Transfusion Medicine Reviews xxx (xxxx) xxx

Contents lists available at ScienceDirect

Transfusion Medicine Reviews

journal homepage: https://www.journals.elsevier.com/transfusion-medicine-reviews

The Research of a Large-Scale Analysis Platform for MNS Blood Group Identification Based on Long-Read Sequencing

Hua Xu*, Xiaomin Su, Qinqin Zuo, Liangzi Zhang, Xiaoyue Chu

Shaanxi Blood Center, Xi'an, China

ARTICLE INFO

Article history:
Available online xxx

Keywords:
MNS blood group
The LRS-PCR
Haplotypes
Novel mutation

ABSTRACT

The objective of this study was to devise a novel approach for determining MNS blood group utilizing long-read sequencing (LRS) and to identify intricate genome variations associated with this blood group system. In this study, a total of 60 blood samples were collected from randomly selected Chinese Han blood donors. The amplification of the full-length sequences of *GYPA* exon 2-7 (11 kb) and *GYPB* exon 2-6 (7 kb) was conducted on the blood samples obtained from these 60 donors. Subsequently, the sequencing of these amplified sequences was performed using the PacBio platform. The obtained sequencing data were then compared with the reference sequence of the human genome (GRCh38) utilizing the pbmm2 software, resulting in the acquisition of the haploid sequences of *GYPA* and *GYPB*. The serological typing prediction was conducted using the International Society of Blood Transfusion (ISBT) database, while the analysis of SNVs sites was performed using deepvariant v1.2.0 software and reference sequence alignment. A total of 60 samples yielded unambiguous high-quality haplotypes, which can serve as a standardized reference sequence for molecular biology typing of MNSs in the Chinese population. In a total of 60 serological samples, the LRS method successfully identified the M, N, S, and s blood group antigens by analyzing specific genetic variations (c.59, c.71, c.72 for *GYPA*, and c.143 for *GYPB*), which aligned with the results obtained through conventional serological techniques. 4 Mur samples that had been previously validated through serology and molecular biology were successfully confirmed, and complete haploid sequences were obtained. Notably, one of the Mur samples exhibited a novel breakpoint, *GYP* (*B1-136-B ψ 137-212-A213-229-B230-366*), thereby representing a newly identified subtype. Single molecule sequencing, which eliminates the necessity for PCR amplification, effectively encompasses GC and high repeat regions, enhancing accuracy in quantifying mutations with low abundance or frequency. By employing LRS analysis of the core region of *GYPA* and *GYPB*, diverse genotypes of MNS can be precisely and reliably identified in a single assay. This approach presents a comprehensive, expeditious, and precise novel method for the categorization and investigation of MNS blood group system.

Introduction

The MNS blood group system is ranked as the second most complex system, following the Rh system. Currently, the ISBT has recorded 50 MNS antigens [1]. These antigens are produced by a gene cluster consisting of *GYPA, GYPB*, and *GYPE*, which exhibit a homologous sequence similarity of over 95 % and possess a comparable exon-intron organization. This gene cluster is located on Chromosome 4q31.22. [2], spanning approximately 350 kb. The MNS system is characterized by the presence of major polymorphic antigens, namely M/N and S/s. These antigens are the result of specific amino acid substitutions occurring at residues 20 and 24 of *GYPA*, as well as a single amino acid substitution at residue 29 of *GYPB*. Furthermore, the occurrence of homologous gene crossover and conversion events between *GYPA* and *GYPB* leads to the formation of various variants of MNS hybrid glycophorins, including *GYP(A-B-A), GYP(B-A-B), GYP(A-B)*, and *GYP(B-A)* [3,4]. In rare instances, gene conversion involving *GYPE* also takes place, resulting in the formation of *GYP(A-E-A)* and *GYP(B-E-B)* [5]. The glycophorin genes encoding the MNS blood group antigens exhibits a high level of genetic polymorphism and complexity. Recent research findings indicate that there exist distinct racial and regional variations in the distribution of allele frequencies within the MNS blood group system, with particular emphasis on the M, N, S, and s antigens. The MNS antigens may stimulate an immune response and the production of specific alloantibodies through transfusion and/or pregnancy. Within the MNS blood group system, there are many low prevalence antigens that arise on different hybrid glycophorin molecules, e.g., GP.Mur. Among them, Mi^a (MNS7) and Mur (MNS10) blood group antigens are considered the most

* Correspondence to: Hua Xu, Blood Center of the Shaanxi Province, Institute of Xi'an Blood Bank, 407# Zhuque Ave, Xi'an, Shaanxi Province 710061, China.
E-mail address: drxuhua@126.com (H. Xu).

https://doi.org/10.1016/j.tmrv.2024.150836

Please cite this article as: H. Xu, X. Su, Q. Zuo et al., The Research of a Large-Scale Analysis Platform for MNS Blood Group Identification Based on Long-Read Sequencing, Transfusion Medicine Reviews, https://doi.org/10.1016/j.tmrv.2024.150836

clinically significant antigens, particularly in southern China and Southeast Asian countries. Their prevalence in these regions is significantly higher compared to the European and American White population [6-10]. Numerous studies have reported that the presence of corresponding antibodies to these antigens can lead to hemolytic transfusion reactions (HTR) and neonatal hemolytic disease [11-13]. Consequently, precise typing of the MNS blood group system holds considerable clinical importance in guiding blood transfusion practices and monitoring antibodies in patients.

