Clinical Chemistry 00:0
1–10 (2025)

Article

Comprehensive Annotation of Complete *ABO* Alleles and Resolution of *ABO* Variants by an Improved Full-Length *ABO* Haplotype Sequencing

Yanling Ying,[a] Jingjing Zhang,[a] Xiaozhen Hong,[a] Wenjing Yuan,[a] Kairong Ma,[a] Xinyu Huang,[a] Xianguo Xu,[a] and Faming Zhu[a,*]

BACKGROUND: Full-length *ABO* haplotype sequencing is crucial for accurate genotyping, reference gene annotation, and molecular mechanism analysis of its variants. However, there is currently a deficiency of comprehensive annotation for full-length *ABO* haplotypes, spanning from the 5′ untranslated region (UTR) to the 3′ UTR.

METHODS: Two sets of specimens (79 random blood donors and 47 *ABO* variants) were tested. The full-length *ABO* gene spanning the 5′ UTR to the 3′ UTR was amplified using an improved one-step ultra-long-range PCR with a pair of PCR suppression primers. A single-molecule real-time library was constructed, and *ABO* haplotype sequencing was performed. Data analysis including basecalling, aligning, variant calling, clustering, and variant annotation were performed.

RESULTS: The amplicon measured 26.1 kb without splicing, representing the most complete *ABO* gene reported to date. The complete *ABO* haplotype sequence was obtained via long-read sequencing. The comprehensive *ABO* reference alleles were obtained and the *ABO* sequence patterns within each allele in a Chinese population were further classified. The full-length *ABO* gene haplotype analysis technique effectively resolved *ABO* variants with structural variations (SVs), including large fragment deletions, inversions, recombination, and chimeras.

CONCLUSIONS: Full-length *ABO* haplotype sequencing filled a gap that was missing with respect to the 3′ UTR sequences of *ABO* alleles and can advance blood group genomic analysis, aiding in *ABO* gene function analysis, evolutionary studies, and the resolution of *ABO* variants.

Introduction

High-throughput sequencing has significantly advanced blood group genomics (1–4). While earlier studies primarily focused on analyzing exonic sequences of genes, the recent trend involves exploring introns and regulatory regions to gain a more comprehensive understanding of gene sequences (4–6). However, comprehensive full-length haplotype data for blood group genes is limited. Full annotation of blood group genes is important for improving the accuracy of blood group variant typing.

The *ABO* blood group system is of great clinical significance, and there are intensified research efforts in this area (7–10). *ABO* genetic events [such as the coding DNA sequence (CDS) variants, intronic variants, insertions, deletions, and recombination] may affect the activity of A or B glycosyltransferase, altering *ABO* phenotypes (11). However, apart from the CDS, data on intronic sequences, 5′ untranslated region (UTR), and 3′ UTR are limited, particularly for full-length *ABO* haplotypes. This limitation in sequence recognition hinders the understanding of glycosyltransferase coding, and inactivating variants of glycosyltransferases are of particular concern. Notably, the 5′ and 3′ UTRs may contain particularly significant regulatory elements affecting translational or posttranscriptional processes, resulting in numerous *ABO* variants (12–14). However, it is technically challenging to obtain the full-length haplotype sequence, including the UTRs associated with their respective alleles, due to the distance between the 5′ UTR and exonic single nucleotide variants (SNVs) in exons 6 and 7, and the variable number of short repeats in the 3′ UTR (12, 13).

Previous studies relied on PCR and Sanger sequencing of specimens that were homozygous for 5′ UTR sequences, which could not be definitively associated with different *ABO* haplotypes (13). The 3′ UTR sequence of the human *ABO* gene is even less well understood, and the little-known sequence information for the 3′ UTR may contain errors (12). Although Morgan et al. provided the reference sequences for 6 major *ABO* alleles, the study lacked data from the 5′ and 3′ UTRs (15). Consequently, there is currently no comprehensive annotation for the full-length *ABO* haplotype sequence

[a]Blood Center of Zhejiang Province, Blood Transfusion Medicine Research Institute, Hangzhou, Zhejiang, People's Republic of China.
*Address correspondence to this author at: Blood Center of Zhejiang Province, Key Laboratory of Blood Safety Research of Zhejiang Province, Jianye Road 789, Hangzhou, Zhejiang Province 310052, People's Republic of China. E-mail zfm00@hotmail.com.
Received May 29, 2024; accepted January 21, 2025.
https://doi.org/10.1093/clinchem/hvaf015

spanning from the 5′ to the 3′ UTR. Filling this gap is important both for blood transfusion medicine and resolving genetic variants (16, 17).

The third-generation single-molecule sequencing technology offers the potential to sequence the full *ABO* gene. Several studies on various blood group genes have utilized this advanced sequencing technique (1, 4–6). However, a challenge lies in achieving longer-length fragments of genes for library construction. Conventional long-range PCR (LR-PCR) struggles with DNA fragments >10 kb (18). A previous study had obtained the haplotype sequence of *ABO* using segment overlap amplification combined with subsequent biological information assembly, but this approach had not covered the UTR regions and the length of amplicons was slightly short (15). Various factors can influence the PCR amplification length (19–21). For instance, inverted terminal repeats with higher melting temperature (T_m) can cause the complementary ends of single-stranded DNA to anneal during PCR, forming duplex stems or panhandle-like structures. This blocks primer binding and leads to PCR suppression (PS), while amplification inhibition of short fragments favors amplification of long fragments (18, 22). In this study, ultra-long-range *ABO* fragments that were 26.1 kb in length were successfully amplified using a single pair of *ABO*-PS-specific primers in a conventional PCR system based on the effect of PCR suppression. Combined with long-read sequencing, this approach enabled a comprehensive annotation of the *ABO* gene and enhanced the capacity to detect *ABO* variants.

Materials and Methods

STUDY SPECIMENS

The study included 2 specimen sets: 79 randomly healthy blood donors from the Blood Center of Zhejiang Province and 47 *ABO* variants from the same center and the hospitals of Hangzhou, Zhejiang Province, China. Ethical approval for this study was granted by the ethics committee of the Blood Center of Zhejiang Province (Zhejiang Blood Center Ethical Review 2023 Study No. 022). All blood donors and patients provided informed consent.

SEROLOGIC TESTING

The *ABO* phenotypes of the donors were determined using an automated blood group analyzer (PK7300, Olympus). In cases with typing discrepancies, further identification was performed by tube tests with various antibody reagents and reverse-typed cells according to our previous report (23). All operational procedures and serologic diagnostic classification adhered to standard methods outlined in the Association for Advancement of Blood & Biotherapies technical manual (24).

