35th Regional Congress of the ISBT, 2025, Milan, Italy

Capture-enrichment of *RHD* and *RHCE* for Nanopore long-read sequencing to resolve complex haplotypes: advantages and challenges

S. Koller[1,2], M. Gueuning[1], G. A. Thun[1], S. Sigurdardottir[2], N. Trost[2], K. Engehausen[2], C. Engström[3], S. Meyer[2]

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

Background
Elucidating the allelic composition of the extended and highly homologous RHD and RHCE genes at the haplotype level, which holds significant transfusion relevance, is challenging. Long-read sequencing is the method of choice for phasing, however, targeted long-read sequencing of genes of interest is still challenging. To date, two approaches have been used for targeted RHD and RHCE sequencing with Nanopore technology: overlapping long-range PCRs (amplicon-based method) and adaptive sampling. However, both methods have limitations: while the PCR approach has the risk of allele drop-out and uneven coverage distribution, the adaptive sampling is cost intensive due to modest enrichment capability (maximal 2-4 samples/PromethION flow cell). To overcome these challenges, we evaluated a long-read capture enrichment approach for targeted sequencing of RHD and RHCE. In addition, the RHAG gene, known to potentially weaken or silence RhD and RhCE expression, was also included in this sequencing panel.

Aims
In this study, we evaluated a long-read capture enrichment Nanopore sequencing protocol for RHD, RHCE and RHAG, with the aim of comparing the sequencing data with well-characterised samples on the haplotype level. These samples had previously been analyzed using either adaptive sampling or the amplicon-based approach on a Nanopore sequencer. For one sample in this trial, only Sanger sequencing data were available.

Methods
Custom probes were designed to capture the genomic region spanning RHD (~57 kb) to RHCE (~59 kb) along with flanking regions of 75 kb on both sides. The RHAG gene including 10 kb flanking regions was also targeted by the probes. In the first proof of concept, 8 samples were pooled for sequencing on one MinION flow cell. Bioinformatic analysis included both a de novo and a reference-based variant calling approach.

Results
The samples showed an evenly distributed coverage of the target region spanning RHD to RHCE with a mean coverage of about 1600x and N50 of 3.7 kb. Within the samples, between 37.8 and 41.7% of the reads mapped to the targeted regions. Phasing of variants was successful within and between RHD and RHCE in a non-hybrid sample. All exonic SNVs were in agreement with previous results and break points in hybrid alleles could be determined with high resolution in a reference-based setting. A

RHCE*02 based reference sequence proved essential to avoid mismappings in RhC positive samples. First attempts to de novo assemble hybrid alleles, as an alternative method that does not rely on reference sequences, were unsuccessful, likely due to the relatively low N50.

Summary/Conclusions

To our knowledge, we represent here for the first time long-read capture enrichment of RHD and RHCE used in combination with Nanopore sequencing. Sequencing data were promising and suggest improvements over amplicon-based sequencing due to more even coverage, in particular for breakpoint determination of hybrid alleles. Compared to adaptive sampling, enrichment by capturing proved advantageous especially in terms of cost. While read length limitations may pose challenges in mapping or phasing, further optimization of the protocol to achieve a higher N50 of up to 7 kb is desirable, making long-read capture enrichment a promising approach for full-length sequencing of RHD and RHCE, as well as other complex blood group systems.