Positioning Multi-Agent Large Language Models as the Future of Bulk Gene Expression Analysis and Cancer Prediction

Anonymous Author(s)
Affiliation
Address
email

Abstract

Bulk RNA sequencing is essential for understanding cancer biology, yet current computational methods struggle with cross-cohort generalization, interpretability, and multi-source integration. We propose **multi-agent architectures built from specialized large language models (LLMs)** as a solution. Unlike monolithic models, our framework integrates **multi-modal inputs**. By assigning complementary tasks to expression-focused, sequence-based, literature-aware, and integrative agents, the system achieves more robust, interpretable, and clinically meaningful insights. We discuss supporting evidence, potential challenges, and a research agenda, emphasizing the paradigm's importance for precision oncology.

1 Introduction

Cancer remains a leading cause of mortality worldwide, with nearly 20 million new cases and 9.7 million deaths (including nonmelanoma skin cancers [NMSCs]) in 2022. About one in five people will develop cancer, and roughly one in nine men and one in twelve women will die from it [2]. Its heterogeneous nature, with diverse molecular subtypes and therapeutic responses, necessitates sophisticated computational approaches for accurate diagnosis, prognosis, and treatment selection [18, 14].

Bulk RNA sequencing (RNA-seq) is a fundamental technology for profiling the transcriptomic landscape of cancer, revealing gene expression patterns, pathway dysregulation, and molecular subtypes [11, 17]. Resources such as The Cancer Genome Atlas (TCGA) provide extensive datasets across cancer types, enabling large-scale computational analyses [19].

Despite these advances, traditional machine learning models often generalize poorly across cohorts and experimental conditions. Efforts to improve interpretability exist [7, 20], but models still lack transparency, limiting clinical applicability.

Large language models (LLMs) have demonstrated remarkable capabilities in natural language processing, code generation, and scientific reasoning [3, 15], as well as in biological applications like protein structure prediction, drug discovery, and genomic analysis [9, 12, 5]. However, their use in bulk transcriptomic analysis, especially within multi-agent frameworks, remains largely unexplored.

We argue that the future of bulk RNA-seq analysis and cancer prediction lies in a specialized multi-agent paradigm powered by large language models (LLMs). Single-model pipelines, while effective for narrow tasks, struggle with scalability, interpretability, and clinical translation. In contrast, a multi-agent architecture where dedicated LLM agents handle quality control, feature extraction, pathway inference, and predictive modeling offers superior performance and transparency. We argue that such distributed, collaborative systems are not optional enhancements but essential for precision oncology, turning computational predictions into actionable, trustworthy insights for patient care.

Submitted to 39th Conference on Neural Information Processing Systems (NeurIPS 2025). Do not distribute.

2 Background and Current Challenges

2.1 Bulk RNA Sequencing in Cancer Research

Bulk RNA-seq technology measures the average gene expression across all cells in a tissue sample, providing a comprehensive snapshot of the transcriptomic state [21]. In cancer research, bulk RNA-seq data has been instrumental in identifying molecular subtypes, predicting treatment responses, and understanding disease mechanisms [6].

The typical bulk RNA-seq analysis workflow involves several computational steps: quality control, read alignment, quantification, normalization, and downstream analysis including differential expression analysis, pathway enrichment, and predictive modeling. Each step introduces potential sources of bias and technical variation that can impact downstream interpretations.

2.2 Limitations of Current Computational Approaches

Traditional computational approaches for bulk RNA-seq include:

- **Statistical Methods:** Tools like DESeq2 [13] and edgeR [16] identify differentially expressed genes. *Limitation:* Cannot capture complex gene-gene interactions.
- **Dimensionality Reduction:** PCA and t-SNE are used for visualization and exploration. *Limitation:* May distort biologically meaningful relationships and is sensitive to noise.
- **Machine Learning Models:** Random forests, SVMs, and neural networks predict cancer outcomes. *Limitations:* Poor cross-cohort generalization, limited interpretability, and difficulty integrating multi-modal data (clinical, genomic, literature).

2.3 Promise and Gaps of LLMs in Biology

Large language models (LLMs), particularly transformer-based architectures, excel at sequential data analysis and have been applied successfully in biology:

- **Protein sequences:** ProtBERT [5] and ESM-2 [12] achieve state-of-the-art structure prediction and functional annotation.
- **Genomic sequences:** [8] and Nucleotide Transformer [4] predict regulatory elements and functional regions.
- **Biomedical text:** BioBERT [10] and SciBERT [1] enhance entity recognition, relation extraction, and literature mining.

