Vox Sanguinis ISBT International Society of Blood Transfusion

P-024 | Complete reference sequences for ABO blood group gene alleles by long-read sequencing

selected for main programme

M. P. Mattle-Greminger[1], M Gueuning[1], G. A. Thun[1], M Wittig[2], A. L. Galati[3], S Meyer[4], J Fuss[2], S Sigurdardottir[4], N Trost[4], Y Merki[4], K Neuenschwander[4], E Gourri[1, 4], Y Busch[3], J Gottschalk[5], A Franke[2], B. M. Frey[5], C Gassner[2, 6], W Peter[3, 7]

[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]Transfusion Medicine, University Hospital of Cologne, Cologne, Germany

Background: Defining high-quality allele reference sequences for blood group genes has become increasingly important, in particular as next-generation sequencing (NGS) is being more widely used in molecular analysis of blood groups. Allele reference sequences need to (i) span the complete gene region including introns, (ii) have a fully resolved haplotype, (iii) offer confirmed serology, and (iv) be well-accessible in a public sequence repository. Generating such sequences is technically challenging. The main obstacle lies in resolving solid haplotypes, as both Sanger sequencing and short-read NGS are restricted by their read length (and therefore physical phase information). Hence, for many blood group genes, allele reference sequences remain rare. Even for ABO, the clinically most important and well-known system, only nine complete human *ABO* gene sequences have been deposited in the NCBI Nucleotide database (accessed 2021-03-30).

Aims: We aimed to establish high-quality reference sequences for common *ABO* alleles observed in Switzerland. To resolve allele haplotypes, we took advantage of the latest 3rd-generation long-read NGS technologies of Oxford Nanopore Technologies (ONT) and Pacific Biosciences (PacBio).

Methods: We selected samples from a large, well-characterized *ABO* genotype dataset (n=25,200) of serologically-typed blood donors from the greater Zurich area (Switzerland), which had been generated previously using MALDI-TOF MS. We aimed to sequence at least 20 alleles for each of the six main *ABO* groups, i.e. *ABO*A1*, *A2*, *B*, *O.01.01*, *O.01.02*, and *O.02*. In total, we selected 78 samples for sequencing. The entire *ABO* gene comprising ~24.2 kb was amplified in two overlapping long-range PCRs (13 kb and 17 kb). PCR amplicons were sequenced using a PCR-free barcoding protocol on a MinION device from ONT. To circumvent potential biases from standard reference-based read mapping, we performed for each sample a *de-novo* (i.e. reference-free) assembly. For cross-validating the ONT sequences, a subset of 12 samples (n=2 for each *ABO* group) was sequenced using PacBio (HiFi sequencing on a Sequel II system) and Illumina (MiSeq).

Results: We established generic long-range PCR reactions covering the entire *ABO* gene, including the large intron 1, which has been difficult to amplify so far. Amplicons from the 78 study samples were sequenced with ONT to a median sequence depth of 1400X. For all samples, both full-length *ABO* haplotype sequences could be resolved. Cross-validation with gold standard HiFi PacBio and Illumina data confirmed high quality of the ONT haplotype sequences. Considerable genetic variation was observed among the six *ABO* groups (*ABO*A1*, *A2*, *B*, *O.01.01*, *O.01.02*, and *O.02*). Within-group genetic diversity was highest for *ABO*O.01.01*, while genetic diversity was particularly low within the groups *ABO*A1* and *ABO*B*. Strikingly, our data revealed four SNVs being putatively specific for *ABO*A1*. Such diagnostic SNVs are currently lacking.

Summary/Conclusions: We have established a large dataset of 156 fully-phased *ABO* sequences for the most common *ABO* alleles in Switzerland. This collection will serve as a valuable reference resource for NGS-based *ABO* genotyping and sequencing. Our data uncovered four putatively *ABO*A1*-specific SNVs, which are currently being studied in detail to verify diagnostic specificity.

