Reference sequences for *ABO* alleles by long-read sequencing reveal putative *A1*-diagnostic variants

Authors: Maja P. Mattle-Greminger[1], Morgan Gueuning[1§], Gian Andri Thun[1§], Michael Wittig[2], Anna-Lena Galati[3], Stefan Meyer[4], Nadine Trost[4], Elise Gourri[1,4], Janina Fuss[2], Sonja Sigurdardottir[4], Yvonne Merki[4], Kathrin Neuenschwander[4], Yannik Busch[3], Peter Trojok[3], Marco Schäfer[3], Jochen Gottschalk[5], Andre Franke[2], Christoph Gassner[2,6], Wolfgang Peter[3,7], and Beat M. Frey[1,4,5]

Affiliations:
[1]Department of Research and Development, Blood Transfusion Service Zurich, Swiss Red Cross, Schlieren, Switzerland
[2]Institute of Clinical Molecular Biology, Christian-Albrechts-University of Kiel, Kiel, Germany
[3]Stefan-Morsch-Foundation, Birkenfeld, Germany
[4]Department of Molecular Diagnostics and Cytometry, Blood Transfusion Service Zurich, Swiss Red Cross, Schlieren, Switzerland
[5]Blood Transfusion Service Zurich, Swiss Red Cross, Schlieren, Switzerland
[6]Institute for Translational Medicine, Private University in the Principality of Liechtenstein, Triesen, Liechtenstein
[7]Institute for Transfusion Medicine, Faculty of Medicine and University Hospital Cologne, University of Cologne, Cologne, Germany

[§]These authors contributed equally to this work

Background: With emerging blood group genomics, reference collections of blood group gene alleles have gained high importance for genotyping and sequencing. For most blood groups, however, such allele haplotype sequences remain rare. The main obstacle lies in resolving haplotypes with both Sanger and short-read sequencing. Using single-molecule, long-read Oxford Nanopore sequencing, we aimed to generate a collection of full-length haplotype sequences for all six main *ABO* allele groups: *ABO*A1, A2, B, O.01.01, O.01.02, and O.02*. Together, these groups cover 99.9% of the genetic diversity at *ABO* in Switzerland.

Methods: We selected 77 samples from a well-characterized *ABO* genotype dataset (n=25,200) of serologically-typed blood donors from the Zurich area (Switzerland). The entire *ABO* gene was amplified in two overlapping generic long-range PCRs (covering ~23.6 kb). Amplicons were sequenced on MinION flow cells of Oxford Nanopore Technologies. For quality validation, two samples per *ABO* group were re-sequenced using an Illumina/PacBio hybrid approach.

Results: All 154 *ABO* sequences could be resolved as full-length haplotypes. Median depth of coverage was 1454x per PCR amplicon. Cross-validation with PacBio/Illumina data confirmed high quality of nanopore sequences. Our haplotype collection uncovered hitherto unknown distinct sequence patterns among *ABO* groups, which had not yet been unveiled due to the very few complete human *ABO* gene sequences available so far. Most genetic diversity was found between, not within *ABO* groups, with a high degree of fixed differences. Phylogenetic tree and haplotype network analyses highlighted distinct clades of each *ABO* group. Strikingly, our data uncovered four genetic variants putatively specific for *ABO*A1*, for which direct diagnostic targets are currently lacking. We validated *A1*-diagnostic sensitivity and specificity using whole-genome data (n=4,872) of a multi-ethnic cohort.

Conclusions: Overall, our nanopore sequencing strategy proved powerful for producing high-quality *ABO* haplotypes and holds promise for generating similar collections for other blood groups. The publicly available collection of 154 haplotypes will serve as a valuable resource for molecular analyses of *ABO*, as well as studies about function and evolutionary history of *ABO*.