However, LLM applications to bulk transcriptomic data remain limited.

3 Our Position: Multi-Agent, Multi-Modal Transcriptomic Analysis

We propose a multi-agent framework grounded in the principle that complex biological problems benefit from specialized expertise and collaborative reasoning. Analogous to interdisciplinary cancer research, this system distributes tasks across agents, each focusing on a specific modality or aspect of bulk RNA-seq analysis. This multi-modal, multi-agent design offers:

- **Modularity:** Agents can be independently trained, updated, and validated.
- **Specialization:** Each agent focuses on a distinct modality, improving performance over generalist models.
- **Robustness:** Distributed agents provide resilience against individual failures.
- **Interpretability:** Agent-specific outputs enhance transparency and biological insight.

3.1 Agent Types and Capabilities

We define four specialized agents handling different data modalities: gene expression matrices, RNA sequences, biomedical text, and integrative synthesis. Table 1 summarizes their objectives, inputs, architectures, and key capabilities.

Table 1: Specialized agents in the multi-agent, multi-modal framework for bulk RNA-seq analysis.

Agent Type	Input Modality	Architecture	Key Capabilities
Gene Expression Agent	Expression matrices (TCGA, other bulk RNA-seq datasets)	Transformer adapted for tabular data with attention mechanisms	Identify latent transcriptomic signatures, classify samples and subtypes, generate synthetic data for augmentation, enable cross-cohort generalization
RNA Sequence Agent	Raw RNA sequences, splice junctions, variants	DNA/RNA LLMs (DNABERT, Nucleotide Transformer)	Predict expression from sequence features, capture structural and regulatory patterns, integrate sequence context with expression data
Literature-Aware Agent	Biomedical text, clinical trials, curated databases	Pretrained biomedical LLMs (BioBERT) fine-tuned for genomics	Provide biological interpretation of expression patterns, validate findings, integrate prior knowledge to improve prediction and interpretability
Integration Agent	Multi-modal outputs from other agents	Multi-modal transformers with cross-attention	Fuse information from all agents, generate consensus predictions, quantify uncertainty, support interpretable and clinically relevant decision-making

3.2 Communication and Coordination

Effective multi-agent analysis relies on structured communication:

- Each specialized agent independently analyzes its data modality.
- Integration agents synthesize outputs to produce consensus predictions.
- Meta-learning mechanisms iteratively refine agent communications.

This architecture ensures robust multi-modal reasoning and interpretable framework for cancer prediction and transcriptomic analysis. Figure 1 shows the overall architecture of the proposed multi-agent framework for bulk RNA-seq analysis.

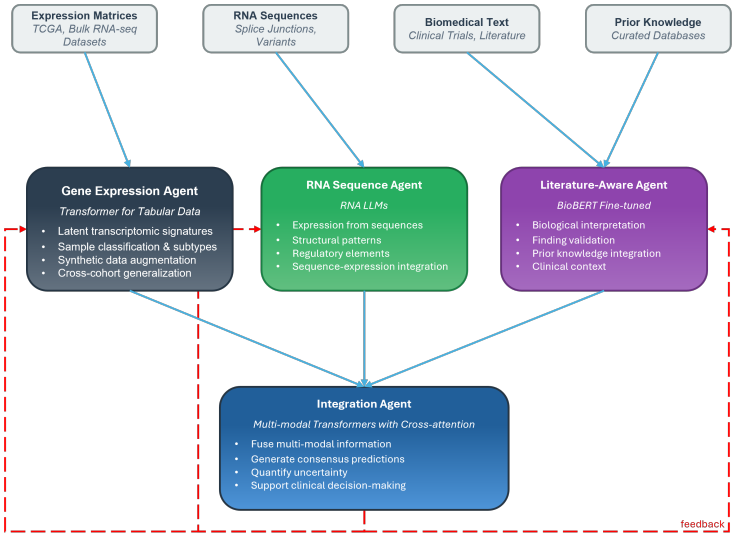

Figure 1: Multi-agent framework integrating gene expression matrices, RNA sequences, biomedical literature, and prior knowledge through specialized agents coordinated by an Integration Agent.