PCR SEQUENCE-BASED TYPING FOR THE FULL CODING REGION OF THE *ABO* GENE

The full coding regions of *ABO*, including exons 1 to 7 were amplified with 3 pairs of primers and sequencing using PCR sequence-based typing (PCR-SBT) according to a previous report (9).

ONE-STEP SUPPRESSION ULR-PCR OF THE FULL-LENGTH *ABO* GENE

A one-step LR-PCR method was improved to amplify the entire *ABO* gene, spanning from the complete 5′ UTR (c.−5001) to the 3′ UTR (c.+1600) flanking region (approximately 26.1 kb) (Fig. 1). A pair of PS primers was designed with a 25-nt arbitrary GC-rich sequence added to the 5′ end of the specific primers. The ultra-long-range PCR (uLR-PCR) system was optimized and performed in 20 μL per reaction containing 1× buffer, 0.5 mM dNTPs, 0.15 μmol/L each primer, 0.5 U KOD FX Neo DNA polymerase, and about 50 ng DNA template. The sequences of PS primers and the PCR cycling details for the PS primer are available in the Supplemental Materials.

LIBRARY PREPARATION AND SMRT SEQUENCING

Single-molecule real-time (SMRT) libraries was constructed according to the manufacturer's guidelines. Briefly, a one-step assay combining DNA damage repair, end-repair, and adapter ligation was employed to generate the presequencing libraries with unique barcode adapters. After purification of the ligation products using 0.6× Ampure PB beads (Beckman Coulter), the uniquely barcoded prelibraries were pooled in equal masses. The final library was prepared using the Sequel Binding Kit 2.2 (Pacific Biosciences) and Internal Control Kit 1.0 (Pacific Biosciences), followed by sequencing on the Sequel II platform (Pacific Biosciences).

BIOINFORMATICS ANALYSIS

Primary analysis utilized SMRTLink software (v.10.1.0, Pacific Biosciences). The reads were initially demultiplexed, with barcode sequences automatically assigned at the end of the runs. Subsequently, subreads were then processed to generate circular consensus sequencing (CCS) reads using the CCS software (Pacific Biosciences). Filtered CCS reads were aligned to the *ABO* reference (NG_006669.2) using pbmm2 (https://github.com/PacificBiosciences/pbmm2, v.1.2.1) (25). SNVs and small indels were identified and called using deepvariant (https://github.com/google/deepvariant/blob/r1.2/docs/deepvariant-quick-start.md, v.1.2.0) (26). The threshold for variation site quality was set to QUAL >30. The large indels and allelic differences were analyzed and manually reviewed using the SnapGene software (v.6.2.2) (27). Alleles were named according to the

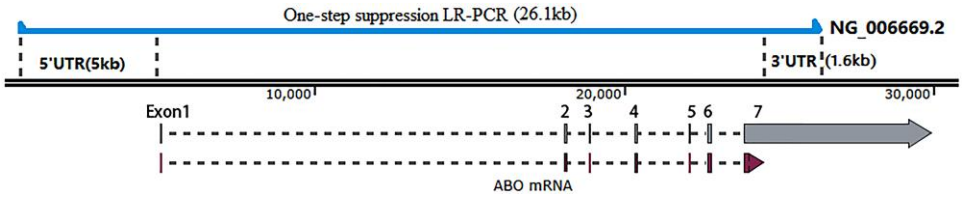

Fig. 1. Genetic structure of *ABO* and the amplification of full-length *ABO* through uLR-PS-PCR. A single pair of PS primers, designed to cover the 5′ to the 3′ UTR (26.1 kb in length), was employed for uLR-PS-PCR (highlighted in blue line). The black double line represents the *ABO* reference sequence NG_006669.2, and the number below the double line represents the base position of the sequence. Exons 1 to 7 are illustrated in gray and the CDS is in dark red. Color figure available at https://academic.oup.com/clinchem.

guidelines of the International Society of Blood Transfusion working group for allele nomenclature (https://www.isbtweb.org/isbt-working-parties/rcibgt.html).

Results

ONE-STEP AMPLIFICATION AND PACBIO SMRT SEQUENCING FOR THE FULL-LENGTH *ABO* GENE

The full-length *ABO* amplicon was approximately 26.1 kb in length without splicing, representing the longest *ABO* amplicon achieved to date (Fig. 1 and Supplemental Fig. 1). Mapping the filtered reads to the *ABO* reference sequence demonstrated complete coverage of the entire *ABO* gene (spanning 5′ to 3′ UTR). Quality control details for sequencing data and reads mapping are provided in Supplemental Table 1.

HAPLOTYPE SEQUENCE OF BLOOD DONORS

Based on the specific exonic variations, 158 *ABO* haplotypes from 79 blood donor specimens were attributed to the dominant International Society of Blood Transfusion alleles, including *ABO*A1.01*, *ABO*A1.02*, *ABO*B.01*, *ABO*O.01.01*, and *ABO*O.01.02*. Two low frequency alleles *ABO*A2.05* and *ABO*O.01.71* were detected among the blood donors. Detailed information on each specimen and corresponding haplotyping results are provided in Supplemental Table 2. The inferred phenotype of these samples from blood donors based on the sequences in the exon regions by this method were consistent with the phenotype by serology method.

DIFFERENT HAPLOIDS OF MAJOR ALLELES FORM A SPECIFIC SEQUENCE PATTERN

The 5 predominant *ABO* alleles in the Chinese population were further subdivided into various subtypes based on intron and regulatory region variations. Figure 2 illustrates the overall sequence pattern of these major alleles. The 2 *ABO*A1.02* subtypes differed mainly by 2 linked variants: c.374+163C>T and c.374+179T>C in intron 6. The *ABO*B.01* allele was similarly categorized based on the variant c.−2287A>G in the 5′ UTR. *ABO*O.01.01* exhibited 2 distinct patterns with additional subtypes distinguished by significant numerous SNVs across a 3.0 kb region from the 5′ UTR to intron 1. *ABO*O.01.02* was classified into 2 subtypes based on the variant c.29−75A>G. Detailed variations of these alleles, including variations in exons, introns, the 5′ UTR, and the 3′ UTR, are listed in Supplemental Table 3. Interestingly, 4 recombinant alleles were discovered in donor specimens. Their structural diagrams of recombination events are shown in Supplemental Fig. 3, and the detailed variations associated with these alleles are provided in Supplemental Table 4.