The hemagglutination technique is commonly used in clinical blood typing for practical identification purposes. However, serological typing has its limitations due to the unavailability of commercialized reagents and experimental experience, which can potentially impact the accuracy of the results. Specifically, the determination of hybrid glycophorins in the MNS system necessitates multiple rare typing reagents that are not commonly found in laboratories [1,4]. Therefore, the molecular methods are deemed more suitable as they allow for the detection of specific variants. Currently, several genotyping methods have been developed to characterize hybrid glycophorins, including PCR-sequence specific primer, high-resolution melting (HRM), multiplex ligation-dependent probe amplification, and next-generation sequencing, among others [14-18]. However, to date, no single molecular genotyping method has the capability to detect and differentiate all variant types, some of which are also impractical for large-scale sample screening. The advent of long-read sequencing (LRS) technologies has eliminated the requirement for DNA fragmentation, enabling the sequencing of repetitive sequences in their entirety and facilitating continuous and comprehensive assembly. Currently, widely employed real-time single-molecule sequencing technology, exemplified by PacBio Single Molecule Real-Time (SMRT), and nanopore sequencing technology, exemplified by Oxford Nanopore Technology (ONT), offer continuous long-reads with an average read length exceeding 10 kb. These advancements present a novel approach for detecting variations in genome structure and identifying homologous genes. They can all cross highly repetitive regions in the genome to achieve more precise and complete detection of nucleic acid molecules, identify large fragments of variation and more transcripts in the genome.

In recent years, the *RHD-RHCE* region has been assembled in an unbiased and comprehensive manner using the *de novo* approach on the PacBio platform [19]. The PacBio sequencing platform significantly reduces the cost when processing multiple samples in a centralized manner, especially its high accuracy, which allows each sample to be sequenced with only a certain number of reads (more than 100 reads per haplotype), greatly reducing the sequencing volume and cost. The flexibility of MinIon devices makes it easier to carry out corresponding tests in clinical settings in the future. However, due to their low accuracy, it is necessary to achieve a sufficient number of reads to improve the accuracy of the area to be tested. At the same time, sequencing errors cannot be eliminated. Therefore, their testing costs will be higher during large-scale testing. So, we are more inclined to use PacBio technology during large-scale screening to achieve the goal of accuracy and low cost. Additionally, precise and annotated reference gene/allele sequences for the Duffy blood group have been established through targeted PacBio HiFi sequencing [20]. Given these advancements, our objective was to employ target sequencing on the PacBio platform to examine the core regions of *GYPA* and *GYPB* with known genotypes, while accurately and cost-effectively distinguishing the MNS antigens in a single analysis. We aimed to enhance the capacity for predicting the phenotype of MNS blood group antigens, establish a robust foundation for addressing clinical transfusion concerns associated with MNS blood group antigens.

Methods

Sample Selection and DNA Preparation

Blood samples from randomly selected Chinese blood donors (n = 60) were collected in the Shaanxi Blood center, located in the West region of China. Monoclonal antibodies (MoAbs; anti-M, Clone 11H2; anti-N, 1422C7; anti-S, MS94; anti-s, P3BER; CE-Immundiagnos tika) were used for MNS phenotyping. Peripheral blood samples were collected, and genomic DNAs were extracted with the QIAamp DNA blood mini kit. The quality of the DNA was analyzed through an electrophoresis assay. The prominent bands should be around 10 kb. The DNA also be quantitated by a dsDNA HS assay kit (ThermoFisher Scientific) with Qubit 4.0 fluorometer (ThermoFisher Scientific).

LR-PCR Amplification

Six sets of primers (Supplemental n) were designed for the amplification of the complete sequences of *GYPA* exon and *GYPB* exon. The amplicons had an approximate size of 10 kb (Figure 1B). To

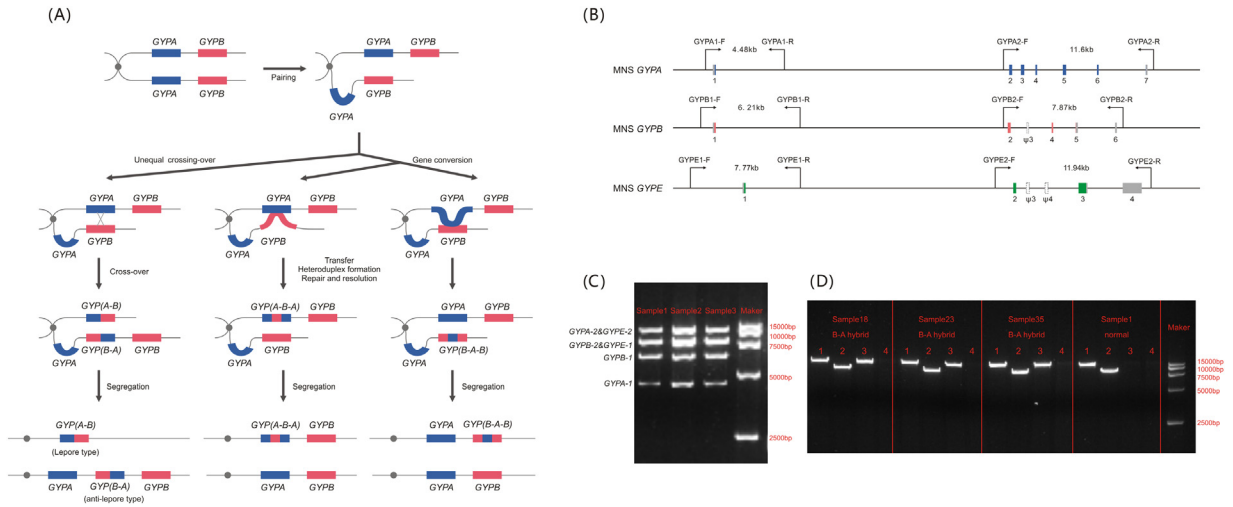

Fig. 1. Strategy for MNS blood group genotyping. (A) Molecular basis of unequal crossing-over event and gene conversion events between homologous *GYPA* and *GYPB*. (B) Position of multiplex PCR primer pairs for long-range PCR. (C) Agarose gel showing the expected amplicons of three samples for six *GYPA*, *GYPB*, and *GYPE* genes fragments in one tube. (D) The amplified products of three *GYP(B-A)* hybrid samples and a normal sample. Lane 1, lane 2, lane 3, lane 4 refer to the PCR product of the primer of *GYPA2-F* and *GYPA2-R*, *GYPB2-F* and *GYPB2-R*, *GYPB2-F* and *GYPA2-R*, *GYPA2-F* and *GYPB2-R*, respectively.

