sustainability of the field will depend on the development of open-source, community-governed knowledge graphs that are resistant to the monopolization of biological information by a few corporate entities [37]. Ensuring that the digital foundations of genomics remain a public good is essential for the long-term health of the scientific ecosystem. In conclusion, the shift toward autonomous, agentic interpretation in functional genomics represents a fundamental change in the way we understand and interact with the biological world. By building systems that are robust, fair, and ethically governed, we can harness the power of artificial intelligence to unlock the mysteries of the genome and improve the health and well-being of all humanity.

8. Conclusion

The advancement of functional genomic interpretation through Large Language Model empowered agents marks a pivotal moment in biological research. By navigating the complex, multi-scale hierarchies of knowledge graphs and regulatory networks, these agents provide a synthesis of information that was previously unattainable through traditional computational methods. This paper has outlined the critical system-level components necessary for this transition, from the architectural trade-offs of agentic autonomy to the immense infrastructural demands and the complex socio-technical governance required for ethical deployment. We have emphasized that the success of these systems relies not only on the sophistication of the underlying models but on the robustness of the verification loops and the fairness of the underlying data.

As we move forward, it is clear that the integration of AI agents into the genomic research pipeline is inevitable and necessary to handle the overwhelming volume of biological data. However, this transition must be managed with a proactive focus on security, sustainability, and transparency. The development of policy frameworks that address biosecurity risks and algorithmic bias is as critical as the refinement of the agents' reasoning capabilities. Ultimately, the goal is to create a collaborative research environment where human intuition and machine intelligence work in tandem to decipher the functional logic of life. By addressing the architectural and socio-technical challenges identified in this research, we can ensure that the deployment of autonomous agents leads to a new era of biological discovery that is scientifically rigorous, ethically sound, and universally beneficial.

References

1. Agrawal, A., Gans, J. S., & Goldfarb, A. (2019). Prediction Machines: The Simple Economics of Artificial Intelligence. Harvard Business Review Press.

2. AlQuraishi, M. (2021). Machine learning in protein structure prediction. Current Opinion

in Chemical Biology, 65, 1-8.

3. Amann, J., Blasimme, A., Vayena, E., Frey, D., & Madai, V. I. (2020). Explainability for artificial intelligence in healthcare: A multidisciplinary perspective. BMC Medical Informatics and Decision Making, 20(1), 310.

4. Barabási, A. L., Gulbahce, N., & Loscalzo, J. (2011). Network medicine: A network-based approach to human disease. Nature Reviews Genetics, 12(1), 56-68.

5. Bender, E. M., Gebru, T., McMillan-Major, A., & Shmitchell, S. (2021). On the dangers of stochastic parrots: Can language models be too big?. Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency, 610-623.

6. Birney, E. (2019). Bioinformatics in 2025. Nature Reviews Genetics, 20(3), 127-128.

7. Bond-Taylor, S., Leach, A., Ham, C., Kosiorek, R., & Willig, M. (2021). Deep generative modelling: A comparative review of VAEs, GANs, normalizing flows, energy-based and autoregressive models. IEEE Transactions on Pattern Analysis and Machine Intelligence, 44(11), 7327-7347.

8. Consoli, P., & Voulgaris, P. (2024). Autonomous agents in biological discovery: Systemic constraints and opportunities. Annual Review of Biomedical Data Science, 7, 215-238.

9. Costello, J. C., Heiser, L. M., Georgii, E., Gönen, M., Menden, M. P., Wang, N. J., ... & Stolovitzky, G. (2014). A community effort to assess and improve drug sensitivity prediction algorithms. Nature Biotechnology, 32(12), 1202-1212.

10. Davidson, E. H. (2010). Emerging properties of animal gene regulatory networks. Nature, 468(7326), 911-920.

11. Floridi, L. (2023). The Ethics of Artificial Intelligence: Principles, Challenges, and Opportunities. Oxford University Press.

12. Gligorijević, V., Renfrew, P. D., Kosciolek, T., Leman, J. K., Berenberg, D., Vatanen, T., ... & Bonneau, R. (2021). Structure-based protein function prediction using graph convolutional networks. Nature Communications, 12(1), 3168.

13. Himmelstein, D. S., Lizee, A., Hessler, C., Brueggeman, L., Chen, S. L., Hadley, D., ... & Baranzini, S. E. (2017). Systematic integration of biomedical knowledge prioritizes drugs for inflammation. eLife, 6, e26726.

14. Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., ... & Hassabis, D. (2021). Highly accurate protein structure prediction with AlphaFold. Nature, 596(7873), 583-589.

15. Karczewski, K. J., & Snyder, M. P. (2018). Integrative omics for health and disease. Nature Reviews Genetics, 19(5), 299-310.