MAIN STRUCTURAL DIFFERENCES IN THE INTRON 1 SEQUENCE AMONG DIFFERENT ALLELIC HAPLOTYPES

Significant structural variations (SVs), particularly involving repetitions and insertion-deletion sequences, were observed in the proximal region of erythroid cell-specific transcription regulatory element within intron 1 (Fig. 3A). Notably, a (TA)n variable number of tandem repeats (VNTR) was observed in the proximal upstream region around +5.8 kb region, where "n" ranges from 11 to 26. Alleles *ABO*A1.02* and *ABO*B1.01* exhibit a higher number of TA repeats compared to *ABO*O* alleles. The median number of TA repeats for *A1.02* and *B1.01* alleles was 21, whereas for *ABO*O* alleles, the median number was 13 (Fig. 3B). Remarkably, the terminating sequence of TA repeats varied among alleles, with *ABO*A*, *B*, and *O.01.01* pattern 1 alleles ending with CACA, a feature absent in other alleles. Figure 3C illustrates the specific characteristics of the (TA)n VNTR sequences, along with insertion and deletion sequences. The detailed variations associated with these alleles are provided in Supplemental Table 3.

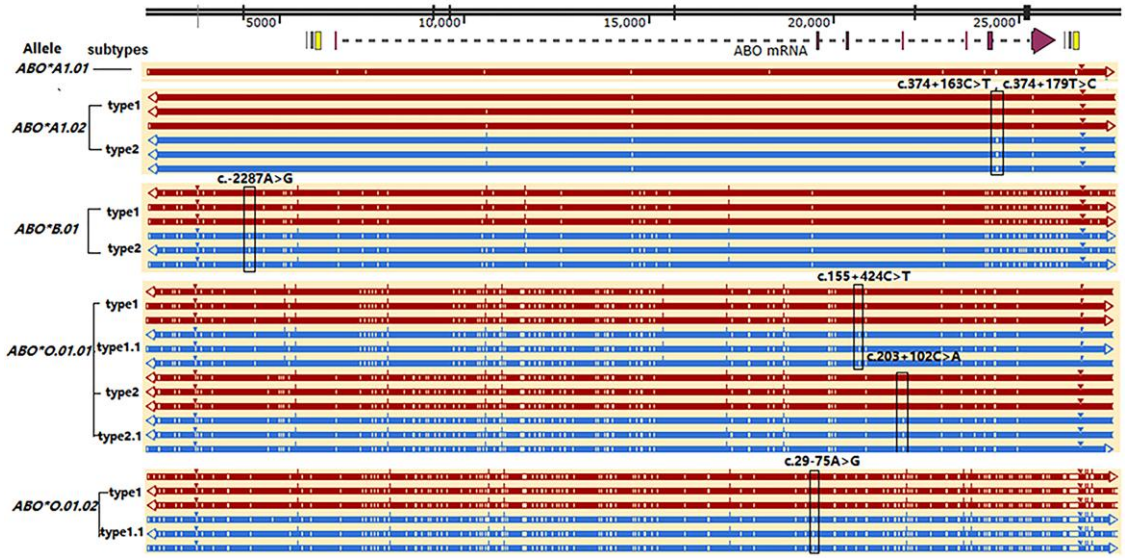

Fig. 2. Alignment of haplotype sequences highlighting distinct sequence patterns of the 5 main *ABO* alleles in the Chinese population. The *ABO* gene structure is depicted at the top of the graph for orientation purposes. There is only one haplotype sequence of *ABO*A1.01*. The remaining alleles are classified into different subtypes based on specific variants, and 3 haplotype sequences were randomly selected for each subtype. The top black double line represents the reference sequence (NG_006669.2), and the number below the double line represents the base position of the sequence. An arrow pointing to the right signifies that the sequence aligns with the *ABO* reference sequence NG_006669.2, while an arrow pointing to the left indicates that the sequence aligns with the complementary strand of the reference sequence. Small yellow vertical bars on the haplotype sequences highlight positions that differ from the *ABO* reference sequence, including SNVs and deletions. Different subtypes of the same allele are separated by dark red and blue, while small black boxes indicate specific distinguishing variants. Color figure available at https://academic.oup.com/clinchem.

SEQUENCE CHARACTERISTICS OF THE 5′ AND 3′ UTR IN DIFFERENT ALLELIC HAPLOTYPES

The characteristic motifs within the 5′ and 3′ UTR were comprehensively delineated across the main alleles. The variations, including various SNVs, 6- and 8-bp repetitions, CBF/NF-Y-binding minisatellite region (43-bp repeat unit), and a 36-bp insertion within different alleles are described in detail (Fig. 4 and Supplemental Table 3). Interestingly, the *ABO*A* allele featured a single 43-bp repeating motif, while other alleles had 4 repeats of this motif. Additionally, distinct NTT endings in motif sequences varied by allelic subtypes. *ABO*O.01.01*-type1 ends in GTT and *ABO*O.01.01*-type2 ends in CTT. Apart from the primary 5′ UTR pattern, some variants exhibited diverse forms of motif sequence characteristics, as illustrated in Fig. 4.

The 3′ UTR sequences (roughly 1.6 kb in length) of different alleles were successfully obtained. The 3′ UTR of *ABO* gene exhibited highly repetitive sequences, notably characterized by numerous (CA)n short repeats. Based on the research by Rie Sano et al., each motif and its subsequent sequence was considered as a distinctive alignment or unit (12). We further categorized the highly repetitive sequence units in the 3′ UTR across different alleles into units 1–14 (Fig. 5). These units are relatively conserved among subtypes of the same allele, with minor variations that may exist among different alleles. Notably, the current *ABO* reference sequence NG_006669.2 has the minimum number of duplicate units in the 3′ UTR, lacking units 11, 12, and 13 compared to the sequence derived from the Chinese population. Among these major alleles, *ABO*O.01.02* displayed the most significant sequence differences, missing units 1, 6, 8, 11, 12, and 13, and adding other unit changes. Three common intervening sequences were identified, characterized by 10 bp (GAAAGAC ACA), 8 bp (CACAGACA), and 7 bp (AATCACA) (Fig. 5).

