Received: 12 March 2024 | Revised: 16 July 2024 | Accepted: 17 July 2024

DOI: 10.1111/vox.13721

ORIGINAL ARTICLE

Vox Sanguinis ISBT International Society of Blood Transfusion

Detecting serologically difficult ABO blood groups using single-molecule real-time sequencing technology

Zhe Wang | Yushuang Chu | Yanlin Xiao | Maohong Bian

Department of Blood Transfusion, The First Affiliated Hospital of Anhui Medical University, Hefei, Anhui, China

Correspondence
Maohong Bian, The First Affiliated Hospital of Anhui Medical University, 218 Jixi Road, Hefei, Anhui, China.
Email: mhbian@126.com

Funding information
Wanjiang Emerging Industries Technology Development Center Industrialization Project, Grant/Award Number: WJ21CYHXM11

Abstract

Background and Objectives: Recently, third-generation long-read sequencing technology has been increasingly applied to the detection of various blood group systems. Because of its long read length and use of single-molecule sequencing, it is capable of obtaining the sequences of blood group genes in their entirety as well as of distinguishing haplotypes. Therefore, here, we collected ABO blood group samples that were difficult to classify serologically and analysed the sequences of the coding regions of the ABO genes as well as the sequences upstream and downstream of the coding regions.

Materials and Methods: Samples with ABO antigen typing and reverse serum typing discrepancies were screened in a total of 21 patients. All samples were subjected to serological testing and preliminary ABO genotyping (polymerase chain reaction with sequence-specific primers [PCR-SSP]), followed by single-molecule real-time (SMRT) sequencing to obtain complete ABO gene sequences. PCR sequence-based typing (PCR-SBT) was performed to validate the results.

Results: Of the 21 samples, 15 had common ABO types, and 6 had rare ABO subtypes. One new allele, *ABO*B.NEW (c.861C>T)*, and one allelic base recombination event was identified. Forty-two haplotype sequences were obtained via SMRT sequencing with intronic single-nucleotide variants (SNVs) specific to the ABO allele, and all of the exon region sequences were consistent with the PCR-SBT results.

Conclusion: SMRT sequencing is capable of accurately obtaining complete ABO gene sequences, distinguishing haplotypes and identifying allelic recombination.

Keywords

ABO subtypes, haplotypes, single-molecule real-time sequencing, single-nucleotide variants

Highlights

- Third-generation long-read sequencing can obtain complete ABO gene sequences and distinguish haplotypes. It can be a good aid in serologically difficult ABO blood grouping.
- A novel allele *ABO*B.NEW (c.861C>T)* was discovered and has been uploaded to GenBank (OR565861).
- A total of 114 specific intronic region single-nucleotide variants for different ABO alleles were found using single-molecule real-time sequencing.

INTRODUCTION

Accurate blood group identification is a prerequisite for ensuring the safety of clinical blood transfusion, and the ABO blood group system, the first blood group system discovered in humans [1], is an important part of clinical blood grouping. With the development of science and technology, the means of blood group testing has expanded from traditional serological methods to molecular biology methods, and the study of blood group genes using polymerase chain reaction with sequence-specific primers (PCR-SSP), Sanger sequencing and next-generation sequencing (NGS) has become more widespread [2–13].

The ABO gene consists of seven exons and six introns and is located on human chromosome 9, 9q34.1-9q34.2, with a total length of approximately 25 kilobases (kb). Exons 6 and 7 encode most of the catalytically active glycosyltransferase region. This is the major part of the ABO gene product that determines the glycosyltransferase function and thus the production of ABO antigens [14]. Only exons 1–7, particularly exons 6 and 7, are tested in most ABO blood group gene sequencing. The complete sequence of the ABO coding region as well as the regulatory regions upstream and downstream of the coding region have been less well studied.

Single-molecule real-time (SMRT) sequencing, a third-generation sequencing technology, yields an average read length of up to 100 kb and highly accurate sequencing results, overcoming challenges in distinguishing haplotypes and multiple recombination events because of its sequencing of single molecules [15–17]. Fichou et al. were the first to apply third-generation sequencing technology to the detection of blood group genes, using this technology to obtain the full-length haplotype reference sequence of the ACKR1 gene (Duffy blood group system) [18]. Zhang et al. used the long-read length of third-generation single-molecule sequencing to assemble the full-length sequence of RHD-RHCE for the first time and study it in detail [19]. Subsequently, Tounsi et al. developed a convenient and rapid technique for full-length sequence detection of RHD genes using third-generation single-molecule sequencing [20]. Ji et al. also elucidated the transcriptional mechanism of the RHD DEL phenotype using third-generation sequencing [21]. Moreover, Gueuning et al. obtained the first full-length standard reference sequences of the six major isoforms of the ABO gene using third-generation sequencing and performed an evolutionary analysis of the gene [22].

The above-mentioned studies demonstrate the successful full-length sequencing of several common subtypes of blood group genes using third-generation sequencing technology, but the acquisition and analysis of complete sequences for rare ABO subtypes remains scarce. Therefore, in this study, ABO blood samples that were difficult to identify via serology were screened for rarer ABO subtypes. The target ABO gene was sequenced using SMRT technology, which not only yielded the complete sequence of the ABO gene from the start codon to the stop codon, including the sequences upstream and downstream of the coding region, but also distinguished the maternal and paternal haplotypes of the ABO gene.

