REFERENCE SEQUENCES FOR THE MOST FREQUENT ABO* ALLELES OBSERVED IN THE ZURICH REGION OF SWITZERLAND

Maja P. Mattle-Greminger[1§*], Anna-Lena Hansen[2§], Stefan Meyer[1], Sonja Sigurdardottir[1], Kathrin Neuenschwander[1], Nadine Trost[1], Yvonne Merki[1], Claudia Portmann[1], Elise Gourri[1], Jochen Gottschalk[3], Beat M. Frey[3], Wolfgang Peter[2], Christoph Gassner[1]

[1]Blood Transfusion Service Zurich, Department of Molecular Diagnostics & Research, SRC, Schlieren, Switzerland; [2]Stefan-Morsch-Stiftung, Birkenfeld, Germany; [3]Blood Transfusion Service Zurich, SRC, Schlieren, Switzerland

[§]These authors contributed equally to this work

Background: The availability of allele reference sequences is essential for molecular blood group genotyping and sequencing, particularly if based on next-generation sequencing (NGS). Such allele references sequences should meet the following criteria: (i) complete gene sequence, including introns, exons and flanking regions, (ii) fully phased (i.e. solid haplotype), (iii), confirmed phenotype, and (iv) deposited in a public sequence database (e.g. GenBank). For many blood group systems, however, the availability of such population-based allele reference sequences is still limited. One of the main technical challenges is solid haplotype information for the entire gene, which is difficult to obtain with both classical Sanger sequencing and short-read NGS. A prime example is ABO, considered the clinically most relevant blood group system. Only 11 complete human *ABO* sequences are currently deposited on GenBank (accessed March 8th 2017). Some other deposited sequences just lack the large intron 1, but the majority of sequences are restricted to the exons 6 and 7.

Aims: We aimed to generate high-quality, fully phased reference sequences for the most frequent ABO* alleles of the greater Zurich area of Switzerland, and therefore providing a reference tool supporting ABO blood group genotyping and sequencing.

Methods: To circumvent the haplotype phasing issue, we selected samples being putatively homozygous for the *ABO* gene locus for NGS. In brief, our methodological strategy encompassed three main steps. First, we performed data mining on a large ABO genotype dataset (n=25,200), which had been generated previously using matrix-assisted laser desorption/ionization, time-of-flight mass spectrometry (MALDI-TOF MS). All individuals were serologically ABO-typed blood donors from the Zurich region. Out of this dataset, we selected 378 individuals being homozygous for three causative SNPs designating A (ISBT wild-type, n=142), B (c.803G>C, n=73), O1 (c.261delG, n=142), and O2 (c.802G>A, n=21) for extended heterozygosity analysis. Second, to detect further potential allelic heterozygosities, we extended genotyping of these samples by testing 12 additional SNPs, including the SNP c.1061delC, which is discriminative for A1 and A2 alleles. Third, we selected 88 samples being homozygous at all genotyped SNPs and amplified the entire *ABO* gene locus with overlapping long-range PCRs. Amplicons were sequenced in paired-end mode on a MiSeq system (Illumina) using a validated pipeline.

Results: Of all 378 individuals with extended ABO SNP genotyping, 258 samples were homozygous at all 15 SNPs ("SNP-homozygous"), while 120 were heterozygous for at least one SNP. Notably, such heterozygous genotypes were exclusively detected among O1 and A alleles, with two different O1 ("O1-h1", "O1-h2") and two predominant A haplotypes (A1 and A2). Among B and O2 haplotypes no polymorphisms were detected. From all 258 SNP-homozygous individuals, we selected 88 for NGS (12 A1A1; 12 A2A2; 12 BB; 20 O1O1-h1; 20 O1O1-h2; 12 O2O2). We obtained sequences of the entire *ABO* gene (~24.6 kb), spanning from the enhancer region (~3.7 kb upstream of exon 1) to the end of exon 7.

Summary: We have generated a collection of high-quality *ABO* sequences for the most frequent Swiss ABO*A1, A2, B, O1, and O2 alleles with confirmed phenotype. These sequences will serve as a valuable reference resource for NGS-based ABO blood group genotyping and sequencing also in other Caucasian populations.