VARIANTS TYPING SOLUTION

Coding Region Variation Analysis. The full-length *ABO* gene sequencing method developed in this study enabled

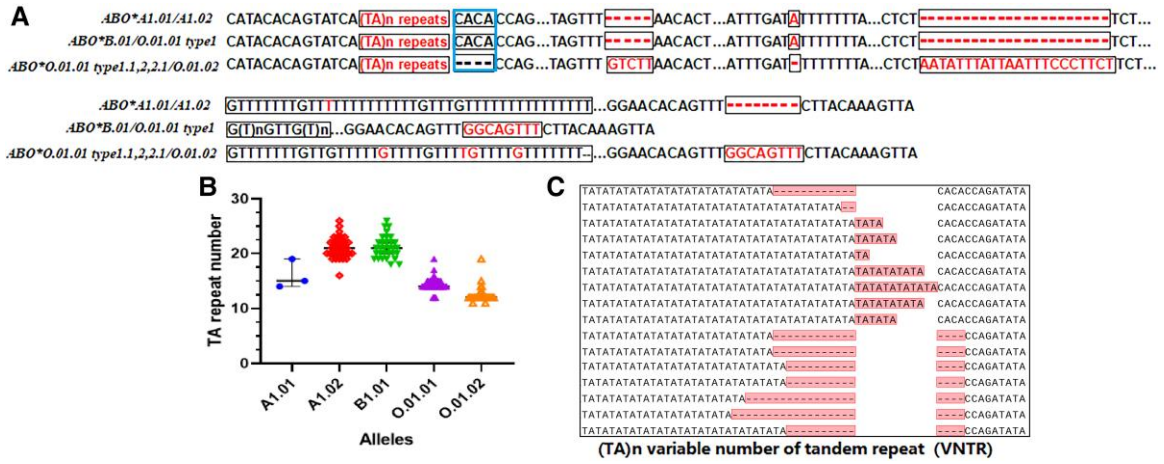

Fig. 3. The main characteristic structural differences of intron 1 sequence between different allelic haplotypes. (A), Major repetitions, insertions, deletions, and structural differences are highlighted in black frames. The red within the black box signifies sequence variations. The different endings of (TA)n are shown in blue boxes. *ABO*A*, *B*, and *O.01.01 type 1* alleles end with CACA, a feature absent in other allele types; (B), TA repeat number discrepancies between different alleles; (C), Sequence diagrams of TA repeat variations in some samples, with differences between alleles visible in the light red box. Color figure available at https://academic.oup.com/clinchem.

direct acquisition of the variant allele haplotype sequences. Among 47 *ABO* variants, 11 specimens revealed distinct CDS variants (Supplemental Table 5). Three novel variants were identified, including the typical *ABO*A1.02* associated with c.563G>A (specimen ID R23028), c.203G>T (specimen ID R23039), and the *ABO*B1.01* allele added c.255C>T (specimen ID R22056). Additionally, *ABO*B1.01−c.203G>T* is linked to c.−119C>T in the remote promoter. Comparing with the sequences in the exon regions, the results of this method were consistent with those of PCR-SBT. However, the PCR-SBT method did not provide complete allele sequence information, therefore other variations in the UTRs or intronic regions found in this method were not able to be compared with the PCR-SBT.

Intron and Promoter SNVs Analysis of Haploids. Thirty-two *ABO* variants were intronic and promoter-region SNVs, including 8 variants at intronic splice sites, 9 within +5.8 kb of the core region of intron 1, 2 within the promoter, and the remaining 13 within other introns (Supplemental Table 5). An interesting finding was the linkage of c.155+5G>A with the c.−2287A>G variant in the distal 5′ UTR, as revealed by full-length haplotype sequencing of *ABO*B3.03* alleles (Fig. 6A). In the +5.8 kb region of intron 1, c.28+5858T>G was identified in an *ABO*A2.05* allele, resulting in the Ael phenotype. Besides specimen R21112, a novel variant c.1−67C>T with *ABO*A1.02* was identified in the promoter. Other variants were scattered across introns, but their connection to antigen alterations remains unclear.

Identification of a Large Deletion in the Erythroid Cell-Specific Transcription Enhancer. A substantial 7337 bp deletion was identified in specimen R22124, associated with the Bel phenotype characterized by extremely weak B antigen expression and serology typing discrepancy. The genotype was determined as *ABO*Bel-new/ABO*O.01.01*. The breakpoints were located at c.28+501A and c.28+7837G, spanning the entire region of the erythroid cell-specific transcription enhancer (Fig. 6B).

Discovery of the Allelic Recombination Structure. Recombination events and hotspots were identified in 3 variants with the ABw phenotype. For instance, specimen R21070 had a recombination between the haplotypes of *ABO*O.01.01* type1.1 and *ABO*B.01*, with the hotspot occurring between variant c.204−9T to c.240−219G (Fig. 6C). The details of all recombination alleles are listed in Supplemental Table 5.

Detection of Chimeras with Multiple Allelic Haplotype Mixing. A case (specimen R18033) with a serological discrepancy due to chimerism was identified (Fig. 6D). The serology results indicated a mixed field (mf) of the A antigen on red blood cells, with complete

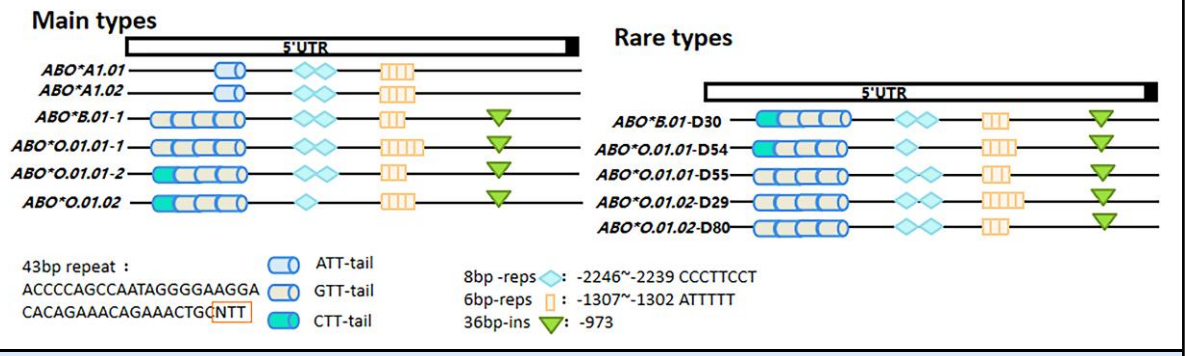

Fig. 4. The characteristic motifs of the haplotype sequences in the 5′ UTR, including the CBF/NF-Y-binding minisatellite region (43-bp repeat unit), 6-bp repetitions, 8-bp repetitions, and a 36-bp insertion. Each motif is represented by a specific shape graph, as indicated in the figure. The color of the cylinders represents different endings of the CBF/NF-Y-binding minisatellite sequence. The number of shapes represents the number of motif sequence repetitions. Color figure available at https://academic.oup.com/clinchem.