MATERIALS AND METHODS

Sample collection

A total of 21 blood samples with incompatible ABO antigen typing and reverse serum typing were obtained from unpaid blood donors at the Blood Centre of Hefei City, Anhui Province, China, and from patients requiring blood transfusion for clinical purposes at the First Affiliated Hospital of Anhui Medical University, Anhui Province, China. Blood was collected intravenously in 3-mL EDTA-k^2 anticoagulation tubes and stored at 4°C. Blood serology was performed after sampling, and the samples were stored at −80°C for DNA extraction within 72 h. This study was approved by the Scientific Ethical Committee of The First Affiliated Hospital of Anhui Medical University. Informed consent was obtained from all participants.

Serological analysis

ABO antigen typing and reverse serum typing was performed on an ORTHO VISION MAX automated blood typing analyser (Ortho-Clinical Diagnostics) using the Ortho BioVue® System ABO-Rh/Reverse Grouping Cassette (Ortho-Clinical Diagnostics). If weak antigens or antibodies are detected, they must be further confirmed by the test tube method and the absorption/elution test. The relevant procedures were performed according to the National Clinical Laboratory Procedures Standards [23].

Genomic DNA extraction

The target genomic DNA was extracted from whole blood using a DNA extraction kit (Tianjin Super Biotechnology Developing Co., Ltd.) in strict accordance with the kit instructions. The concentration of the extracted DNA was determined using a Qubit 4 fluorometer (ThermoFisher Scientific), and the DNA concentration suitable for sequencing analysis was 30–50 ng/μL, with A260/A280 values of 1.6–2.0.

ABO genotyping using the PCR-SSP method

Target genomic DNA amplification and ABO blood group genotyping were performed using the Human Erythrocyte ABO Blood Group Genotyping Kit (PCR-SSP) (Tianjin Super Biotechnology Developing Co., Ltd.). The handling was carried out strictly in accordance with the kit instructions.

Sequencing the exon region of the ABO gene with the polymerase chain reaction sequence-based typing method

The PCR sequence-based typing (PCR-SBT) method belongs to the Sanger sequencing technology. Exons 1–7 of the ABO gene were

amplified and sequenced in both directions using the BigDye Terminator Cycle V3.1 sequencing kit (Application Biosystems, Foster City, CA, USA). The primer sequences are referenced to the Reference [24]. Sequence data were analysed using the SEQscape2.1 software (Application Biosystems, Foster City, CA, USA). The ABO alleles were determined from the nucleotide sequences of the polymorphic loci according to the International Society of Blood Transfusion (ISBT) standards for erythrocyte immunogenetics and blood group terminology [25] (Reference sequence:GenBank ID: NG_006669.2).

SMRT sequencing and bioinformatics analysis of the ABO gene

Three primer pairs were designed based on the ABO gene sequence (GenBank ID: NG_006669.2) [26]. The sequence from the start codon to the stop codon was covered by the overlapping amplicons of the primer pairs (9, 9.5 and 11.5 kb) (Figure 1, Table 1). Amplification was performed using the KOD FX enzyme (TOYOBO). The PCR amplification system and parameters are shown in Tables 2 and 3.

The construction of SMRT libraries was performed via a one-step method. DNA damage repair, end repair and adapter ligation were combined in one step to generate pre-sequencing libraries containing unique barcode adapters. The reaction mixtures were prepared before use and were 10 μL in total, containing 4 μL of PCR product, 5 μmol/L barcoded adaptor (Integrated DNA Technologies), 1× T4 DNA ligase buffer (Enzymatics), 1 mmol/L ATP (New England Biolabs), 200 μmol/L dNTP (New England Biolabs), 2.5 units of T4 polynucleotide kinase (Enzymatics), 0.75 units of T4 DNA polymerase (Enzymatics) and 180 units of T4 DNA ligase (HC) (Enzymatics).

A total of 120–250 ng of purified PCR product was mixed with the enzyme mixture. The reaction mixture was then incubated at 37°C for 20 min, 25°C for 15 min and 65°C for 10 min. Afterwards, exonuclease I (Enzymatics) and exonuclease III (Enzymatics) were added to remove the failed ligation products, and the final prelibrary was purified with 0.6× AMPure PB beads (Enzymatics). For multiple sample sequencing, prelibraries were pooled together according to equal masses. After pooling, the prelibraries were purified two times with 0.45× AMPure PB beads.

The final library was subsequently subjected to sequencing with the Sequel Binding Kit 2.2 (Pacific Biosciences) and the Internal Control Kit 1.0 (Pacific Biosciences). A total of 150 pM DNA–polymerase complexes were finally loaded and sequenced with the Sequel II platform (Pacific Biosciences) in a 20-h movie.

Primary analysis of the output data was carried out with the SMRTLink v10.1.0 software (Pacific Biosciences). The raw reads were first demultiplexed, and barcode sequence analysis was carried out automatically at the end of the runs, followed by subread analysis to generate circular consensus sequencing (CCS) reads using the CCS software application (filtration criterion: quality value>19). Filtered CCS reads were aligned to the human reference genome (GRCh38) with pbmm2 to specifically batch the desired blood-related gene CCS fragments. The target CCS reads were realigned to the reference genome (ISBT) using pbmm2. For the identification of single-nucleotide variants (SNVs) and small indels, variant calling was carried out with DeepVariant v1.2.0 (https://github.com/google/deepvariant/blob/r1.2/docs/deepvariant-quick-start.md). All alleles were named according to the standards of the ISBT working group,

TABLE 2 Polymerase chain reaction (PCR) amplification system.

	Single sample (μL)
KOD neo FX buffer	10.5
dNTPs	4
Total primer	9
gDNA	1
KOD Neo FX	0.5

TABLE 3 Polymerase chain reaction (PCR) parameters.