Please cite this article as: H. Xu, X. Su, Q. Zuo et al., The Research of a Large-Scale Analysis Platform for MNS Blood Group Identification Based on Long-Read Sequencing, Transfusion Medicine Reviews, https://doi.org/10.1016/j.tmrv.2024.150836

account for the high similarity between the two genes, primers with 3′ end variations were designed within the intronic region to generate distinct amplicons for each gene. LR-PCR was conducted using KOD FX Neo (TOYOBO) and the PCR cycling conditions were set as per the manufacturer's instructions. A 2-step cycle was employed, with a 10-minute extension time per cycle for a total of 30 cycles. The PCR products were subsequently verified through agarose gel electrophoresis and further purified with 0.6x Ampure PB beads (Pacific Biosciences).

Library Preparation and SMRT Sequencing

The SMRT libraries were constructed using a one-step method, wherein pre-sequencing libraries containing unique barcode adapters were generated by combining DNA damage repair, end-repair, and adapter ligation in a single step. Prior to usage, a reaction master mixture was prepared, consisting of 10 µl, including 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 combined with the enzyme mixture. The resulting reaction mixes were then incubated at specific temperatures: 37°C for 20 minutes, 25°C for 15 minutes, and 65°C for 10 minutes. Subsequently, exonuclease I (Enzymatics) and exonuclease III (Enzymatics) were introduced to eliminate any unsuccessful ligation products. The final pre-library was purified using 0.6× Ampure PB beads. In the case of multiple sample sequencing, the pre-libraries were combined based on equal masses. Following the pooling process, the pre-libraries underwent purification twice using 0.45× Ampure PB beads.

The sequencing enzymes and primers used in the final library were bound using the Sequel Binding Kit 2.2 (Pacific Biosciences) and Internal Control Kit 1.0 (Pacific Biosciences). Subsequently, 150 pM DNA-polymerase complexes were loaded onto the Sequel II platform (Pacific Biosciences) and sequenced for a duration of 20 hours.

Data Analysis and DNA Variant Calling

The primary data analysis was conducted using SMRT Link v10.1.0 software (Pacific Biosciences). Initially, the raw reads were demultiplexed and the barcode sequences were automatically processed at the conclusion of the runs. Subsequently, the subreads were analyzed to generate HiFi reads utilizing the HiFi software application. The filtered HiFi reads were then aligned to the human reference genome (GRCh38) using pbmm2 alignments in order to identify the desired *GYPA* and *GYPB* CCS fragments. The target HiFi reads were subsequently re-aligned to the reference genome (ISBT *GYPA* (NCBI GeneNo: 2993) and *GYPB* (NCBI GeneNo: 2994) gene) using pbmm2. To identify single-nucleotide variations (SNVs) and small indels, variant calling was performed using deep variant v1.2.0 (https://github.com/google/deepvariant/blob/r1.2/docs/deepvariant-quick-start.md). The alignments of variant and wild-type molecules were visualized using Integrative Genomics Viewer. The resulting data were compiled, and individual haplotypes, including gene crossover and conversion region haplotypes, were generated using the SnapGene software module. This software is available for download and use on the official website (https://www.snapgene.com/). The predicted phenotyping results were analyzed by comparing the sequences of *GYPA* and *GYPB* to MNS ISBT blood group alleles v4.1 (https://www.isbtweb.org/resource/002mnsalleles.html).

Results

LR-PCR Amplification and Sequencing of GYPA, GYPB, *and* GYPE *Core Region*

The assay was developed with the purpose of sequencing the core region of *GYPA*, *GYPB*, and *GYPE* using PacBio LR HiFi sequencing. The promoter regions and entire exons of the three genes were amplified and sequenced in order to ascertain the MNS genotypes and potential recombination forms (Figure 1B). The multiplex LR-PCR was conducted by combining all primers in a single PCR reaction, allowing for the simultaneous detection of all target regions and recombination forms as reported by ISBT (Figure 1A and C). For example, the genotype *GYP(B-A)* was selectively amplified using specific *GYPB* forward and *GYPA* reverse primers. Subsequently, the amplified products underwent PacBio target sequencing, and the resulting data were analyzed using a customized variant calling pipeline, as illustrated in Figure 1D.

MNS Genotyping Based on LRS

In this study, a total of 60 samples were selected for the purpose of conducting a comparative analysis of the MNS antigen phenotype. These samples were randomly chosen from Chinese Han blood donors. Initially, the samples underwent serological typing to determine their M/N and S/s phenotypes. The results, as presented in Table 1, indicated that out of the 60 samples, 7 were M+N-, 35 were M+N+, 18 were M-N+, 3 were S+s+, 56 were S-s+, and 1 was a rare S+s-. Subsequently, all 60 samples were subjected to target amplification and sequencing. The obtained data were then subjected to further analysis in order to identify the haplotypes for each sample. Based on the haplotypes of *GYPA* and *GYPB*, a customized variant identification assay was developed to predict the

Table 1
Predicted phenotype by serology and long-read sequencing in all selected 60 samples.

Samples	Genotype	Predicted phenotype	Phenotype by serology M	N	S	s
13	*GYPA*02/GYPA*01;GYPB*04/GYPB*04*	M+/N+/s+	+	+	-	+
1	*GYPA*01/GYPA*02;GYPB*03/GYPB*04*	M+/N+/S+/s+	+	+	+	+
16	*GYPA*02/GYPA*02;GYPB*04/GYPB*04*	N+/s+	-	+	-	+
13	*GYPA*01/GYPA*02;GYPB*04/GYPB*04*	M+/N+/s+	+	+	-	+
1	*GYPA*01/GYPA*02;GYPB*04/GYPB*04*	M+/N+/s+	+	+	+	+
1	*GYPA*02/GYPA*02/GYP*401;GYPB*04/GYPB*04*	N+/s+/St(a+)	-	+	-	+
4	*GYPA*01/GYPA*01;GYPB*04/GYPB*04*	M+/s+	+	-	-	+
2	*GYPA*01/GYPA*02;GYP*501/GYPB*04*	M+/N+/Mur+/s+	+	+	-	+
2	*GYPA*02/GYPA*01/GYP*401;GYPB*04/GYPB*04*	M+/N+/s+/St(a+)	+	+	-	+
3	*GYPA*02/GYPA*01;GYPB*04/GYP*501*	M+/N+/s+/Mur+	+	+	-	+
1	*GYPA*02/GYPA*02;GYPB*03/GYPB*04*	N+/S+/s+	-	+	+	+
1	*GYPA*01/GYPA*01;GYPB*03/GYPB*03*	M+/S+	+	-	+	-
2	*GYPA*01/GYPA*01;GYPB*04/GYP*501*	M+/s+/Mur+	+	-	-	+