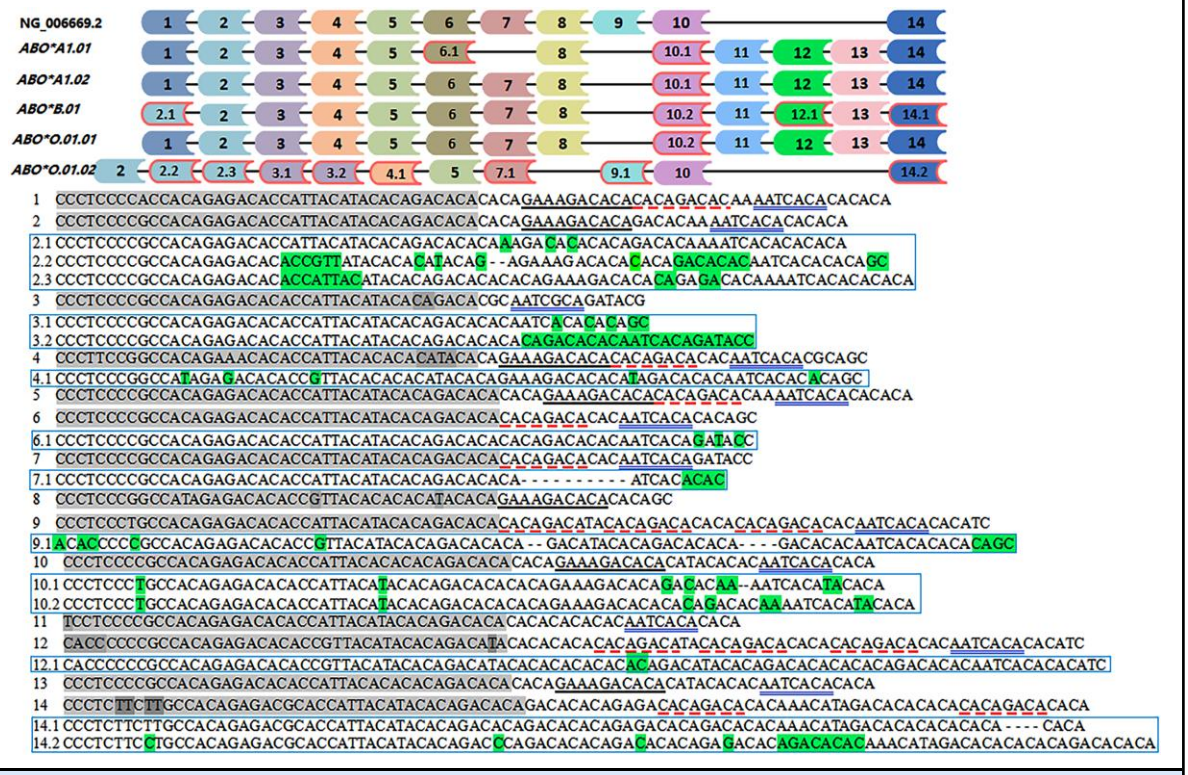

Fig. 5. The characteristic motifs of the 3′ UTR haplotype sequence of the human *ABO* gene. The haplotype sequences of the 3′ UTR of the human *ABO* gene were categorized into distinctive alignments or units, labeled as 1–14, displayed in fusiform shapes with different colors. Variability in unit composition among different alleles is noted, highlighted by a light red box indicating the variant units. Nucleotides distinct from the consensus sequence of the repetitive motifs are indicated on a green background. A repetitive motif consisting of a 40-bp sequence is marked with gray background. The intervening sequences are recognized by a 10-bp sequence (black underlined), an 8-bp sequence (red dotted line), or a 7-bp sequence (double blue underlined). The small dash represents a base deletion. Color figure available at https://academic.oup.com/clinchem.

