Nanopore Sequencing in Human Genetic Studies: Application to Structural Variant Detection

Catarina Silva[1,2] & Luís Vieira[1,2]
[1]Departamento de Genética Humana, Instituto Nacional de Saúde Doutor Ricardo Jorge, Lisboa
[2]ToxOmics - Centro de Toxicogenómica e Saúde Humana, Nova Medical School, Faculdade de Ciências Médicas, Universidade Nova de Lisboa, Lisboa

Background: Nanopore sequencing is a recent technology which allows direct real-time sequencing of DNA or RNA molecules and production of read lengths as long as the size of the original fragments. It is based on the application of an ionic current to move DNA strands through an array of nanopores embedded in an electrically-resistant membrane. An ASIC chip placed below this membrane is responsible for recording signal changes in the nanopore, which are then translated into a specific base sequence. Although major improvements have been made in the genomic short-read sequencing technologies, structural variants (SV) detection still stands as a challenge. The characteristics of nanopore sequencing make it well suited for whole genome sequencing and as a preferential tool to identify SV, namely those associated with tumours. In this work we present a complete workflow to detect SV in tumour samples using nanopore sequencing.

Methods: We used a test tumour sample to prepare DNA libraries using the rapid sequencing kit (Oxford Nanopore Technologies-ONT) and sequenced those in a single R9.4 flow cell on the MinION device (ONT). We implemented a pipeline for SV detection using multiple bioinformatics tools: MinKNOW software (ONT) (run set up and data acquisition), Albacore pipeline (ONT) (basecalling), LAST aligner (mapping) and Picky (SV detection and analysis).

Results: The use of the rapid sequencing kit allowed a fast preparation of sequencing libraries when compared with standard library preparation procedures. The MinION generated a total of 2.34 Gb of DNA. The highest number of reads were obtained during the first 8-10h of the sequencing run. Overall, 87.8% of reads had a quality (Q) value >7 (quality threshold) and 54.2% had a Q value >10. The longest read obtained had 74369 bases and the mean read length was of 4508 bases. All reads with a length >50 kb had Q values >= 7. A total of 3470 SV were identified, including deletions, duplications, translocations, insertions and inversions.

Conclusions: Long-read sequencing technology is becoming an important tool as it poses itself as a powerful complementary method that suppresses the inherent limitations of other technologies. Nanopore sequencing is a fast and sensitive approach of great potential for human genomics research and for use in a clinical perspective, namely in the detection of SV in the human genome. A critical issue in nanopore sequencing is DNA quality and integrity because longer size reads make

This work was supported by Toxomics (Centre for Toxicogenomics and Human Health), Nova Medical School and is a result of the GenomePT project (POCI-01-0145-FEDER-022184), supported by COMPETE 2020 - Operational Programme for Competitiveness and Internationalisation (POCI), Lisboa Portugal Regional Operational Programme (Lisboa2020), Algarve Portugal Regional Operational Programme (CRESC Algarve2020), under the PORTUGAL 2020 Partnership Agreement, through the European Regional Development Fund (ERDF), and by Fundação para a Ciência e a Tecnologia (FCT).