Temperature (°C)	Time	Cycle number (×)
94	2 min	1
98	10 s	25
68	12 min	
68	10 min	1
8	∞	

FIGURE 1 Three primer pairs were designed covering the 5′-UTR to 3′-UTR of the ABO gene, and three overlapping amplicons of 9, 9.5 and 11.5 kb were amplified.

TABLE 1 Primer design for ABO gene single-molecule real-time (SMRT) sequencing.

Sequence primers	Forward primer	Reverse primer
Primer 1	5′catccctttcaccttggcattt3′	5′agctacattgaccagagagaga3′
Primer 2	5′gcccaccaaaactccctggaa3′	5′ccagttcctgccagggagagga3′
Primer 3	5′gtgtgaaactcatcaaaacc3′	5′cgcagggattgcagtgagg3′

which develops and maintains guidelines for blood group antigens and allele nomenclature [25].

RESULTS

ABO serology, PCR-SSP, PCR-SBT and SMRT sequencing results from 21 samples

Fifteen common ABO phenotypes were identified (nine group A, three group B and three group AB). Six rare ABO subtypes were identified with the following genotypes: *A1.02/B.NEW(c.28G>A)*, *cisAB.05/B.NEW(c.861C>T)*, *A1.02/B3.01*, *BEL.03/O.01.02*, *A1.02/BW.07* and *BA.04/O.01.01*. The specific test results for each method for these samples are shown in Table 4. Given that the PCR-SSP kit was designed with primers specific for only common ABO subtypes, the typing could not be interpreted in three samples (1, 4, 5), and the typing did not match the final genotype in three samples (2, 3, 6). Four samples (1, 2, 3, 5) for which the PCR-SBT results disagreed with the SMRT sequencing results are described in detail in the following sections.

TABLE 4 Serology, polymerase chain reaction with sequence-specific primers (PCR-SSP), PCR sequence-based typing (PCR-SBT) and single-molecule real-time (SMRT) sequencing results from 21 samples.

		ABO serology						
		Forward		Reverse				
Sample	Phenotype	Anti-A	Anti-B	A cell	B cell	PCR-SSP	PCR-SBT	SMRT
1	AB_{weak}	4+	mf	–	–	\	A1.02/B.01	A1.02/B.NEW
2	$A_{weak}B$	2+	4+	1+	–	AB	B(A)new/B.01	cisAB.05/B.NEW
3	AB3	4+	–	4+	4+	AB	A1.02/B.01	A1.02/B3.01
4	Bel	–	–	4+	–	\	BEL.03/O.01.02	BEL.03/O.01.02
5	AB_{weak}	4+	–	–	–	\	A1.02/B.01	A1.02/BW.07
6	B(A)	2+	4+	1+	–	BO_1	BA.04/O.01.01	BA.04/O.01.01
7	A	4+	–	2+	2+	AA	A1.02/A1.02	A1.02/A1.02
8	AB	4+	4+	3+	2+	AB	A1.02/B.01	A1.02/B.01
9	B	–	4+	+	+	BO_1	B.01/O.01.01	B.01/O.01.01
10	A	–	–	–	4+	AO_2	A1.02/O.01.02	A1.02/O.01.02
11	A	–	–	–	4+	AO_1	A1.02/O.01.01	A1.02/O.01.01
12	AB	4+	4+	–	1+	AB	A1.01/B.01	A1.01/B.01
13	B	–	4+	3+	1+	BO_2	B.01/O.01.02	B.01/O.01.02
14	A	4+	–	+	–	AO_1	A1.02/O.01.01	A1.02/O.01.01
15	AB	4+	4+	+	+	AB	A1.02/B.01	A1.02/B.01
16	B	–	4+	+	+	BO_2	B.01/O.01.02	B.01/O.01.02
17	A	4+	–	3+	3+	AO_1	A1.02/O.01.01	A1.02/O.01.01
18	A	4+	–	3+	3+	AO_1	A1.02/O.01.01	A1.02/O.01.01
19	A	4+	–	2+	3+	AO_2	A1.02/O.01.02	A1.02/O.01.02
20	A	4+	–	3+	3+	AO_2	A1.02/O.01.02	A1.02/O.01.02
21	A	4+	–	2+	3+	AO_1	A1.01/O.01.01	A1.01/O.01.01

TABLE 5 The results of the six ABO subtypes with variations in the exon region by single-molecule real-time (SMRT) sequencing.

Sample	Phenotype	Allele name	Location	Nucleotide change	Amino acid change	GenBank ID/rsID
1	Bweak	ABO*B.NEW(c.28G>A)	Exon 1	c.28G>A	p.Gly10Arg	JN652594
2	B	ABO*B.NEW(c.861C>T)	Exon 7	c.861C>T	p.Ala287Ala	OR565861
3	B_3	ABO*B3.01	Exon 7	c.1054C>T	p.Arg352Trp	rs56390333
4	Bel	ABO*BEL.03	Exon 7	c.502C>T	p.Arg168Trp	rs573234689
5	Bweak	ABO*BW.07	Exon 7	c.1055G>A	p.Arg352Gln	rs1019994127
6	B(A)	ABO*BA.04	Exon 7	c.640A>G	p.Met214Val	rs964984014

Haplotype sequence analysis of ABO alleles

SMRT sequencing of the ABO gene in 21 samples yielded 42 haplotype sequences, including 2 *ABO*A1.01*, 14 *ABO*A1.02*, 6 *ABO*B.01*, 7 *ABO*O.01.01*, 6 *ABO*O.01.02*, 1 *ABO*B.NEW (c.28G>A)*, 1 *ABO*B.NEW (c.861C>T)*, 1 *ABO*cisAB.05*, 1 *ABO*B3.01*, 1 *ABO*BEL.03*, 1 *ABO*BW.07* and 1 *ABO*BA.04*. The exonic SNVs of each haplotype sequence are shown in Table S1, with key variant sites in bold. Two alleles were found that are not yet listed in the official ABO (ISBT001) blood group 624 allele list (v1.1 17102) [25], and the corresponding new variant sites are highlighted in red.

