# Long-read capturing for Nanopore sequencing RH and MNS blood group loci to resolve full haplotypes

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

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

**Background:** Determining the complete allelic composition of the highly homologous genes in the RH and MNS blood group systems is essential for transfusion medicine, yet this remains challenging due to their complex genomic organisation. Despite their ability of phasing genetic variants, existing targeted approaches for Nanopore sequencing – such as overlapping long-range PCR and adaptive sampling – encounter limitations, including allele drop-out, uneven coverage, and modest enrichment efficiencies. To overcome these challenges, capture-enrichment protocols for long reads were evaluated. In a first approach, *RHD* and *RHCE* were targeted, complemented by the *RHAG* gene, which can modulate RhD/RhCE expression. The homologous genes of the *GYP* locus were added as targets in a second protocol, together with further optimisations with respect to output and enrichment of captured on-target reads.

**Methods:** Custom capture probes were designed to span the genomic interval from *RHD* (~57 kb) to *RHCE* (~59 kb) and the *RHAG* gene (~31 kb), complemented by at least 10 kb of flanking regions on either side. The probe panel was later extended to the 3 genes of the *GYP* locus (~290 kb, including flanking regions). For the first proof-of-concept study, eight samples were pooled and sequenced on a MinION flow cell. In a second approach, we doubled samples per flow cell and applied adaptive sampling 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:** Both approaches produced evenly distributed coverage over the targeted regions with an average depth of ~1600× and ~800× (with 16 samples), respectively. N50 read length was in the range of 3.2 to 3.7 kb. Within the second approach, pores with adaptive sampling produced an almost two-fold higher on-target coverage with similar data output compared with pores without adaptive sampling. Variant phasing within and between *RHD* and *RHCE* was successful in most non-hybrid samples, but failed when alleles were little polymorphic. In a reference-based setting, hybrid allele breakpoints could be defined with high resolution—particularly when a *RHCE*02* reference sequence was used to reduce mismappings in RH2-positive samples. Initial attempts at *de novo* assembling hybrid alleles looked particularly promising for *GYP*401* alleles, but were impeded by the moderate read length distribution.

**Conclusions:** This study demonstrates that long-read capture enrichment combined with Nanopore sequencing can effectively analyze complex alleles of the MNS and RH blood group systems. The method offers significant advantages over amplicon-based techniques by providing more uniform coverage and improved breakpoint determination, and it is more cost-effective than adaptive sampling. Further protocol optimizations to obtain longer reads (a test run already reached N50 of ~5.0 kb) will strengthen the use of this strategy for sequencing complex blood group genes. It will particularly reduce mapping ambiguity and improve phasing in reference-based variant calling pipelines, as well as facilitate *de novo* assembly of complex hybrid alleles.