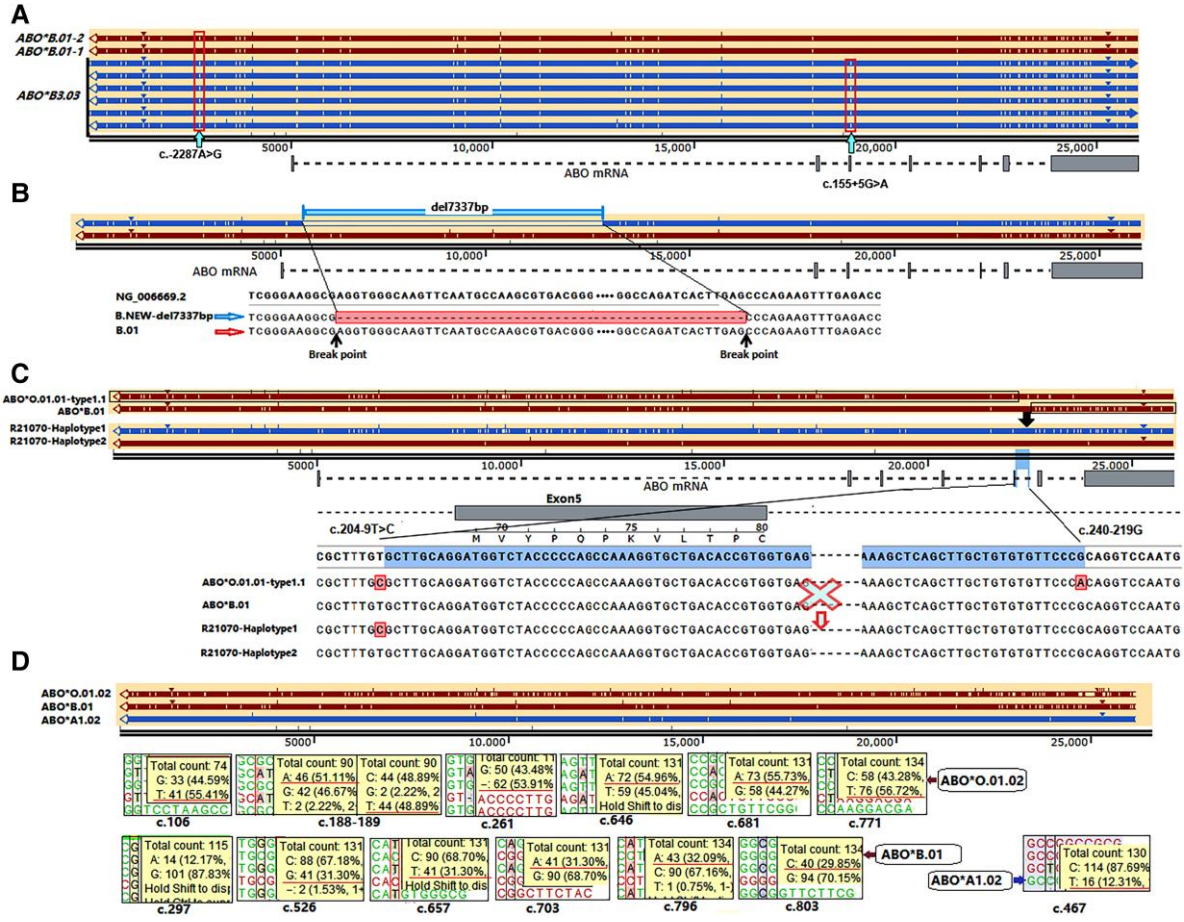

Fig. 6. Haplotype sequence analysis of variants. Small yellow vertical bars on the haplotype sequences highlight positions that are different from the *ABO* reference sequence (NG_006669.2). The black double line represents the reference sequence, and the numbers below the double line represent the base position of the sequence. (A) c.155+5G>A is linked with the c.−2287A>G variant located in the distal 5′ UTR based on the full-length haplotype sequence of the *ABO*B3.03* alleles. The red haplotype sequences represent 2 subtypes of the *B.01* allele, and the blue sequences are the *ABO*B3.03* alleles. The red boxes with arrows pointing at them indicate each variant location. (B) Identification of a large deletion in the erythroid cell-specific transcription enhancer. The light blue bar represents the deleted region, and light red bar highlights the deleted sequences. The deletion breakpoints are located at c.28+501A and c.28 +7837G. (C) Identification of an *ABO* recombination event of the haplotype. In sample R21070, the recombination event occurred near to exon 5. The top 2 red alleles are the normal *ABO*O.01.01-type1* and *ABO*B.01* alleles. The black box part of the sequences was recombined to form a recombinant allele R21070-haplotype1 (allele highlighted in blue). The sequence with a light blue background is the region where recombination occurred, and the small red box is the hotspot base of recombination. (D) Detection of chimeras with multiple haplotype sequences. The specificity variants in CDS for each allele are listed in the box and underlined in red. A black oval box indicates allele names, and color of the small arrow pointing to the CDS position corresponds to the color of the full-length haplotype sequences above. Color figure available at https://academic.oup.com/clinchem.

agglutination of a minority of cells (Supplemental Table 5). Three distinct allelic haplotypes, *ABO*B.01*, *ABO*O.01.02*, and *ABO*A1.02*, were isolated by clustering of full-length haplotype sequences with frequencies of approximately 30%, 40%, and 12%, respectively.

Discussion

In this study, an innovative uLR-PCR technique for capturing the full-length *ABO* gene was developed. The *ABO* haplotype sequence spanning from the 5′ to 3′ UTR without splicing were successfully analyzed. The long-read length not only resolved haplotype phasing ambiguity but also enabled accurate alignment within tandem repeats and homologous genes. This capability enabled the reliable detection of SVs, including large fragment deletions, inversions, recombination, and so on. However, this method also has certain limitations, notably the requirement for high-quality DNA to amplify ultra-long fragments, which can be challenging for aged specimens or degraded DNA. In additional, only variants identified via long-range sequencing in the exons and splice sites were able to be compared with PCR-SBT. The variants found in UTRs and most intronic sequences could not be verified by a secondary method.

While previous reports on *ABO* gene amplification may have certain limitations, our study provides a global overview of the *ABO* gene structure, significantly advancing the understanding of its genetic architecture. The complete sequences for the predominant *ABO* alleles in the Chinese population, including the common *ABO*A1.02* allele, expanded on the sequence patterns described by Gueuning et al. (15). Moreover, an in-depth analysis of sequence patterns based on the most extensive dataset of *ABO* sequences was conducted. Previously, our laboratory identified specific SNVs in the intron region of different *ABO* alleles using next-generation sequencing (NGS) technology (28); however, assigning SNVs to specific allele patterns proved challenging. Gueuning et al. elucidated the pattern differences in *ABO* alleles using segmented amplification of the *ABO* gene combined with Oxford Nanopore Technologies (ONT), but did not provide detailed annotations for specific intron SNVs (15). In this study, we further identified the specific distinguished SNVs and subdivided the sequence patterns within each allele. For instance, within the *ABO*A1.02* pattern, the c.374+163C>T and c.374 +179T>C intron 6 variants were detected in numerous individuals and consistently linked, suggesting a nonrandom occurrence or structured pattern. Similarly, fixed SNVs were regularly found within the *ABO*B.01*, *ABO*O.01.01*, and *AB0*O.01.02* alleles, thereby serving as distinguishing markers for their respective subgroups.

The SNV correlations observed in this study may enhance our understanding of *ABO* molecular diagnostics in transfusion and transplantation medicine (29–32). For instance, a link was found in the *ABO*B.03* allele sequence between c.155+5G>A and c.−2287A>G in the distal 5′ UTR. The c.−2287A>G variant served as a distinguishing SNV for the *ABO*B.01* sequence pattern subgroup, indicating the potential specificity of the *B.03* subgroup derived from the *ABO*B.01*-type2 pattern.

Numerous studies have highlighted the association of the *ABO* gene with human disease susceptibility, underscoring its pivotal role in transfusion medicine and transplantation (29–32). Previous research primarily focused on exon sequences and often analyzed the correlation between single SNP loci and diseases (32). The complete sequencing of the *ABO* gene and the construction of its sequence patterns open the way for new insights into the *ABO* blood group's association with diseases. It is also noteworthy that identifying specific distinguishing SNVs could refine the nomenclature of *ABO* alleles, drawing inspiration from the multilevel naming system of HLA alleles for more informative naming (33).

Additionally, obtaining the sequence of the 3′ UTR region filled a gap in the downstream regulatory sequence pattern of the human *ABO* gene. The challenge of acquiring a complete 3′ UTR sequence has left a void in extensive human *ABO* gene data, which was primarily derived from cell lines and a limited number of homozygote individuals. Our analysis revealed the relative conservation of the 3′ UTR sequence within the same allele despite its high repetitiveness. Adjusting the unit structure classification and common intervening sequences reported by Rie Sano et al., we generated a 3′ UTR structure map corresponding to different alleles in the Chinese population. Previous investigations have indicated that the suppressor region downstream of the *ABO* gene may negatively affect promoter transcription. The stability of *ABO* transcripts may depend on a concise stem ring structure (34). Therefore, a comprehensive investigation of *ABO* 3′ UTR terminus properties is essential for understanding mRNA stability regulation. The information collected in this study serves as a vital resource for exploring the role of the 3′ UTR in *ABO* transcription.

An interesting VNTR composed of (TA)n in intron 1 exhibited noteworthy differences in both the number of repeats and their ending sequence among alleles. A distinct pattern of TA repeats was observed, distinguishing the Chinese population's sequences from those reported by Gueuning et al. (15), notably lacking the CACA ending. TA VNTR, a common feature in the human genome, is associated with genomic rearrangements in evolution and is enriched near breakpoints involved in cancer translocation and deletion (35).

