Extending long-read capture enrichment for Nanopore sequencing to the MNS blood group locus to resolve complex haplotypes

S Koller[1,2], M Gueuning[1], GA Thun[1], S Sigurdardottir[2], S Meyer[1,2]

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

Introduction: Accurately assessing the allelic composition of the highly homologous genes in the RH and MNS blood group systems is challenging due to frequent genetic rearrangements. Owing to limitations of the currently most successful targeted Nanopore sequencing approaches (long-range PCR and adaptive sampling), we recently tested a capture-enrichment strategy targeting *RH* genes, which showed promising results. Here, we aimed to optimise output and workflow of this method by including genes from the MNS system (adding up to a total of 6 targeted genes: *RHD*, *RHCE*, *RHAG*, *GYPA*, *GYPB* and *GYPE*), doubling sample size per flow cell and trying to further enrich captured on-target reads.

Methods: Capture probes were designed to target: (I) the full MNS locus, including *GYPE* (~290 kb); (II) the full *RH* locus (~178 kb); and (III) the *RHAG* gene (~51 kb), all with >10 kb flanking regions on each side. Sixteen samples were pooled and sequenced on a MinION flow cell, including 4 with suspected *GYP*401* MNS alleles. To further increase enrichment, adaptive sampling was performed on half of the pores. Coverage uniformity, phasing accuracy and resolution of hybrid alleles were evaluated using both *de novo* and reference-based variant-calling pipelines.

Results: The sequencing run generated 27.3 Gb in total, with comparable data output from pores with and without adaptive sampling. However, the former produced more than twice as many reads (9.63M vs 4.07M), resulting in higher on-target coverage overall (~14,900x vs 7,700x). Per sample, mean coverage across the targeted regions ranged from 476x to 1051x with an N50 of ~3.2 kb. Initial *de novo* assembly of *GYP*401* hybrid alleles looked very promising despite the moderate read-length distribution, but further molecular analyses are needed to confirm breakpoint assignment. Variant phasing within and between *RHD* and *RHCE* was successful in most (non-hybrid) samples, but failed when alleles were little polymorphic.

Conclusion: This study corroborates that long-read capture enrichment combined with Nanopore sequencing is effective for analysing MNS and RH hybrid alleles. It provided more uniform coverage and improved breakpoint resolution than amplicon-based methods, and was much more cost-effective than adaptive sampling. Protocol optimisation to obtain longer reads (a test run already reached N50 of ~5.0 kb) will further reduce mapping ambiguity and improve phasing, strengthening the use of this strategy for sequencing complex blood group genes.