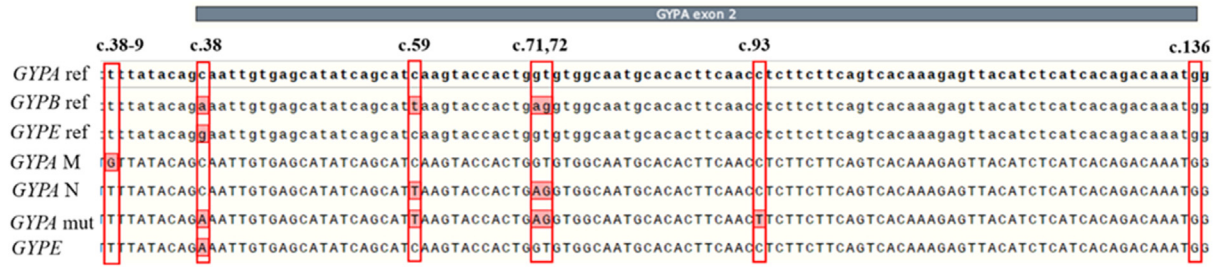

Fig. 2. An alignment of informative nucleotide polymorphisms for *GYPA*, *GYPB*, GYPE, and *GYPA* M allele, *GYPA* N allele reference sequences and *GYPA*, *GYPE* variants sequence. *GYPA* mutant indicated that the N+ allele reference sequence with *c.38 C > A, c.93 C > T* variants. *GYPE* refer to the *GYPE* sequence with *c.38 G > A* variants, which were detected in almost all test sample haplotypes.

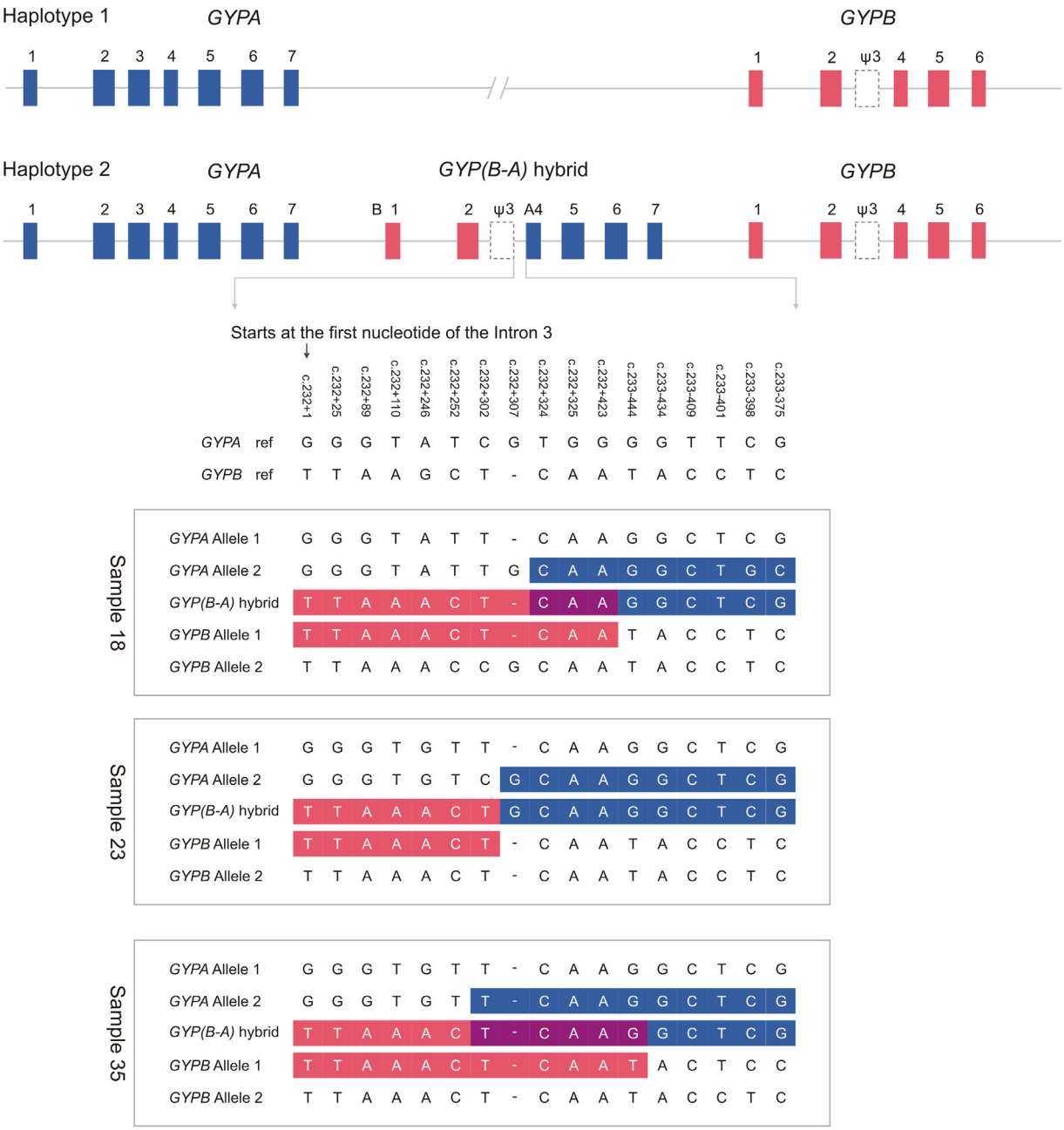

Fig. 3. Proposed alignment of *GYPA* and *GYPB* gene for three *GYP(B-A)* hybrid Sample. Breakpoint regions between *GYPB* and *GYPA* are indicated by purple background. Note: reference is *GYPA* (NG007470.3).

