VS-17-7

FULLY-PHASED REFERENCE SEQUENCES FOR ABO BLOOD GROUP GENE ALLELES BY LONG-READ NANOPORE SEQUENCING: PUTATIVE ABO*A1-SPECIFIC SINGLE-NUCLEOTIDE VARIANTS REVEALED

M. P. Mattle-Greminger[1], G. A. Thun[1], M. Gueuning[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]

1 Blood Transfusion Service Zurich (SRC), Department of Research and Development, Schlieren, Schweiz
2 Christian-Albrechts-University of Kiel, Institute of Clinical Molecular Biology, Kiel, Deutschland
3 Stefan-Morsch-Foundation, Birkenfeld, Deutschland
4 Blood Transfusion Service Zurich (SRC), Department of Molecular Diagnostics and Cytometry, Schlieren, Schweiz
5 Blood Transfusion Service Zurich (SRC), Schlieren, Schweiz
6 Private University in the Principality of Liechtenstein, Institute for Translational Medicine, Triesen, Liechtenstein

Background: Molecular blood group genotyping and sequencing require allele reference sequences. For many blood group genes, however, com-plete allele sequences remain rare. The main obstacle lies in resolving haplotypes. We aimed to generate fully-phased reference sequences for all six major ABO allele groups: ABO*A1/A2/B/O1.01/O1.02/O2. To resolve allele haplotypes, we used the latest 3rd-generation long-read sequenc-ing technologies of Oxford Nanopore Technologies (ONT) and Pacific Biosciences (PacBio).

Methods: We selected 78 samples from a large, well-characterized ABOgenotype dataset (n=25,200) of serologically-typed blood donors from the greater Zurich area (Switzerland), which had been generated pre-viously using MALDI-TOF mass spectrometry. The entire ABO gene (~23.3 kb) was amplified in two overlapping long-range PCRs (13 kb and 17 kb) and amplicons were sequenced with ONT. For cross-validating ONT sequences, a subset of 12 samples (n=2 for each ABO group) was sequenced using gold standard long-read PacBio HiFi and short-read Illumina. ONT data was analyzed using a reference-based read mapping pipeline as well as a de-novo (i.e. reference-free) assembly pipeline to cir-cumvent potential biases from standard single-reference-based mapping.

Results: Median ONT sequencing depth was 4200x per sample. Cross-validation with PacBio and Illumina data confirmed high quality of ONT sequences. The de-novo assembly pipeline outperformed the single- reference-based read mapping. For all samples, both full-length ABOhaplotype sequences could be resolved. Most of the genetic diversity was observed between, not within ABO groups. Within-group diversity was highest for ABO*O.01.01 (π=0.00048) and lowest for ABO*B (π=0.00004). Phylogenetic tree and network analyses showed distinct clustering of each ABO allele group with a high proportion of fixed SNVs. Most strikingly, our data revealed four SNVs being putatively specific for ABO*A1.01 (ISBT reference allele). Such diagnostic SNVs are currently lacking.

Conclusion: We have generated a large dataset of 156 fully-phased sequences for all six major ABO allele groups (ABO*A1/A2/B/O1.01/O1.02/O2). They will serve as a valuable reference resource for ABOgenotyping and sequencing. Nanopore sequencing provided high-quality sequences and was powerful for resolving haplotypes. Our data uncov-ered four putatively ABO*A1-specific SNVs, which would finally allow for A1-specific genotyping. The SNVs are currently being studied in detail to verify diagnostic specificity