

# Review 1

**PC member:** Nesrine Mhadhbi

**Time:** Nov 21, 15:34 UTC

## Review

### 1. General Comment

The extended abstract presents a comparative study of two approaches for RNA-seq based cancer classification: a sequence-based biological language model (BioLLM) applied directly to RNA sequences, and a Graph Foundation Model (GFM) operating on compacted de Bruijn graphs. The primary aim is to assess whether explicit graph-level reasoning improves tumor prediction compared to a sequence-only approach. The study also investigates the impact of neural positional encoding on graph representation and classification performance.

### 2. Strengths

- The research question is clearly defined and highly relevant to bioinformatics and tumor classification.
- The integration of biological language models with GFMs is innovative and well-motivated, with a solid discussion of equivariance and positional encoding properties.
- The methodology is well-structured, including dataset description, tokenization, graph construction, and model adaptation.
- The expected results section logically discusses potential improvements in performance and interpretability with the graph-based approach.

### 3. Areas for Improvement

- More details on dataset size and specific evaluation metrics would strengthen the study.
- Some sentences are long and technical, which may reduce readability for non-specialists.
- The practical implications of the study for clinical applications could be further clarified.

### 4. Writing Quality

Overall, the abstract is well-organized with clear sections on background, problem statement, methodology, and expected results. Minor rewording could improve readability and flow.

**Use of LLMs:** Yes, somewhat, but the use of LLMs is reasonable.

### Reviewer's confidence

2: Partly, I may be missing some concepts or elements of the state-of-the-art, but I got the main idea.

### Usage of LLM

2: Yes, a bit.

### Confidential remarks for the program committee

The study is well-introduced and demonstrates solid understanding of language and graph models in bioinformatics. Providing more information on evaluation metrics and clinical relevance would further strengthen the document.