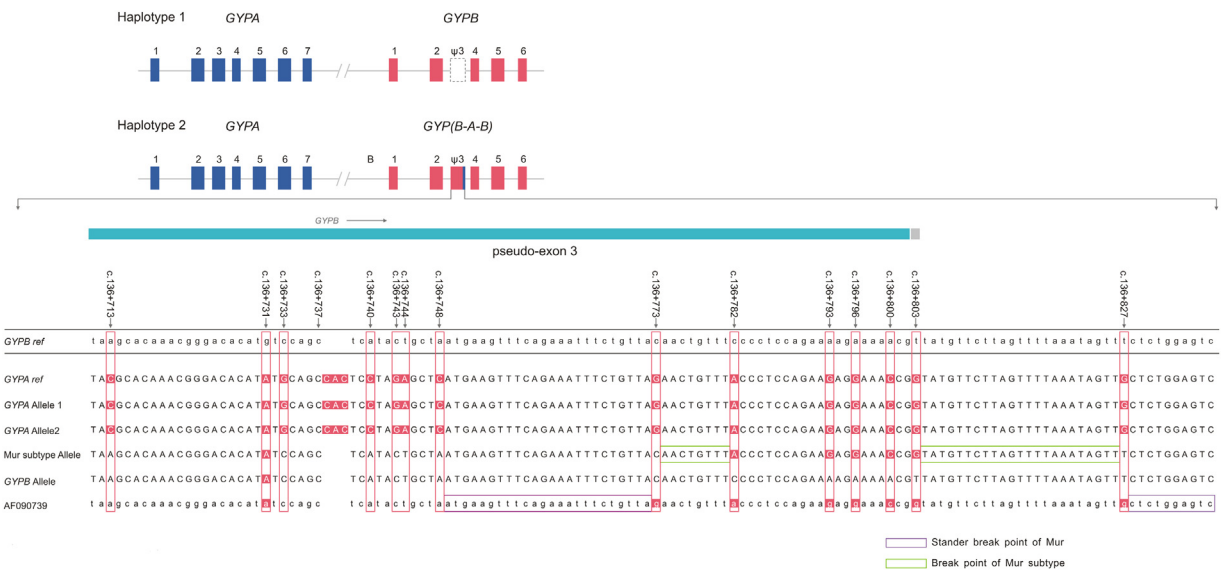

Fig. 4. An alignment of informative nucleotide polymorphisms for *GYPA* and *GYPB* reference sequences, *GYPA* allele 1 and *GYPA* allele 2csequences, *GYP*(*B-A-B*) allele sequences, *GYPB* allele1 sequences and standard Mur sequences (AF090739). *Note*: Reference is *GYPB* (NG_007483.2).

M/N and S/s phenotypes. The sequencing results were consistent with serological typing (Table 1).

Hybrid Glycophorins Genotyping

The haplotype sequence results were obtained through hybrid glycophorins genotyping. A total of 60 samples were analyzed, and the alignment of informative nucleotide polymorphisms was performed for *GYPA, GYPB, GYPE, GYPA* M allele, *GYPA* N allele reference sequences, and *GYPA, GYPE* variants sequence. The presence of a *GYPA* mutant was indicated by the N+ allele reference sequence with c.38 C > A and c.93 C > T variants. Additionally, the *GYPE* sequence with c.38 G > A variants, which were detected in nearly all haplotypes of the test samples, was referred to as *GYPE* (Figure 2). Three of the samples contain the *GYP(B-A)* with a single unequal crossing over at intron 3 of *GYPB*. The cross-over points for the St^a samples No. 18, 23, and 35 are located at 252-490, 302-307, and 252-500, respectively, as depicted in Figure 3. In addition to the four known GP. Mur samples, three other samples were discovered to contain the *GYP(B-A-B)* gene. A novel allele of the *GYP*Mur* subtype, specifically *GYP* (*B1-136-Bψ137-212-A213-229-B230-366*), was identified. The replacement breakpoints of *GYPB* were determined to be located at c.205_212 upstream and c.232+1_24 downstream, as illustrated in Figure 4. Notably, this novel allele exhibits a unique amino acid substitution (c.204G > C, Thr) compared to the reference *GYP*Mur* (AF090739). Gene conversion events between *GYPB* and *GYPE* have been documented in rare S-s- blood phenotypes [5]. In this particular investigation, sample No. 24 exhibits the presence of part intron 3 of *GYPE* within *GYPB* intron 4, thereby resulting in the formation of GYPA recombines with GYPE at the GYP (B-E-B) genotype position of the intron. The point of crossover occurs between nucleotides *GYPB* c3+361-368 and c3+630-1083 (as depicted in Figure 5). Furthermore, sample No. 47 reveals the existence of GYPB recombines with GYPE at the GYP (A-E-A) hybrid allele position. The replacement sequence of *GYPA* commences between c.137-240 and c.137-101, and concludes between c.137-64 and c.137-58 (as illustrated in Figure 6). In summary, 306 and 343 intronic single nucleotide polymorphisms (SNPs) were detected in *GYPA* and *GYPB* respectively.

Table 2
Position of intronic variations and their reference SNP number detected in more than 90% of GYPA allele.