Significantly, haplotype sequencing with complete genes effectively resolved the *ABO* variants, particularly in identifying novel alleles arising from distant variants outside exons 6 and 7, as well as detecting complex SVs. This technique offers significant advantages over PCR-SBT and NGS (28, 36). Moreover, it resolved challenging cases that existing technologies could not. In one case, a 7337 bp deletion in the erythroid cell-specific transcription enhancer was identified, marking it the largest reported fragment deletion to date (37). Conventional segmented PCR amplification may fail to detect large fragment

deletions in heterozygotes. In cases of phenotype–genotype discrepancy, a microchimera was detected. Despite being defined as an AwB phenotype, only an *ABO*B.01/ABO*O.01.02* heterozygote was initially detected via PCR-SBT. Notably, approximately 12% of the *ABO*A1.02* allele was eventually identified, confirming a small population of A cells chimerism with B cells. Distinguishing chimeras causing *ABO* discrepancy from A3 or B3 variants is challenging due to mixed-field agglutinations (38), but our method effectively resolved this issue.

The method also proved effective in exploring the recombinant hot spots in specimens carrying recombinant alleles. This versatility extends beyond *ABO*, proving valuable for detecting recombination events in the Rh and MNS blood group system (39, 40). Obtaining all alleles for *ABO* ambiguities holds significant implications, especially for monitoring the prognosis of patients undergoing *ABO*-incompatible hematopoietic stem cell transplantations and for optimizing transfusion strategies in clinical settings (29–32).

In conclusion, we have developed an effective haplotype sequencing method for the full-length *ABO* gene using a single pair of *ABO* PS primers with long-read sequencing. The approach has enhanced our understanding of the *ABO* gene sequence and effectively addresses *ABO* variant challenges, benefiting clinical transfusions and transplants. As we navigate the present era and foreseeable future, the integration of "blood transfusion omics" and "multi-omics" is imminent, with genomics emerging as the cornerstone. Our research will promote the study on the regulatory mechanisms governing blood group systems within the field of transfusion medicine.

Data Availability

DNA sequencing data have been submitted to public repository SRA. The project accession number: PRJNA1164930.

Supplemental Material

Supplemental material is available at *Clinical Chemistry* online.

Nonstandard Abbreviations: UTR, untranslated region; SVs, structural variations; CDS, coding sequences; SNVs, single nucleotide variants; LR-PCR, long-range PCR; PS, PCR suppression; PCR-SBT, PCR sequence-based typing; uLR-PCR, ultra-long-range PCR; SMRT, single-molecule real-time; CCS, circular consensus sequencing; VNTR, variable number of tandem repeats; NGS, next-generation sequencing.

Human Genes: *ABO*, alpha 1-3-N-acetylgalactosaminyltransferase and alpha 1-3-galactosyltransferase; *RHD*, Rh blood group D antigen; *MNS*, MNS antigen system gene.

Author Contributions: *The corresponding author takes full responsibility that all authors on this publication have met the following required criteria of eligibility for authorship: (a) significant contributions to the conception and design, acquisition of data, or analysis and interpretation of data; (b) drafting or revising the article for intellectual content; (c) final approval of the published article; and (d) agreement to be accountable for all aspects of the article thus ensuring that questions related to the accuracy or integrity of any part of the article are appropriately investigated and resolved. Nobody who qualifies for authorship has been omitted from the list.*

Yanling Ying (Data curation-Lead, Funding acquisition-Lead, Investigation-Lead, Methodology-Lead), Jingjing Zhang (Funding acquisition-Equal, Investigation-Equal), Xiaozhen Hong (Investigation-Equal), Wenjing Yuan (Data curation-Equal, Investigation-Equal), Kairong Ma (Data curation-Equal, Investigation-Equal), Xinyu Huang (Investigation-Equal), Xianguo Xu (Investigation-Equal, Methodology-Equal, Project administration-Equal), and Faming Zhu (Funding acquisition-Equal, Methodology-Equal, Project administration-Lead)

Authors' Disclosures or Potential Conflicts of Interest: *Upon manuscript submission, all authors completed the author disclosure form.*

Research Funding: This research was supported by Zhejiang Provincial Natural Science Foundation of China under Grant (LTGY23H080003 [to Y. Ying], LTGY24H080005 [to J. Zhang]), National Natural Foundation of China (82070195, to F. Zhu), and the Science Research Foundation of Zhejiang Healthy Bureau (2022KY140 [to Y. Ying], 2023KY663, 2024KY937 [to J. Zhang]). This work was sponsored by the Zhejiang Provincial Program for the Cultivation of High-Level Innovative Health Talents (to Y. Ying).

Disclosures: None declared.

Role of Sponsor: The funding organizations played no role in the design of study, choice of enrolled patients, review and interpretation of data, preparation of manuscript, or final approval of manuscript.

Acknowledgments: We thank the blood donors and patients in this study, and assistance in data analysis provided by Engineer Xinyu Xu from Baiyi technology, China.

References

1. Jadhao S, Davison C, Roulis EV, Lee S, Lacaze P, Riaz M, et al. Using whole-genome sequencing to characterize clinically significant blood groups among healthy older Australians. Blood Adv 2022;6:4593–604.
2. Zhang Z, An HH, Vege S, Hu T, Zhang S, Mosbruger T, et al. Accurate long-read sequencing allows assembly of the duplicated RHD and RHCE genes harboring variants relevant to blood transfusion. Am J Hum Genet 2022;109:180–91.
3. Fichou Y, Audrézet MP, Guéguen P, Le Maréchal C, Férec C. Next-generation sequencing is a credible strategy for blood group genotyping. Br J Haematol 2014;167:554–62.
4. Fürst D, Tsamadou C, Neuchel C, Schrezenmeier H, Mytilineos J, Weinstock C. Next-generation sequencing technologies in blood group typing. Transfus Med Hemother 2020;47:4–13.
5. Gueuning M, Thun GA, Trost N, Schneider L, Sigurdardottir S, Engström C, et al. Resolving genotype-phenotype discrepancies of the

