Abstract

SeqBIM

Nanopore MinION long read sequencer: an overview of its error landscape

Clara Delahaye[1]*, Jacques Nicolas[2]

[1,2] *Univ Rennes, Inria, CNRS, IRISA F-35000, Rennes, France*
***Corresponding author**: clara.delahaye@irisa.fr

Abstract

Third generation ONT's (Oxford Nanopore Technologies) sequencer provides longer DNA fragments (mean read length often over 10 kB) than usual second generation sequencers. However, this technology is also more error-prone [1], with currently around 6% of error on raw reads. Many articles worked on read correction methods (there even is a tool to assess error correction methods [2]), while few addressed the detailed characterization of observed errors [3], as the frequent (almost monthly!) updates in ONT chemistry and softwares hinder the task. The MinION sequencer is now getting more stable. We propose here an up-to-date view of its error landscape, using state-of-the-art flowcell and basecaller. We worked on bacterial and human data to get an overview of Nanopore sequencing error biases.

As opposed to usual NGS, Nanopore sequencing does not requires PCR amplification, thus one expects that this technology would not suffer from GC bias. Yet we found that it is actually a decisive factor linked to sequencing errors. In particular, low-GC reads have almost 2% fewer errors than high-GC reads. Nanopore sequencers are also known to struggle sequencing accurately repeated regions (homopolymers or regions with short repeats). Our work highlighted that for these regions, being the source of about half of all sequencing errors, the error profile also depends on the GC content and shows mainly deletions, although there are some reads with long insertions. Another interesting finding is that the quality measure offers valuable information on the error rate as well as the abundance of reads.

Overall we hope this work will help designing more accurate methods for error correction.

References

[1] C. L. Ip, M. Loose, J. R. Tyson, M. de Cesare, B. L. Brown, M. Jain, R. M. Leggett, D. A. Eccles, V. Zalunin, J. M. Urban, et al. Minion analysis and reference consortium: Phase 1 data release and analysis. *F1000Research*, 4, 2015.

[2] C. Marchet, P. Morisse, L. Lecompte, A. Lefebvre, T. Lecroq, P. Peterlongo, and A. Limasset. ELECTOR: evaluator for long reads correction methods. *NAR Genomics and Bioinformatics*, 2(1), 11 2019. lqz015.

[3] R. R. Wick, L. M. Judd, and K. E. Holt. Performance of neural network basecalling tools for oxford nanopore sequencing. *Genome Biology*, 20(1):129, 2019.