Targeted long-read Nanopore sequencing of the blood group genome by adaptive sampling

M. Gueuning[1], G. A. Thun[1], S. Meyer[2], M. P. Mattle-Greminger[1]

[1] Blood Transfusion Service Zurich (SRC), Department of Research and Development, Schlieren, Switzerland
[2] Blood Transfusion Service Zurich (SRC), Department of Molecular Diagnostics and Cytometry, Schlieren, Switzerland

Background: Oxford Nanopore Technologies (ONT) has recently introduced a novel approach to long-read sequencing which eliminates off-target reads, i.e. DNA fragments not covering a region of interest, in real-time during sequencing. This so called 'adaptive sampling' allows selective enrichment of regions of interest, for instance the complete blood group genome. The process is entirely computationally and does not require the development of laborious wet-lab enrichment protocols. Furthermore, it can be adapted to any genomic region in a few minutes, making it exceptionally versatile. Adaptive sampling does not impose restrictions on DNA fragment length, which is particularly advantageous for phasing variants across target loci and for resolving paralogous gene regions encompassing structural variants such as found in the RH and MNS blood group systems.

Aim: The objective of this study was to explore the effectiveness of Nanopore adaptive sampling with regard to sequencing blood group genes. We applied the sequencing strategy to a set of selected samples, focused on complex structural variants.

Methods: Genomic high-molecular weight DNA was used to build Nanopore sequencing libraries. Each sample was run on a separate MinION flow cell (ONT). To increase total sequencing output, flow cells were washed and re-loaded with library several times. The reference FASTA file conveying the genomic regions to enrich, contained all known red cell blood group genes (n=39), the transcription factor genes *GATA* and *KLF1*, as well as the human platelet antigen genes (n=7). We included 50 kb flanking regions for each gene to increase the chance of retrieving long on-target reads. In sum, we targeted ~ 7 Mb. Reads were mapped to the novel human reference genome (T2T-CHM13v2.0) and, in the case of complex structural variation, also *de-novo* assembled. To assess variant calling accuracy, ONT sequencing results were compared to pre-typed genetic data where applicable.

Results: The mean length of sequenced reads was high with N50 of >30 kb. The maximum read length achieved was over 500 kb. Reads were categorized as either on- or off-target within the first 500 bp (~1 second of sequencing). From the ~10% on-target fraction (>100,000 reads per sample), only ~2% mapped to our regions of interest. This resulted in an expected enrichment of around 5-10x. All target genes were fully covered by at least 5x coverage, with a median coverage of 15x across all genes. Although this was not yet sufficient for reliable single-nucleotide variant calling (~10% of expected calls below quality threshold), it allowed in combination with very long reads (~10 reads >100 kb for the RH and MNS locus per sample) to resolve complex structural variants.

Conclusions: Nanopore adaptive sampling has emerged as a promising tool for cost-effective and straightforward long-read sequencing of all blood group genes on single MinION flow cells. A first evaluation showed still suboptimal coverage for general variant calling, but intriguing potential to resolve complex structural variants in the paralogous RH and MNS regions, a task in which conventional molecular techniques usually fail. Since adaptive sampling is still in the early stages of development, it is anticipated that enrichment efficiency and thus sequencing coverage will further increase.

ISBT 2023