4 Arguments Supporting Our Position

The multi-agent, multi-modal framework provides several key advantages for cancer prediction, analysis, and translational relevance:

- **Enhanced Predictive Performance:** By integrating complementary information from expression patterns, RNA sequences, and literature knowledge, the system captures a more comprehensive view of cancer biology. Literature-aware agents provide biological constraints that improve generalization across cohorts, while ensemble-like predictions from multiple agents enable robust uncertainty quantification.
- **Synthetic Data Generation:** Gene Expression Agents can produce biologically plausible synthetic samples, which help address class imbalance in rare cancer subtypes, enable privacy-preserving analyses, and support hypothesis generation by exploring uncharted regions of expression space.
- **Interpretability and Clinical Relevance:** Each agent produces outputs that can be mechanistically interpreted, fostering expert trust and facilitating translational application of predictions in clinical settings.

5 Counterarguments and Rebuttals

While multi-agent systems offer significant benefits, critics raise several concerns. We address these challenges as follows:

- **Technical Challenges:** Critics may argue that multi-agent systems are computationally expensive, sensitive to heterogeneous data, and difficult to coordinate. We respond that:
  - Advances in distributed computing, model compression, and efficient attention mechanisms reduce computational burden.
  - Standardized preprocessing pipelines and data harmonization strategies mitigate issues arising from data heterogeneity.
  - Careful system design ensures effective coordination and communication between agent.
- **Scientific Challenges:** Skeptics point to risks such as hallucination, overfitting, and poor reproducibility. Our countermeasures include:
  - Rigorous validation and benchmarking using curated datasets and expert oversight to ensure reliability and biological relevance.
  - Monitoring and evaluation to prevent overfitting and to identify fske predictions.

6 Future Directions

Looking forward, we identify key areas to advance multi-agent, multi-modal systems:

- **Technical Priorities:**
  - Optimize agent architectures, communication protocols, and coordination mechanisms for transcriptomic analysis.
  - Improve training efficiency via transfer learning, few-shot learning, and continual learning strategies.
  - Develop comprehensive evaluation frameworks assessing both predictive performance and biological interpretability.
- **Biological Validation:**
  - Systematically validate predictions using cell lines, patient samples, and clinical cohorts.
  - Engage domain experts to assess clinical relevance of outputs.

7 Conclusion

We position **multi-agent LLM architectures as a paradigm shift** for bulk RNA-seq analysis and cancer prediction. By integrating expression data, sequence features, and literature knowledge through specialized agents, these systems overcome current barriers of generalization, interpretability, and integration. We argue this approach is not just a technical advance, but a critical step toward truly precise and clinically relevant oncology.

References

[1] Iz Beltagy, Kyle Lo, and Arman Cohan. Scibert: A pretrained language model for scientific text. *arXiv preprint arXiv:1903.10676*, 2019.

[2] Freddie Bray, Mathieu Laversanne, Hyuna Sung, Jacques Ferlay, Rebecca L Siegel, Isabelle Soerjomataram, and Ahmedin Jemal. Global cancer statistics 2022: Globocan estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA: a cancer journal for clinicians*, 74(3):229–263, 2024.

[3] Yupeng Chang, Xu Wang, Jindong Wang, Yuan Wu, Linyi Yang, Kaijie Zhu, Hao Chen, Xiaoyuan Yi, Cunxiang Wang, Yidong Wang, et al. A survey on evaluation of large language models. *ACM transactions on intelligent systems and technology*, 15(3):1–45, 2024.

[4] Hugo Dalla-Torre, Liam Gonzalez, Javier Mendoza-Revilla, Nicolas Lopez Carranza, Adam Henryk Grzywaczewski, Francesco Oteri, Christian Dallago, Evan Trop, Bernardo P de Almeida, Hassan Sirelkhatim, et al. Nucleotide transformer: building and evaluating robust foundation models for human genomics. *Nature Methods*, 22(2):287–297, 2025.

[5] Ahmed Elnaggar, Michael Heinzinger, Christian Dallago, Ghalia Rehawi, Yu Wang, Llion Jones, Tom Gibbs, Tamas Feher, Christoph Angerer, Martin Steinegger, et al. Prottrans: towards cracking the language of life's code through self-supervised learning. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 44:7112–7127, 2021.

[6] Maxence Gélard, Guillaume Richard, Thomas Pierrot, and Paul-Henry Cournède. Bulkrnabert: Cancer prognosis from bulk rna-seq based language models. *bioRxiv*, pages 2024–06, 2024.

