Point-by-point responses

We would like to thank the reviewers **R1** and **R3** for their thorough review of our manuscript. Please find our point-by-point response below.

R1.General The authors have addressed all the comments and concerns I raised in the previous review. I recommend that the paper be published in its current form.

Response We sincerely thank the reviewer for their thoughtful and constructive feedback on our manuscript introducing DeepChopper. We appreciate the recognition of our work's significance and innovation in addressing chimera artifacts in nanopore direct RNA sequencing.

R2.General The manuscript describes genomic language model based method for detecting chimeric reads in direct RNA sequencing. Overall, the manuscript is well written and thoughtfully organised, where the comments are also well addressed. I have a few questions remain after reading the manuscript, which either arise from reading the manuscript or following up to previous question which I felt more can be done to address those questions.

Response

R2.Q1 The manuscript focuses on direct RNA seq, I wonder how generalizable is the approach to cDNA seq data, would a lot effort be needed? It would be good if authors could discuss for this application given there are a lot cases where Nanopore cDNA samples are generated instead.

Response

R2.Q2 The novelty of this approach, is the application of genomic language models, I agree. However, the advantage of this model vs conventional non-large language model is not discussed enough for why this approach should be favored. One point related is the testing data being too easy, which is shown as clear bimodal distribution as presented in response figure 6, any challenging scenarios that could benefit from deepchopper but not other methods?

Response

R2.Q3 (more specific) Line 059-064: The chimera artifacts in dRNA-seq resulted from ligation (biases? I assume) or software missing open pore signal. The citation seems to describe more in a multiplexing context. I wonder is there any paper discussing this for regular ONT dRNA-seq instead? Also, the part software missing open pore signals is from paper in 2020, specifically referring to MinKNOW, which has been outdated, have the authors checked whether this is still a problem with the recent dRNA-seq?

Response

R2.Q4 Related to the handling of long-reads, as questioned by Reviewer 1, 32kb misses a few genes, for example SMAD2, LTRT, TTN (TTN actually is longer than 32kb?), etc? How are these genes dealt with then? And what impact will happen to these genes? In terms of expression levels? The read length referred to citation in the response is rather outdated, and the annotation with respect is also quite old.

Response

R2.Q5 Following Review 1 Q4 regarding RNA004, for practicality, it would be good if the authors could show fine-tuned results of deep-chopper for RNA004 in terms of performance to be more relevant in direct RNA sequencing context.

Response

R2.Q6 Following Reviewer 1 Q5 response, the authors demonstrated it in different data size, in terms of practicality, an data of size 20/25 million would also be interested to be included for this guidance.

Response

R2.Q7 Figure 1 c). Sliding window worth more explanation for better understanding. For example, from step 3 to step 4, the bottom 3rd base is refined in step 3 from 1 to 0, however, in step 4, the refined based is not used. That means this refinement strategy is always using initial prediction for refinement, what is the advantage and disadvantage of this?

Response

R2.Q8 Following R2.Q2 response, why comparison results from response figure differ so much between deepchopper and other methods.

Response

R2.Q9 Following R2.Q3 respons for the first part related to validating performance of internal adapter detection in real data, I like the example shown in Figure 5, which quite clearly show that this is a internal adapter, I wonder the sequences before the low base-quality, is a real artefact or fusion reads, and in general, what is the prevalence of the presence of such internal adapter in this real dataset to get an idea of how big the issue is in real data? Related to the second question, as to the justification of using a genomic language model over alignment-based methods, a comparison of the genomic language model vs alignment-based methods (not necessarily minimap2, but some methods build upon minimap2 alignment), for example, the recently online paper Breakinator tool (Heinz, Jakob M., Matthew Meyerson, and Heng Li. "Detecting Foldback Artifacts in Long Reads." bioRxiv (2025): 2025-07.) and flair-fusion (which aims to identify fusion by removing chimeric reads). It would worth discussion to highlight the advantage of using genomic language models in this context.

Response

R2.Q10 Following up to R2 Q4, the response does not explain how different thresholds of this classification decision. In the context of 2 labels, the sum of the 2 adapter class probability would always be 1, that is, the classification of 1 label probability being greater than the other label is still 0.5 threshold. Please clarify and state scenarios where the 2 probabilities does not sum to 1, cause it's hard to understand why in the case of 2 labels, the sum would not be 1. Related to this, response figure 6 is very clean indeed, I wonder is there some kind overfitting or testing data is too simple? Please elaborate for challenging scenarios.

Response