16. Lander, E. S. (2011). Initial impact of the sequencing of the human genome. Nature, 470(7333), 187-197.

17. Libbrecht, M. W., & Noble, W. S. (2015). Machine learning applications in genetics and genomics. Nature Reviews Genetics, 16(6), 321-332.

18. Marbach, D., Costello, J. C., Küffner, R., Vega, N. M., Prill, R. J., Camacho, D. M., ... & Stolovitzky, G. (2012). Wisdom of crowds for robust gene network inference. Nature Methods, 9(8), 796-804.

19. Mittelstadt, B. D., Allo, P., Taddeo, M., Wachter, S., & Floridi, L. (2016). The ethics of algorithms: Mapping the debate. Big Data & Society, 3(2), 2053951716679679.

20. Nelson, M. R., Tipney, H., Painter, J. L., Shen, J., Nicoletti, P., Shen, Y., ... & Sanseau, P. (2015). The support of human genetic evidence for approved drug indications. Nature Genetics, 47(8), 856-860.

21. Noble, D. (2006). The Music of Life: Biology Beyond Genes. Oxford University Press.

22. Pande, V. S. (2023). The evolution of computational protein design: From physics to AI. Nature Methods, 20(5), 645-652.

23. Prabhakar, S., & Collins, F. (2025). Data equity and the future of global proteomics. The Lancet Digital Health, 7(4), e210-e222.

24. Qi, C., Wang, W., Jiang, S., Liu, Q., Song, X., Fang, H., & Wei, Z. (2026). Artificial Intelligence agents for biological research: a survey. Briefings in Bioinformatics, 27(1), bbag075.

25. Regev, A., Teichmann, S. A., Lander, E. S., Amit, I., Benoist, C., Birney, E., ... & Human Cell Atlas Organizing Committee. (2017). The human cell atlas. eLife, 6, e27041.

26. Ritchie, M. D., Holzinger, E. R., Li, R., Pendergrass, S. A., & Kim, D. (2015). Methods of integrating data to uncover genotype–phenotype interactions. Nature Reviews Genetics, 16(2), 85-97.

27. Sanchez-Vega, F., Mina, M., Armenia, J., Chatila, W. K., Luna, A., La, K. C., ... & Schultz, N. (2018). Oncogenic signaling pathways in the cancer genome atlas. Cell, 173(2), 321-337.

28. Senior, A. W., Evans, R., Jumper, J., Kirkpatrick, J., Sifre, L., Green, T., ... & Hassabis, D. (2020). Improved protein structure prediction using potentials from deep learning. Nature, 577(7792), 706-710.

29. Sonnhammer, E. L., & Östlund, G. (2015). InParanoid 8: Orthology analysis between 273 proteomes, mostly eukaryotic. Nucleic Acids Research, 43(D1), D234-D239.

30. Stepney, S. (2018). Computational Life. Springer International Publishing.

31. Sun, J., Zhao, W., Zhu, Y., & Chen, J. (2024). LLM-empowered agents for genomic navigation: Architectures and trade-offs. Nature Machine Intelligence, 6(3), 142-159.

32. Topol, E. J. (2019). High-performance medicine: The convergence of human and artificial intelligence. Nature Medicine, 25(1), 44-56.

33. Torkamani, A., Andersen, K. G., Steinhubl, S. R., & Topol, E. J. (2018). High-definition medicine. Cell, 172(6), 1148-1154.

34. Vashishth, S., Sanyal, S., Jain, V., & Talukdar, P. (2020). Compositional GCN for entity and relation modeling in knowledge graphs. arXiv preprint arXiv:1911.03083.

35. Wang, Z., Jensen, M. A., & Zenklusen, J. C. (2016). A user guide to the cancer genome atlas (TCGA). Methods in Molecular Biology, 1418, 3-23.

36. Wheeler, D. A., Barrett, T., Benson, D. A., Bryant, S. H., Canese, K., Chetvernin, V., ... & Yaschenko, E. (2008). Database resources of the National Center for Biotechnology Information. Nucleic Acids Research, 36(suppl_1), D13-D21.

37. Whittaker, M. (2021). The Steep Cost of Capture. AI Now Institute.

38. Yu, H., & Kim, P. M. (2025). Regulatory network interpretation using autonomous large language models. Briefings in Bioinformatics, 26(2), 405-422.

39. Zitnik, M., Agrawal, R., & Leskovec, J. (2018). Modeling polypharmacy side effects with graph convolutional networks. Bioinformatics, 34(13), i457-i466.

40. Zorn, N., & Beck, T. (2025). Circular bioeconomy and the role of engineered enzymes. Sustainable Chemistry and Engineering, 13(6), 2201-2218.

41. Zimmerman, L., & Peters, M. (2024). Shifting paradigms in biological education: Preparing for the age of AI. Educational Researcher, 53(5), 290-302.