Kidd blood group system using long-read Nanopore sequencing. Biomedicines 2024; 12:225.
6. Fichou Y, Berlivet I, Richard G, Tournamille C, Castilho L, Férec C. Defining blood group gene reference alleles by long-read sequencing: proof of concept in the ACKR1 gene encoding the Duffy antigens. Transfus Med Hemother 2020;47:23–32.
7. Yamamoto F. Molecular genetics of the *ABO* histo-blood group system. Vox Sang 1995;69:1–7.
8. Abegaz SB. Human *ABO* blood groups and their associations with different diseases. Biomed Res Int 2021;2021: 6629060.
9. Ying YL, Hong XZ, Xu XG, Chen S, He J, Zhu FM, Xie XY. Molecular basis of *ABO* variants including identification of 16 novel *ABO* subgroup alleles in Chinese Han population. Transfus Med Hemother 2020;47:160–6.
10. Fennell K, Keller MA, Villa MA, Paccapelo C, Kucerakova M, Rosochova J, et al. New *ABO* intron 1 variant alleles. Immunohematology 2021;37:178–84.
11. Yamamoto F. A historical overview of advances in molecular genetic/genomic studies of the *ABO* blood group system. Glycoconj J 2022;39:207–18.
12. Sano R, Nakajima T, Takahashi K, Kubo R, Yazawa S, Kominato Y. The 3′ flanking region of the human *ABO* histo-blood group gene is involved in negative regulation of gene expression. Leg Med (Tokyo) 2011; 13:22–9.
13. Yan L, Zhu F, Liu Y, Xu X, Hong X. Sequences variations in 5′-flanking region of *ABO* gene and correlation with *ABO* alleles in the indigenous Chinese. Vox Sang 2008;94:227–33.
14. Seltsam A, Wagner FF, Grüger D, Gupta CD, Bade-Doeding C, Blasczyk R. Weak blood group B phenotypes may be caused by variations in the CCAAT-binding factor/NF-Y enhancer region of the *ABO* gene. Transfusion 2007;47:2330–5.
15. Gueuning M, Thun GA, Wittig M, Galati AL, Meyer S, Trost N, et al. Haplotype sequence collection of *ABO* blood group alleles by long-read sequencing reveals putative A1-diagnostic variants. Blood Adv 2023;7:878–92.
16. Gassner C, Olsson ML, Lane WJ, Hyland CA. Novel or not? Reference alleles, genes, and genomes to unmask the true nature of the *ABO**AW.10 allele associated with weak A phenotype. Transfusion 2022;62:721–4.
17. Wheeler MM, Johnsen JM. The role of genomics in transfusion medicine. Curr Opin Hematol 2018;25:509–15.
18. Zhao Z, Xie X, Liu W, Huang J, Tan J, Yu H, et al. STI PCR: an efficient method for amplification and de novo synthesis of long DNA sequences. Mol Plant 2022;15: 620–9.
19. Cheng S, Chang SY, Gravitt P, Respess R. Long PCR. Nature 1994;369:684–5.
20. Cheng S, Fockler C, Barnes WM, Higuchi R. Effective amplification of long targets from cloned inserts and human genomic DNA. Proc Natl Acad Sci U S A 1994;91:5695–9.
21. Chua EW, Maggo S, Kennedy MA. Long fragment polymerase chain reaction. Methods Mol Biol 2017;1620:65–74.
22. Dai ZM, Zhu XJ, Chen Q, Yang WJ. PCR-suppression effect: kinetic analysis and application to representative or long-molecule biased PCR-based amplification of complex samples. J Biotechnol 2007; 128:435–43.
23. Ying Y, Hong X, Xu X, Ma K, Chen S, He J, et al. Identification of a novel B allele with a nucleotide deletion (c.3_4 del G) in the *ABO* gene associated with a Bx phenotype individual. Transfusion 2019; 59:793–4.
24. Arthur CM, Olsson ML, Stowell SR. *ABO* and other carbohydrate blood group systems. In: Cohn CS, Delaney M, Johnson S, editors. Technical manual. 21st Ed. Bethesda (MD): Association for the Advancement of Blood & Biotherapies; 2023. p. 305–37.
25. Krajaejun T, Patumcharoenpol P, Rujirawat T, Kittichotirat W, Tangphatsornruang S, Lohnoo T, Yingyong W. PacBio long read-assembled draft genome of Pythium insidiosum strain Pi-S isolated from a Thai patient with pythiosis. BMC Res Notes 2023; 16:271.
26. Huang PJ, Chang JH, Lin HH, Li YX, Lee CC, Su CT, et al. DeepVariant-on-spark: small-scale genome analysis using a cloud-based computing framework. Comput Math Methods Med 2020;2020:7231205.
27. George SS, Pimkin M, Paralkar VR. Construction and validation of customized genomes for human and mouse ribosomal DNA mapping. J Biol Chem 2023;299: 104766.
28. He Y, Hong X, Zhang J, He J, Zhu F, Huang H. Analysis of the genomic sequence of *ABO* allele using next-generation sequencing method. Front Immunol 2022;13: 814263.
29. Krawczyk M, Mihalache F, Höblinger A, Acalovschi M, Lammert F, Zimmer V. Pancreatic cancer risk variant *ABO* rs505922 in patients with cholangiocarcinoma. World J Gastroenterol 2011;17:4640–2.
30. Watchko JF. *ABO* hemolytic disease of the newborn: a need for clarity and consistency in diagnosis. J Perinatol 2023;43: 242–7.
31. Ataca Atilla P, Akkus E, Atilla E, Gokmen N, Civriz Bozdag S, Kocak Toprak S, et al. Effects of *ABO* incompatibility in allogeneic hematopoietic stem cell transplantation. Transfus Clin Biol 2020;27: 115–21.
32. Zhou Y, Wang Y, Ni H, Deng W, Liu D, Xu J, et al. The potential significance of *ABO* genotyping for donor selection in kidney transplantation. Front Immunol 2020;11: 608716.
33. Hurley CK. Naming HLA diversity: a review of HLA nomenclature. Hum Immunol 2021; 82:457–65.
34. Griesemer D, Xue JR, Reilly SK, Ulirsch JC, Kukreja K, Davis JR, et al. Genome-wide functional screen of 3′UTR variants uncovers causal variants for human disease and evolution. Cell 2021;184:5247–60.e19.
35. Kaushal S, Wollmuth CE, Das K, Hile SE, Regan SB, Barnes RP, et al. Sequence and nuclease requirements for breakage and healing of a structure-forming (AT)n sequence within fragile site FRA16D. Cell Rep 2019;27:1151–64.e5.
36. Jiao LX, Zhang JY, Chen L, Xu Y, Yu XL, Han Y, et al. Gene identification of rare B(A) blood group. Transfus Apher Sci 2017;56: 855–7.
37. Ogasawara K, Miyazaki T, Ito S, Yabe R, Uchikawa M, Enomoto T, et al. The B allele with a 5.8 kb deletion in intron 1 of the *ABO* gene is the major allele in Japanese individuals with B_m and A_1 B_m phenotypes. Vox Sang 2018;113:393–6.
38. Wu PC, Lin YH, Tsai LF, Chen MH, Chen PL, Pai SC. *ABO* genotyping with next-generation sequencing to resolve heterogeneity in donors with serology discrepancies. Transfusion 2018;58:2232–42.
39. Flegel WA. Molecular genetics of RH and its clinical application. Transfus Clin Biol 2006;13:4–12.
40. Hassan SN, Thirumulu Ponnuraj K, Mohamad S, Hassan R, Wan Ab Rahman WS. Molecular detection of glycophorins A and B variant phenotypes and their clinical relevance. Transfus Med Rev 2019;33: 118–24.