- Abstract and Introduction:

- **Long-read sequencing offers a number of advantages over short-read sequencing.**
- **Although Short-reads are effective in cost, accurate, and supported by analysis tools and pipelines, sequencing nucleic acid polymers in short fragments is difficult.**
- **Long-read technologies are good at accuracy and generate reads in excess of 10 kb, so the characteristics of long-read data should be focused on the analysis tools.**
- **This paper will study error correction, base modification detection, and long-read transcriptomics analysis and the challenges.**
- **It will Describe the principles of long-read data analysis.**
- **It also introduces open-source catalogue of long-read analysis tools: long-read-tools.org.**
- **short-read sequencers such as Illumina's NovaSeq, HiSeq, NextSeq, and MiSeq instruments ; BGI's MGISEQ and BGISEQ models ; or Thermo Fisher's Ion Torrent sequencers.**
- **long-read sequencing technologies such as Pacific Biosciences' (PacBio) single-molecule real-time (SMRT) sequencing and Oxford Nanopore Technologies' (ONT) nanopore sequencing.**
- **Long-reads can also improve de novo assembly, mapping certainty, transcript isoform identification, and detection of structural variants.**
- **we check available tools to deal with long-read sequencing projects.**

- Related Work:

I) Pollard MO, Gurdasani D, Mentzer AJ, Porter T, Sandhu MS. Long reads: their purpose and place. Hum Mol Genet. 2018; 27(R2):234–41. https://doi.org/10.1093/hmg/ddy177.

II) Burgess DJ. Genomics: next regeneration sequencing for reference genomes. Nat Rev Genet. 2018; 19(3):125. https://doi.org/10.1038/nrg.2018.5.

III) Bentley DR, Balasubramanian S, Swerdlow HP, Smith GP, Milton J, Brown CG, et al.Accurate whole human genome sequencing using reversible terminator chemistry. Nature. 2008; 456(7218):53–9. https://doi.org/10.1038/nature07517.

IV) Goodwin S, McPherson JD, McCombie WR. Coming of age: ten years of next-generation sequencing technologies. Nat Rev Genet. 2016; 17(6):33351. https://doi.org/10.1038/nrg.2016.49.

V) Jeon SA, Park JL, Kim J-H, Kim JH, Kim YS, Kim JC, et al.Comparison of the MGISEQ-2000 and Illumina HiSeq 4000 sequencing platforms for RNA sequencing. Genomics Inform. 2019; 17(3):e32.

VI) Rothberg JM, Hinz W, Rearick TM, Schultz J, Mileski W, Davey M, et al.An integrated semiconductor device enabling non-optical genome sequencing. Nature. 2011; 475(7356):348–52. https://doi.org/10.1038/nature10242.

VII) Quail M, Smith ME, Coupland P, Otto TD, Harris SR, Connor TR, et al.A tale of three next generation sequencing platforms: comparison of Ion torrent, Pacific Biosciences and Illumina MiSeq sequencers. BMC Genomics. 2012; 13(1):341. https://doi.org/10.1186/1471-2164-13-341.

- **There are open-source static catalogues (e.g. github.com/B-UMMI/long-read-catalog), custom pipelines developed by individual labs for specific purposes, and others that attempt to generalise them for a wider research community.**
- **For this purpose, we introduce long-read-tools.org, a timely database that comprehensively collates tools used for long-read data analysis.**
- **In addition to true long-read sequencing technologies (SMRT and nanopore), we include synthetic long-read strategies. We include them in our database for completeness but indicate when they have been superseded or are no longer maintained.**
- **long-read-tools.org is an open-source project under the MIT License, whose code is available through GitHub [161].**

- Result:

- **Nanopore reads-based transcriptomes are more recent , and work is still needed to understand the characteristics of these data (e.g. coverage bias, sequence biases, reproducibility). Certain isoform assembly pipelines predict a large number of unannotated isoforms requiring validation and classification.**
- **Long reads theoretically confer huge advantages over short reads for transcript-level differential expression, however the low-level of replication and modest read counts obtained from long-read transcriptomic experiments are currently limiting.**
- **Until throughput increases and price decreases sufficiently, hybrid approaches that use long reads to define the isoforms expressed in the samples and short reads to get enough counts for well-powered differential expression may be successful; these do not yet exist.**