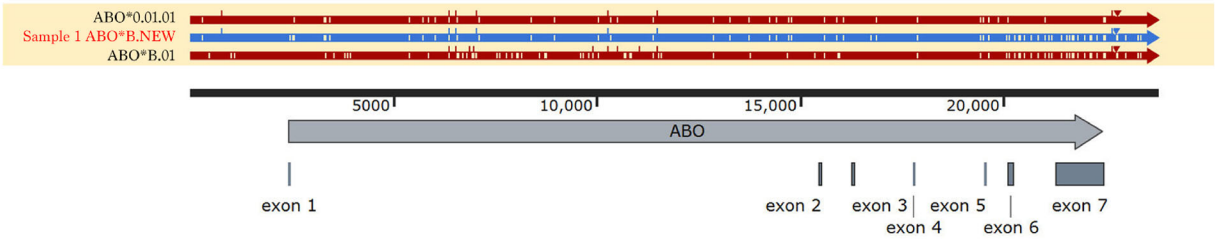

FIGURE 2 Schematic diagram of allelic recombination in sample 1. The sequences from intron 1 to intron 5 has the characteristics of the ABO*O.01.01 but the sequences from exon 6 to exon 7 have characteristics for ABO*B.01. The recombination event maybe happens from c.240-219G>A to c.240-25A>G.