Position	gDNA	Location	cDNA	Ref	Frequency
4:144119810	25942	Intron2	c.137-29	T > A	98.33%
4:144119809	25943	Intron2	c.137-28	G > C	98.33%
4:144119362	26390	Intron3	c.232+324	T > C	100.00%
4:144119361	26391	Intron3	c.232+325	G > A	94.17%
4:144119263	26489	Intron3	c.232+423	G > A	97.50%
4:144119161	26591	Intron3	c.233-409	T > C	100.00%
4:144118506	27246	Intron4	c.271+208	C > A	99.17%
4:144118106	27646	Intron4	c.271+608	A > G	91.67%
4:144117805	27947	Intron4	c.272-866	G > A	95.83%
4:144117776	27976	Intron4	c.272-837	G > A	98.33%
4:144116827	28925	Intron5	c.357+27	G > C	100.00%
4:144116033	29719	Intron5	c.357+821	A > G	100.00%
4:144115486	30264	Intron5	c.358-721	A > AAG	93.33%
4:144115487	30265	Intron5	c.358-720	T > A	93.33%
4:144115458	30294	Intron5	c.358-691	T > C	93.33%
4:144114975	30777	Intron5	c.358-208	A > G	93.33%
4:144112987	32765	Intron6	c.437-1577	A > T	100.00%
4:144112923	32829	Intron6	c.437-1513	G > A	100.00%
4:144112657	33095	Intron6	c.437-1247	T > C	95.00%
4:144112636	33116	Intron6	c.437-1226	G > C	100.00%
4:144112507	33245	Intron6	c.437-1097	A > G	95.00%
4:144114919	35657	3′UTR	c.453+1299	A > T	100.00%
4:144112532	35682	3′UTR	c.453+1324	AT > A	100.00%

Variation in MNS Sequences

Haplotype sequences of *GYPA, GYPB*, and *GYPE* were utilized to identify exon and intron variants. An overview of genetic variations in MNS among haplotypes is presented in Table 2. In comparison to the *GYPA* and *GYPB* reference sequence hg38, a total of nine *GYPA* homozygous SNPs and four *GYPB* homozygous SNPs were observed in all 60 samples, specifically associated with the MNS blood group reference allele. Additionally, 14 *GYPA* SNPs and 24 *GYPB* SNPs were detected in over 90 % of haplotypes (Tables 2 and 3). It is worth noting that multiple intronic SNPs are potentially specific to certain haplotypes (Tables 4 and 5). For instance, the analysis revealed that 23 single nucleotide variant (SNVs) were exclusively present in s+ haplotypes, while absent in S+ haplotypes

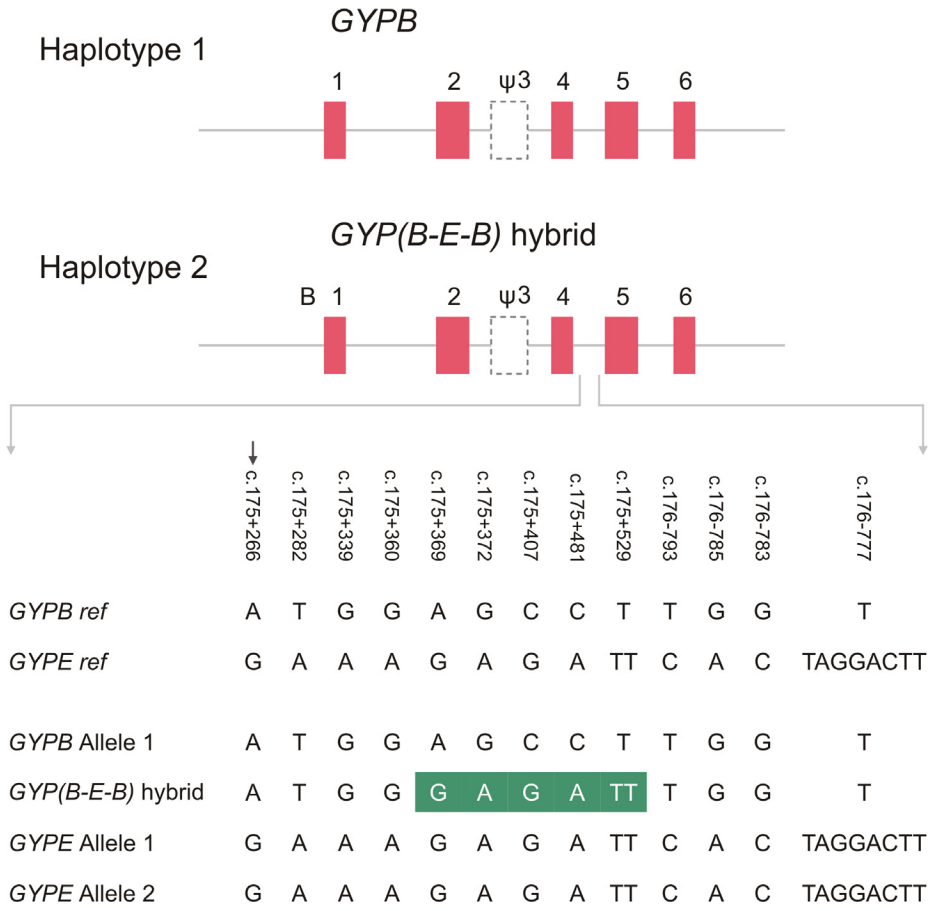

Fig. 5. An alignment of informative nucleotide polymorphisms for *GYPB, GYPE* reference sequences and *GYPB* allele 1, *GYP(B-E-B)* hybrid allele, *GYPE* allele1. *GYPE* allele 2 of the sample 21. *Note*: Reference is *GYPB* (NG_007483.2).

Table 3
Position of intronic variations and their reference SNP number detected in more than 90% of GYPB allele.

Position	gDNA	Location	cDNA	Ref	Frequency
4:144000454	23890	Intron2	c.136+731	G > A	100.00%
4:143997972	26372	Intron3	c.176-338	G > T	96.67%
4:143997906	26438	Intron3	c.176-272	C > T	96.67%
4:143997747	26597	Intron3	c.176-113	C > T	100.00%
4:143997740	26604	Intron3	c.176-106	T > C	96.67%
4:143997714	26630	Intron3	c.176-80	T > C	96.67%
4:143998049	26965	Intron4	c.270+161	A > ATT	96.67%
4:143997132	27212	Intron4	c.270+408	C > G	100.00%
4:143996999	27345	Intron4	c.270+541	C > G	96.67%
4:143996886	27457	Intron4	c.271-583	AC > A	96.67%
4:143996791	27553	Intron4	c.271-487	A > T	96.67%
4:143996792	27554	Intron4	c.271-486	G > GC	96.67%
4:143996657	27687	Intron4	c.271-353	A > G	96.67%
4:143996636	27708	Intron4	c.271-332	T > C	96.67%
4:143996586	27758	Intron4	c.271-282	C > T	95.83%
4:143996463	27881	Intron4	c.271-159	A > G	96.67%
4:143996361	27983	Intron4	c.271-57	T > G	100.00%
4:143995648	28696	3′UTR	c.276+651	G > A	96.67%
4:143995467	28877	3′UTR	c.276+832	T > G	96.67%
4:143995211	29133	3′UTR	c.276+1088	A > G	96.67%
4:143995193	29151	3′UTR	c.276+1106	T > G	96.67%
4:143995170	29174	3′UTR	c.276+1129	G > A	96.67%
4:143995071	29273	3′UTR	c.276+1228	T > G	96.67%
4:143994873	29471	3′UTR	c.276+1426	C > T	96.67%
4:143994811	29533	3′UTR	c.276+1488	T > C	96.67%
4:143994627	29717	3′UTR	c.276+1672	G > A	96.67%
4:143994375	29969	3′UTR	c.276+1924	A > T	96.67%
4:143994316	30028	3′UTR	c.276+1983	G > A	96.67%