[7] Vikas Hassija, Vinay Chamola, Atmesh Mahapatra, Abhinandan Singal, Divyansh Goel, Kaizhu Huang, Simone Scardapane, Indro Spinelli, Mufti Mahmud, and Amir Hussain. Interpreting black-box models: a review on explainable artificial intelligence. *Cognitive Computation*, 16(1):45–74, 2024.

[8] Yanrong Ji, Zhihan Zhou, Han Liu, and Ramana V Davuluri. Dnabert: pre-trained bidirectional encoder representations from transformers model for dna-language in genome. *Bioinformatics*, 37(15):2112–2120, 2021.

[9] John Jumper, Richard Evans, Alexander Pritzel, Tim Green, Michael Figurnov, Olaf Ronneberger, Kathryn Tunyasuvunakool, Russ Bates, Augustin Žídek, Anna Potapenko, et al. Highly accurate protein structure prediction with alphafold. *nature*, 596(7873):583–589, 2021.

[10] Jinhyuk Lee, Wonjin Yoon, Sungdong Kim, Donghyeon Kim, Sunkyu Kim, Chan Ho So, and Jaewoo Kang. Biobert: a pre-trained biomedical language representation model for biomedical text mining. *Bioinformatics*, 36(4):1234–1240, 2020.

[11] Xinmin Li and Cun-Yu Wang. From bulk, single-cell to spatial rna sequencing. *International journal of oral science*, 13(1):36, 2021.

[12] Zeming Lin, Halil Akin, Roshan Rao, Brian Hie, Zhongkai Zhu, Wenting Lu, Nikita Smetanin, Robert Verkuil, Ori Kabeli, Yaniv Shmueli, et al. Evolutionary-scale prediction of atomic-level protein structure with a language model. *Science*, 379(6637):1123–1130, 2023.

[13] Michael I Love, Wolfgang Huber, and Simon Anders. Moderated estimation of fold change and dispersion for rna-seq data with deseq2. *Genome biology*, 15(12):550, 2014.

[14] Corbin E Meacham and Sean J Morrison. Tumour heterogeneity and cancer cell plasticity. *Nature*, 501(7467):328–337, 2013.

[15] Humza Naveed, Asad Ullah Khan, Shi Qiu, Muhammad Saqib, Saeed Anwar, Muhammad Usman, Naveed Akhtar, Nick Barnes, and Ajmal Mian. A comprehensive overview of large language models. *ACM Transactions on Intelligent Systems and Technology*, 16(5):1–72, 2025.

[16] Mark D Robinson, Davis J McCarthy, and Gordon K Smyth. edger: a bioconductor package for differential expression analysis of digital gene expression data. *bioinformatics*, 26(1):139–140, 2010.

[17] Amarinder Singh Thind, Isha Monga, Prasoon Kumar Thakur, Pallawi Kumari, Kiran Dindhoria, Monika Krzak, Marie Ranson, and Bruce Ashford. Demystifying emerging bulk rna-seq applications: the application and utility of bioinformatic methodology. *Briefings in bioinformatics*, 22(6):bbab259, 2021.

[18] Samra Turajlic, Andrea Sottoriva, Trevor Graham, and Charles Swanton. Resolving genetic heterogeneity in cancer. *Nature Reviews Genetics*, 20(7):404–416, 2019.

[19] John N Weinstein, Eric A Collisson, Gordon B Mills, Kenna R Shaw, Brad A Ozenberger, Kyle Ellrott, Ilya Shmulevich, Chris Sander, and Joshua M Stuart. The cancer genome atlas pan-cancer analysis project. *Nature genetics*, 45(10):1113–1120, 2013.

[20] Qian Xu, Wenzhao Xie, Bolin Liao, Chao Hu, Lu Qin, Zhengzijin Yang, Huan Xiong, Yi Lyu, Yue Zhou, and Aijing Luo. Interpretability of clinical decision support systems based on artificial intelligence from technological and medical perspective: A systematic review. *Journal of healthcare engineering*, 2023(1):9919269, 2023.

[21] Wenbin Ye, Qiwei Lian, Congting Ye, and Xiaohui Wu. A survey on methods for predicting polyadenylation sites from dna sequences, bulk rna-seq, and single-cell rna-seq. *Genomics, proteomics & bioinformatics*, 21(1):67–83, 2023.