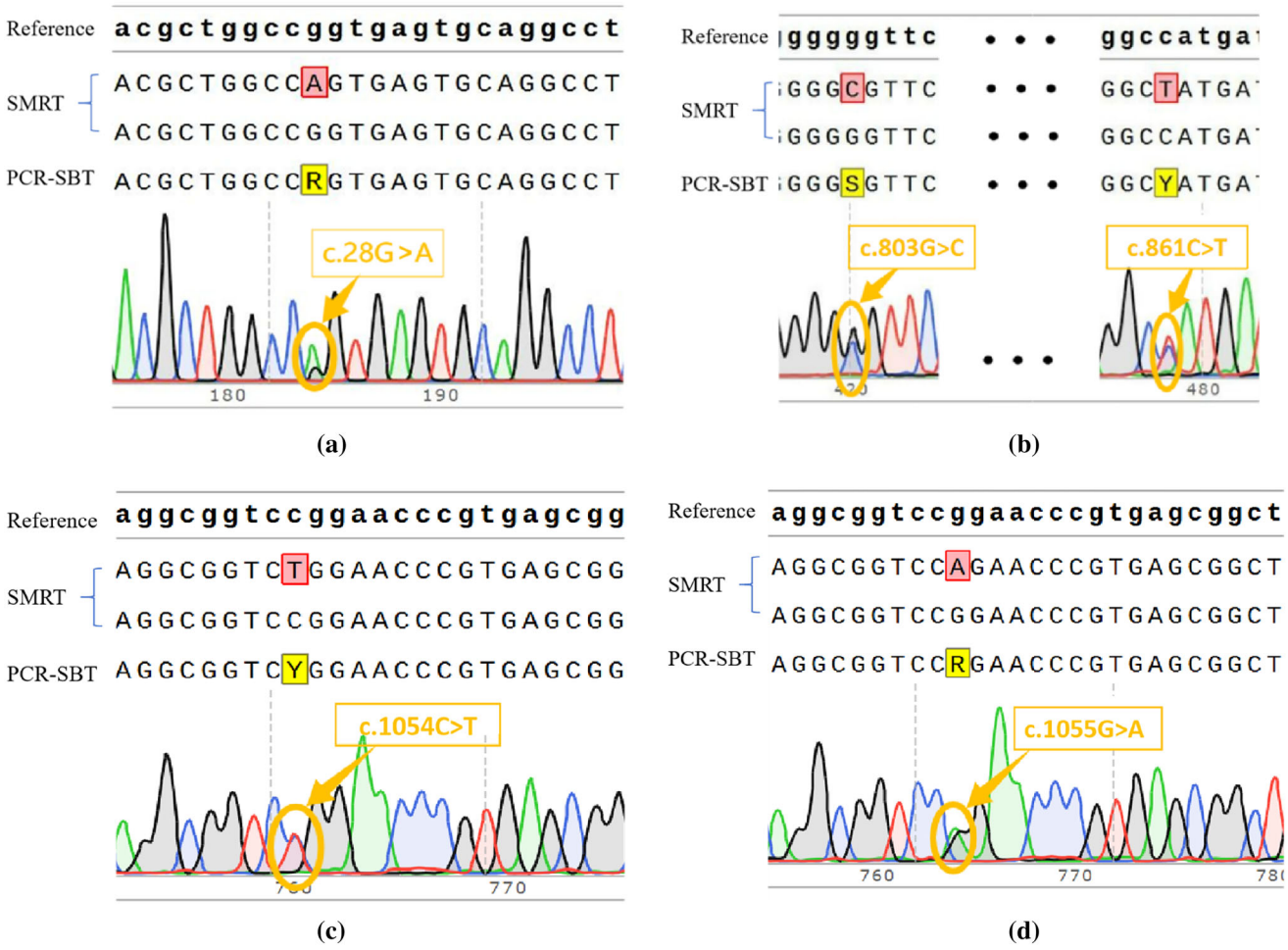

FIGURE 3 (a) In sample 1, the polymerase chain reaction sequence-based typing (PCR-SBT) method was unable to determine the haplotype because of c.28G+A heterozygosity. (b) In sample 2, the PCR-SBT method was unable to determine the haplotype because of c.803G+C and c.861C+T heterozygosity. (c) In sample 3, the PCR-SBT method was unable to determine the haplotype because of c.1054C+T heterozygosity. (d) In sample 5, the PCR-SBT method was unable to determine the haplotype because of c.1055G+A heterozygosity. SMRT, single-molecule real-time.

TABLE 6 The specific intronic region single-nucleotide variants (SNVs) for different ABO alleles (Reference sequence:GenBank ID: NG_006669.2).

Genotype	Location	Nucleotide change	RS ID	Genotype	Location	Nucleotide change	RS ID
ABO*A1.02	Promoter region	c.1-1201A>G		ABO*O1.02	Promoter region	c.1-840G>T	
	Promoter region	c.1-1326_1-1321del			Promoter region	c.1-593C>T	
	Promoter region	c.1-1011_1-975del			Intron1	c.28+175C>T	rs537895
	Intron1	c.28+748C>T	rs532436		Intron1	c.28+3660C>A	rs663054
	Intron1	c.28+1180C>T	rs507666		Intron1	c.28+3840G>A	rs551100
	Intron1	c.28+3981G>A	rs550057		Intron1	c.28+5888C>A	rs596141
	Intron1	c.28+5152G>T	rs9411378		Intron1	c.28+5984C>A	rs66697526
	Intron1	c.29-4303G>A	rs2519093		Intron1	c.28+6043T>C	rs488775
ABO*B.01	Promoter region	c.1-1948G>A			Intron1	c.28+6467T>C	rs574311
	Intron1	c.28+1429G>A	rs8176644		Intron4	c.203+359T>G	rs626792
	Intron1	c.28+4268A>G	rs79343853		Intron4	c.203+751T>G	rs638756
	Intron1	c.28+5163G>T	rs587611953		Intron4	c.204-512C>T	rs517414
	Intron1	c.28+6123T>C	rs8176662		Intron4	c.204-221G>A	rs514708
	Intron1	c.29-4746T>G	rs8176671		Intron4	c.204-192T>C	rs641943
	Intron1	c.29-4618G>A	rs8176672		Intron4	c.204-177T>G	rs641959
	Intron1	c.29-3938C>T	rs13299342		Intron5	c.240-249C>T	rs8176714
	Intron1	c.29-86G>A	rs8176693		Intron5	c.240-105C>A	rs8176717
	Intron5	c.240-25A>G	rs75179845		Intron5	c.240-28G>A	rs8176718
	Intron6	c.374+42G>T	rs8176722		Intron6	c.374+89T>A	rs2073825
	Intron6	c.374+271A>G	rs8176730		Intron6	c.374+188G>A	rs8176726
	Intron6	c.374+280C>T	rs2073823		Intron6	c.374+226C>T	rs8176727
	Intron6	c.375-425A>G	rs8176733		Intron6	c.374+235C>G	rs8176728
	Intron6	c.375-152G>A	rs7855255		Intron6	c.374+493T>C	rs8176732
	3′-UTR	c.1065+31G>A	rs8176751		Intron6	c.375-336G>A	rs8176734
	3′-UTR	c.1065+194_1064+197del			Intron6	c.375-42A>G	rs8176736
	3′-UTR	c.1065+312G>A	rs187099314		Intron6	c.375-40G>A	rs8176737
	3′-UTR downstream region	c.1065+824G>C			3′-UTR	c.1065+328_1065+350del	
	3′-UTR downstream region	c.1065+1338G>A			3′-UTR	c.1065+352_1065+386del	
	3′-UTR downstream region	c.1065+1406C>T			3′-UTR	c.1065+216A>C	rs373302536
ABO*O1.01	Promoter region	c.1-1471C>G			3′UTR	c.1065+228C>G	rs370952072
	Intron1	c.28+3285C>T	rs8176649		3′-UTR	c.1065+229A>C	rs199555421
	Intron1	c.28+3568G>A	rs7046674		3′-UTR	c.1065+345G>A	rs62636488
	Intron1	c.28+5951C>T	rs7036642		3′-UTR	c.1065+351G>A	rs62636487
	Intron1	c.29-6491T>A	rs8176668		3′-UTR	c.1065+359G>C	rs112981202
	Intron1	c.29-2184A>G	rs8176681		3′-UTR	c.1065+364T>C	rs62641788
	Intron1	c.29-1727G>A	rs8176682		3′-UTR	c.1065+372C>T	rs111926917
	Intron1	c.29-1206A>G	rs8176686		3′-UTR	c.1065+376A > G	rs62641786

(Continues)

TABLE 6 (Continued)