(refer to Table 5). In total, 306 and 343 intronic SNPs were identified in *GYPA* and *GYPB* respectively; however, it is noteworthy that the majority of these SNPs did not exhibit conservation across all haplotypes.

Discussion

The genes encoding the MNS blood group antigens exhibit a significant degree of polymorphism and intricate mutations within established human blood type genes. While certain MNS blood group antibodies are naturally occurring IgM class antibodies and typically inactive at 37 °C, it is imperative to acknowledge the significance of IgG and IgM+IgG class immune antibodies in clinical settings. Recent research has indicated that immune anti M, anti N, anti S, and anti-Mur can all lead to hemolytic transfusion reactions and neonatal hemolytic disease, particularly among populations in East and Southeast Asia [21-24]. Furthermore, GP.Mur is a prevalent hybrid glycoprotein within the MNS antigen system depending on the population, expressing five low-frequency antigens: Mur, Mi^a, MUT, Hill, and MINY. The frequencies of anti-"Mur" and anti-"Mi^a" antibodies in South China were found to be higher than those of all other antibodies, indicating their significant roles in HTR [25,26]. There is a growing body of evidence reporting adverse effects such as neonatal hemolysis, HTR, and fetal edema associated with anti-"Mur" and anti-"Mi^a" antibodies [27,28]. It is plausible to suggest that HTR could be prevented by refraining from using blood containing specific Mur and Mi^a antigens in patients receiving chronic transfusions. The MNS blood group system

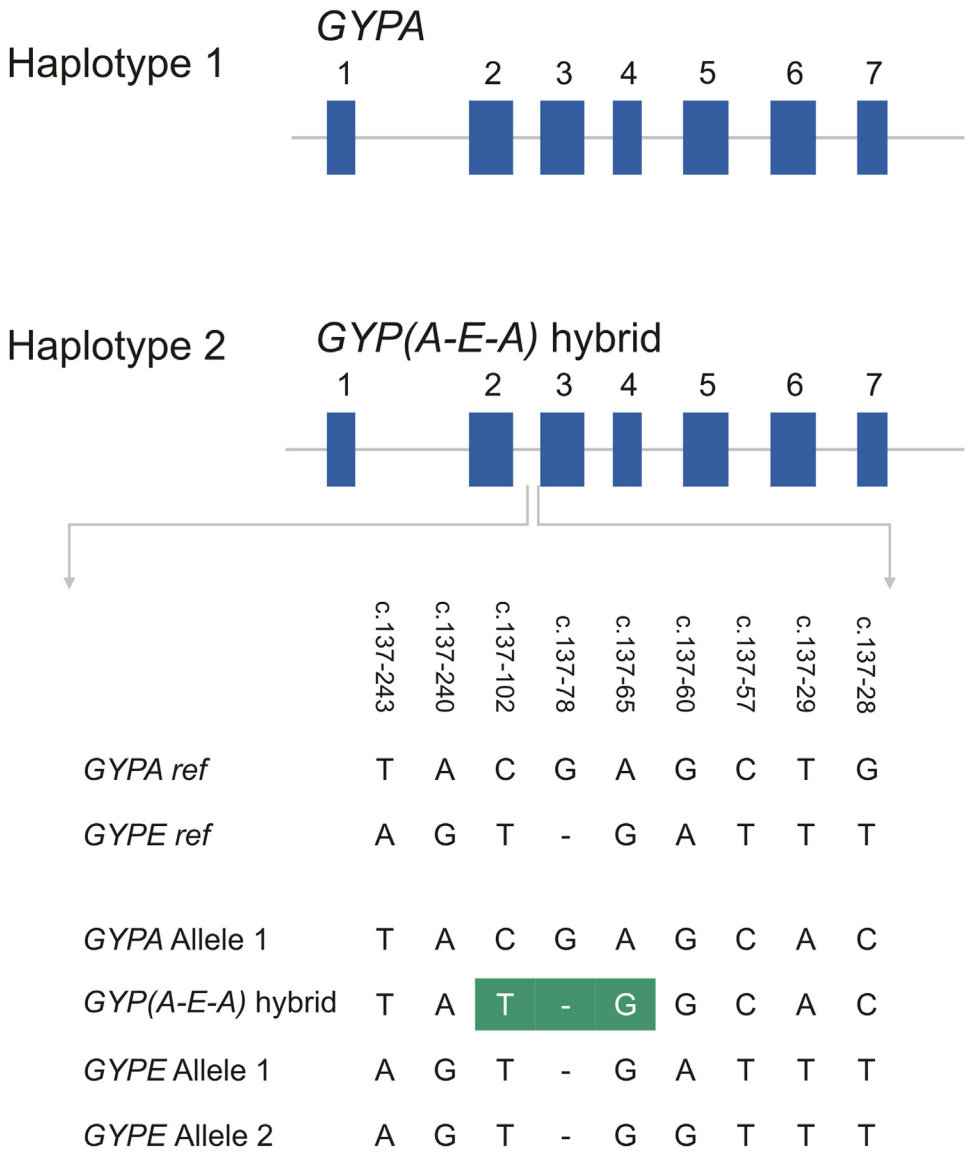

Fig. 6. An alignment of informative nucleotide polymorphisms for *GYPA, GYPE* reference sequences and *GYPA* allele 1, *GYP(A-E-A)* hybrid allele, *GYPE* allele1. *GYPE* allele 2 of the sample 53.

antigens and antibodies hold particular clinical importance in ensuring the safety of clinical blood transfusion practices in China and other areas of the world where there is a large Chinese population. Therefore, the precise identification of the MNS blood group has the potential to be beneficial in mitigating neonatal hemolysis, HTR, and fetal edema, among other related complications.