Genotype	Location	Nucleotide change	RS ID	Genotype	Location	Nucleotide change	RS ID
	Intron1	c.29-1054_ c.29-1038del			3′-UTR	c.1065+384A>G	rs62641785
	Intron1	c.29-746T>C	rs8176690		3′-UTR	c.1065+414C>T	rs113820458
	Intron1	c.29-658G>A	rs8176691		3′-UTR	c.1065+431G>T	rs62641782
	Intron2	c.99-363C>T	rs2073828		3′-UTR	c.1065+443C>T	rs7466265
	Intron2	c.99-356C>G	rs2073827		3′-UTR	c.1065+449A>G	rs113403969
	Intron2	c.99-186C>A	rs2073826		3′-UTR downstream region	c.1065+1337C>T	
	Intron3	c.155+575C>T	rs8176702		3′-UTR downstream region	c.1065+1411A>C	
	Intron4	c.203+115C>A	rs8176707		3′-quantitative reverse transcription PCR UTR downstream region	c.1065+1442C>T	
	Intron4	c.204-9T>C	rs4962040				
	Intron5	c.240-219G>A	rs8176715				
	Intron6	c.375-269G>A	rs7873635				

A novel B allele and an allelic recombination were identified

We obtained six haplotypes of the ABO subtypes using SMRT sequencing. The locations, nucleotide changes and amino acid changes in key variant sites are shown in Table 5. The novel allele *ABO*B.NEW (c.861C>T)*, which is a synonymous variant located in the exon 7 region, was identified in sample 2, but it is not yet clear whether this variant affects protein expression. The full sequence has been uploaded to GenBank (OR565861) by the authors. In addition, the *ABO*B.NEW(c.28G>A)* haplotype of sample 1 showed the c.28G>A variant, which was not indexed by ISBT but was submitted to BGMUT upon its discovery by Cai's team at Shanghai Blood Station [27, 28]. The characteristic sequence of *ABO*O1.01* appeared in the intron 1 to intron 5 segment of *ABO*B.NEW(c.28G>A)* haplotype, and allelic recombination may have occurred (Figure 2).

Reasons for discrepancies between the PCR-SBT and SMRT sequencing results in four samples

Four samples showed discrepancies between the PCR-SBT and SMRT sequencing results and were therefore analysed further (Figure 3). The PCR-SBT method is not able to accurately identify which haplotype the variant site is located on, but rather relies on the probability calculations of the biosignature software to infer the final genotype. Therefore, the difference in the final interpretation is not a result of inaccurate SMRT sequencing. On the contrary, SMRT sequencing can be used to sequence both haplotypes of a chromosome pair separately and accurately distinguish the haplotype in which the variant site is located, which is an advantage over traditional sequencing methods.

The specific intronic region SNVs for different ABO alleles

Intronic regions and regions upstream and downstream of coding regions were analysed for 42 ABO haplotype sequences. The SNVs in these regions that were specific for different ABO alleles were counted. Among them, *ABO*A1.02*, *ABO*B.01*, *ABO*O1.01* and *ABO*O1.02* had 8, 21, 19 and 66 specific intronic region SNVs, respectively, as shown in Table 6.

Comparative mapping of ABO allele haplotypes

Variants in 42 ABO haplotype sequences were summarized and plotted for comparison (Figure 4). For various ABO subtype genes, specific variant sites located in exonic regions as well as in intronic regions can be visualized. Notably, related sequences upstream and downstream of the coding region were also obtained via SMRT sequencing, and these sequences also exhibited specific sequence patterns.

DISCUSSION

With the rapid development of molecular biology techniques, an increasing number of studies on ABO genes have been reported.

FIGURE 4 Comparative mapping of ABO allele haplotypes.

However, most of these studies are still limited to exon sequences and do not include the analysis of introns or sequences upstream and downstream of the coding region; moreover, reports on the full-length genes of rare ABO subtypes are even scarcer. The total length of the ABO blood group gene is approximately 25 kb, of which intron 1 is the longest, at up to 13 kb, making it difficult to obtain the complete gene sequence using conventional sequencing techniques. Although studies have successfully used NGS technology to obtain the full-length sequence of the ABO gene, haplotype analysis for the ABO gene must still be limited to samples of individuals with a homozygous genotype, as NGS technology cannot distinguish between haplotypes [13]. Compared to previous generations of sequencing technology, the average read length of third-generation sequencing technology is as high as 100 kb, and therefore this sequencing technology can successfully distinguish between haplotypes [15–17], which is a unique advantage when used to study ABO gene sequences. In addition, Oxford nanopore technology (ONT) is equally capable of long-read sequencing and distinguishing haplotypes. However, due to the high error rate (the raw error rate is around 5%, and R10.3 chip delivers up to 96% sequencing accuracy), ONT is not

suitable for accurate sequencing of ABO genes but is more suitable for rapid sequencing in emergency situations. Although SMRT sequencing has base read errors, these errors are random and can be self-corrected by Circular Consensus Sequencing (sequencing accuracy >99.9%). Therefore, in the present study, the complete sequence of the coding region of the ABO gene, as well as sequence upstream and downstream of the coding region, of the 21 samples was obtained using SMRT sequencing and differentiated into 42 haplotypes. Comparison with PCR-SBT showed consistent detection results, demonstrating the high accuracy of SMRT sequencing.

Full-length sequences were obtained from six samples with rare ABO subtypes using SMRT sequencing. These complete ABO subtype sequences have been poorly accessed and studied to date. A new ABO allele (*ABO*B.NEW, c.861C>T*) with a key synonymous variant was identified. In a recent study by Clark, it was found that synonymous variants could alter a protein folding mechanism in vivo, leading to changes in cellular fitness [29]. However, there are no studies mentioning whether synonymous variants in ABO alleles affect antigen expression. So whether the c.861C>T variant affects protein expression needs to be further investigated. In sample 1, an ABO allele not included in the ISBT (*ABO*B.NEW, c.28G>A*) had a variant profile and phenotype consistent with those of the study by Cai's team [27], who evaluated the expression of the cDNA containing the variant by quantitative reverse transcription PCR and concluded that the c.28G>A variant affects splicing [28]. We also performed an elution test on this sample and confirmed the weak expression of B antigens. However, the present study revealed that the characteristic sequence of *ABO*O1.01* appeared in this haplotype, located in the segment from intron 1 to intron 5, which was considered a possible allelic recombination and was also consistent with the findings of He's team [13]; therefore, the molecular mechanism of this isoform needs to be further investigated.

Different ABO genes have specific SNVs, whether in the exonic region, the intronic region or upstream and downstream of the coding region. However, studies on specific SNVs in intronic regions or upstream and downstream of coding regions are rare. Forty-two full-length ABO haplotype sequences were distinguished by SMRT sequencing, and a total of 114 specific SNVs in the intronic region and upstream or downstream of the coding region were analysed using these ABO haplotype collections. These specific SNVs are useful for evaluating ABO allele splicing and genotyping during sequencing, as well as for the detection of ABO gene recombination and other structural variants, and for the design of specific primers for ABO gene analysis. For example, a recent study by Matzhold et al. established an allele-specific long-range SBT method based on four SNP positions in introns 1, 2 and 4 of the ABO gene that were found to be suitable for distinguishing the different ABO haplotypes [30]. For some rare ABO subtypes, there may also be some specific intronic SNVs and sequence patterns [31–38]. However, due to the insufficient sample size, statistical analysis could not be performed, and additional samples of the same genotype need to be collected for full sequence acquisition and analysis in subsequent studies. Moreover, the samples collected to date remain limited; to establish a full-length standardized reference sequence set of common ABO gene subtypes in the Chinese population, a larger number of samples covering a wider range of genetic diversity should be collected in subsequent studies.

In summary, SMRT sequencing allows access to accurate and complete ABO gene sequences, differentiation of haplotypes and identification of allelic recombination. The 7 rare and 35 common ABO haplotype full-length sequences collected in this study and made publicly available should provide a valuable resource for molecular diagnosis and genetic evolution studies of ABO genes in Chinese populations.

ACKNOWLEDGEMENTS

This work was supported by the Wanjiang Emerging Industries Technology Development Center Industrialization Project (grant number: WJ21CYHXM11).

Z.W. performed the research and wrote the first draft of the manuscript, Z.W. and Y.C. designed the research study and collected the sample, Z.W., Y.C. and Y.X. acquired and analysed the data and M.B. supervised the research and reviewed and edited the manuscript.

CONFLICT OF INTEREST STATEMENT

The authors declare no conflicts of interest.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available from the corresponding author upon reasonable request.

ORCID

Zhe Wang https://orcid.org/0009-0001-0806-8603

REFERENCES

1. Hu L. Clinical blood transfusion test. 3rd ed. Chichester, China: China Medical Science and Technology Press; 2015.
2. Chen D-P, Wen Y, Lu J-J, Tseng C-P, Wang W-T. Rapid rare ABO blood typing using a single PCR based on a multiplex SNaPshot reaction. J Formos Med Assoc. 2019;118:395–400.
3. Zhao H, Miao T, Yan F, Liu Y. The serological and genetic characterization of CisAB blood group in a Chinese family. Transfus Apher Sci. 2015;53:220–4.
4. Wang C, Li S, Li M, Wu X, Xing X, Lyu R. Identification and genotyping of difficult blood groups in the patients with phymatosis. J Exper Hematol. 2013;21:780–4.
5. Kim M-S, Kim JS, Park H, Chung Y, Kim H, Ko D-H, et al. The first case of Para-Bombay blood type encountered in a Korean tertiary hospital. J Korean Med Sci. 2019;34:e258.
6. Rattanapan Y, Charong N, Narkpetch S, Chareonsirisuthigul T. Genotyping of the rare Para-Bombay blood group in southern Thailand. Hematol Transfus Cell Ther. 2023;45:449–55.
7. Suzuki Y, Isa K, Ogasawara K, Kikuchi Y, Yabe R, Tsuno N-H, et al. A novel c.166A>T (p.Thr56Ser) mutation in GYPB*S accounting for unusual S antigen expression. Vox Sang. 2019;114:171–3.
8. Datta SS, Rophina M, Scaria V. Molecular analysis and transfusion management in a rare case of *cis*-AB blood group: a report from India. Transfus Clin Biol. 2024;31:31–5.

9. Orzińska A, Guz K, Mikula M, Kulecka M, Kluska A, Balabas A, et al. A preliminary evaluation of next-generation sequencing as a screening tool for targeted genotyping of erythrocyte and platelet antigens in blood donors. Blood Transfus. 2018;16:285–92.
10. Schoeman EM, Lopez GH, McGowan EC, Millard GM, O'Brien H, Roulis EV, et al. Evaluation of targeted exome sequencing for 28 protein-based blood group systems, including the homologous gene systems, for blood group genotyping. Transfusion. 2017;57:1078–88.
11. Fichou Y, Mariez M, Le Maréchal C, Férec C. The experience of extended blood group genotyping by next-generation sequencing (NGS): investigation of patients with sickle-cell disease. Vox Sang. 2016;111:418–24.
12. Dezan MR, Ribeiro IH, Oliveira VB, Vieira JB, Gomes FC, Franco LAM, et al. *RHD* and *RHCE* genotyping by next-generation sequencing is an effective strategy to identify molecular variants within sickle cell disease patients. Blood Cells Mol Dis. 2017;65:8–15.
13. 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.
14. Zhang Y, Xu H, Zhou H. Principles of erythrocyte blood grouping and detection strategies. 2nd ed. Chichester: People's Health Press; 2019.
15. Carneiro MO, Russ C, Ross MG, Gabriel SB, Nusbaum C, DePristo MA. Pacific biosciences sequencing technology for genotyping and variation discovery in human data. BMC Genomics. 2012;13:375.
16. Rhoads A, Au KF. PacBio sequencing and its applications. Genomics Proteomics Bioinformatics. 2015;13:278–89.
17. Quick J, Loman NJ, Duraffour S, Simpson JT, Severi E, Cowley L, et al. Real-time, portable genome sequencing for Ebola surveillance. Nature. 2016;530:228–32.
18. 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 Hemotherapy. 2020;47:23–32.
19. 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.
20. Tounsi WA, Lenis VP, Tammi SM, Sainio S, Haimila K, Avent ND, et al. Rh blood group D antigen genotyping using a portable nanopore-based sequencing device: proof of principle. Clin Chem. 2022;68:1196–201.
21. Ji Y, Luo Y, Wen J, Sun Y, Jia S, Ou C, et al. Patients with Asian-type DEL can safely be transfused with RhD-positive blood. Blood. 2023;141:2141–50.
22. Gueuning M, Thun GA, Wittig M, Galati A-L, 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.
23. Department of Medical Science, Ministry of Health, People's Republic of China. National Clinical Laboratory Procedures. 3ed ed. Chichester: Southeast University Press; 2006.
24. Cai X, Jin S, Liu X, Shen W, Lu Q, Wang J, et al. Molecular genetic analysis for the B x subgroup revealing two novel alleles in the ABO gene. Transfusion. 2008;48:2442–7.
25. The International Society of Blood Transfusion. ABO (ISBT 001) blood group allele. https://www.isbtweb.org/fileadmin/user_upload/Working_parties/WP_on_Red_Cell_Immunogenetics_and/001_ABO_Alleles_v1.2.pdf. Last accessed 18 Jan 2024.
26. Wang C, Tang Y, Zhang P, Xiong L, Chen W, Lv X. Detection and phenotype analysis of a novel Ael blood group allele. Vox Sang. 2024;119:74–8.
27. Cai X, Jin S, Liu X, Fan L, Lu Q, Wang J, et al. Molecular genetic analysis of ABO blood group variations reveals 29 novel ABO subgroup alleles. Transfusion. 2013;53:2910–6.
28. Cai X, Qian C, Wu W, Lei H, Ding Q, Zou W, et al. An exonic missense mutation c.28G>A is associated with weak B blood group by affecting RNA splicing of the ABO gene. Transfusion. 2017;57:2140–9.
29. Walsh IM, Bowman MA, Soto Santarriaga IF, Rodriguez A, Clark PL. Synonymous codon substitutions perturb cotranslational protein folding in vivo and impair cell fitness. Proc Natl Acad Sci USA. 2020;117:3528–34.
30. Matzhold EM, Drexler-Helmberg C, Helmberg W, Wagner A, Wagner T. Allele-specific long-range sequencing as a method for ABO haplotyping in clinical blood group diagnosis and immunohematology research. Mol Genet Genomic Med. 2024;12:e2286.
31. Sano R, Nakajima T, Takahashi K, Kubo R, Kominato Y, Tsukada J, et al. Expression of ABO blood-group genes is dependent upon an erythroid cell–specific regulatory element that is deleted in persons with the Bm phenotype. Blood. 2012;119:5301–10.
32. Nakajima T, Sano R, Takahashi Y, Kubo R, Takahashi K, Kominato Y, et al. Mutation of the GATA site in the erythroid cell–specific regulatory element of the ABO gene in a Bm subgroup individual. Transfusion. 2013;53:2917–27.
33. Takahashi Y, Isa K, Sano R, Nakajima T, Kubo R, Takahashi K, et al. Presence of nucleotide substitutions in transcriptional regulatory elements such as the erythroid cell-specific enhancer-like element and the ABO promoter in individuals with phenotypes A3 and B3, respectively. Vox Sang. 2014;107:171–80.
34. Tao C, Xiao J, Hu Y, Huang C, Sun J, Li M, et al. A novel B allele with c.28 + 5885C>T substitution in the erythroid cell–specific regulatory element identified in an individual with phenotype B3. Transfusion. 2017;57:1318–9.
35. Liu F, Li G, Li J, Gui R, Luo Y, Zhou M. A novel mutation eliminates GATA-1 and RUNX1-mediated promoter activity in galactosyltransferase gene. Transfus Med Hemother. 2022;49:331–7.
36. Ying Y, Hong X, Xu X, Ma K, He J, Zhu F. A novel mutation +5904 C>T of RUNX1 site in the erythroid cell-specific regulatory element decreases the ABO antigen expression in Chinese population. Vox Sang. 2018;113:594–600.
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 Bm and A1Bm phenotypes. Vox Sang. 2018;113:393–6.
38. Hayakawa A, Sano R, Takahashi Y, Kubo R, Harada M, Omata M, et al. RUNX1 mutation in a patient with myelodysplastic syndrome and decreased erythrocyte expression of blood group A antigen. Transfusion. 2020;60:184–96.

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Wang Z, Chu Y, Xiao Y, Bian M. Detecting serologically difficult ABO blood groups using single-molecule real-time sequencing technology. Vox Sang. 2024;119:1096–105.