Serology remains the prevailing technique in immunohematological diagnostics due to its simplicity, specificity, sensitivity, and safety, rendering it appropriate for clinical diagnostic settings. Nonetheless, serological methods prove time-consuming and impractical for extensive screening of blood donors. However, automation for serology typing may overcome this issue. Additionally, the acquisition of specific antiserum from immunized patients, donors, or mice incurs substantial costs. Furthermore, variations in serological activity may arise among reagents sourced from different manufacturers. In addition, the accurate classification of blood types in patients who have recently undergone blood transfusions and those who have received multiple transfusions poses a challenge when using serological methods [29,30]. Presently, there is a lack of established techniques for genotyping very complex variant alleles in the MNS blood groups. The routine and precise identification of the MNS blood group through serological methods is hindered by the absence of commercially available antibodies capable of typing the low-frequency antigens of hybrid glycophorins. And the discovery of new hybrid glycophorins is a time-consuming, intricate, and arduous process. The molecular typing is getting increasingly important in this instance.

To date, a range of molecular techniques have been employed to address the intricacy of hybrid glycophorins [31-33]. These techniques offer the advantages of automation and multiplexing, enabling the identification of both uncommon blood types and novel alleles [34]. For the detection of GP.Mur, screening can be conducted using sequence-specific primer PCR (SSP-PCR) or melting-curve analysis (HRM), followed by confirmation through Sanger sequencing of exon 2-4 of *GYPA* and *GYPB*. The difficulty in distinguishing *GYPB* from *GYPA N+* arises from the presence of a region in *GYPB* that exhibits significant sequence homology to the *GYPA N* allele. Consequently, the accuracy of predicting the N+ antigen using next-generation sequencing (NGS) and SNV geno-

Table 4
GYPA intronic SNPs strongly linked with M+ allele or N+ allele.

Position	gDNA	Location	cDNA	Ref	Linked with N+
4:144117474	28278	Intron4	c.272-535	A > G	98.59%
4:144114635	31117	Intron6	c.436+54	C > T	98.59%
4:144109349	25831	Intron2	c.137-140	CTT > C	97.18%
4:144119384	26368	Intron3	c.232+302	C > T	94.37%
4:144119379	26372	Intron3	c.232+306	AG > A	94.37%
4:144112490	33262	Intron6	c.437-1080	T > C	95.77%
4:144120192	25560	Intron2	c.136+298	G > A	95.77%
4:144120174	25578	Intron2	c.136+316	C > A	95.77%
4:144119066	26686	Intron3	c.233-314	G > A	94.37%
4:144120597	25155	Intron1	c.38-9	T > G	100%
4:144119883	25869	Intron2	c.137-102	C > T	100%
4:144119906	25892	Intron2	c.137-79	AG > A	100%
4:144119846	25906	Intron2	c.137-65	A > G	100%
4:144119631	26121	Intron3	c.232+55	A > G	100%
4:144117879	27873	Intron4	c.271+835	C > T	98.59%
4:144112379	33373	Intron6	c.437-969	T > C	98.59%

Table 5
GYPB intronic SNPs strongly linked with S+ allele or s+ allele.

Position	gDNA	cDNA	Ref	Linked with s+
4:143997972	26372	c.176-338	G > T	100%
4:143997906	26438	c.176-272	C > T	100%
4:143997740	26604	c.176-106	T > C	100%
4:143997714	26630	c.176-80	T > C	100%
4:143998049	26965	c.270+161	A > ATT	100%
4:143996999	27345	c.270+541	C > G	100%
4:143996886	27457	c.271-583	AC > A	100%
4:143996791	27553	c.271-487	A > T	100%
4:143996792	27554	c.271-486	G > GC	100%
4:143996657	27687	c.271-353	A > G	100%
4:143996636	27708	c.271-332	T > C	100%
4:143996463	27881	c.271-159	A > G	100%
4:143995648	28696	c.276+651	G > A	100%
4:143995467	28877	c.276+832	T > G	100%
4:143995211	29133	c.276+1088	A > G	100%
4:143995193	29151	c.276+1106	T > G	100%
4:143995170	29174	c.276+1129	G > A	100%
4:143995071	29273	c.276+1228	T > G	100%
4:143994873	29471	c.276+1426	C > T	100%
4:143994811	29533	c.276+1488	T > C	100%
4:143994627	29717	c.276+1672	G > A	100%
4:143994375	29969	c.276+1924	A > T	100%
4:143994316	30028	c.276+1983	G > A	100%
4:143996586	27758	c.271-282	C > T	100%
4:143999004	25340	c.175+407	C > G	100.00%
4:143999123	25459	c.175+526	A > AT	100.00%
4:143999748	24596	c.137-299	T > G	100.00%
4:143999732	24612	c.137-283	T > G	100.00%
4:143999485	24859	c.137-36	T > A	100.00%
4:143999443	24901	c.143	C > T	100.00%
4:143999327	25017	c.175+84	T > A	100.00%
4:143999005	25339	c.175+406	C > T	100.00%
4:143998585	25759	c.175+826	C > T	100.00%
4:143998542	25802	c.175+869	C > T	100.00%
4:143998534	25810	c.175+877	T > C	100.00%
4:143998417	25927	c.176-783	G > C	100.00%
4:143994688	29656	c.276+1611	G > A	100.00%

type is relatively low, at 86.1% and 85.6%, respectively [18]. However, advancements in sequencing throughput and precision have enabled LRS technology to generate continuous sequences spanning tens of thousands to several megabases. The increasing maturity of this technology offers a promising approach for analyzing complex repetitive regions in genome sequences and assembling high-quality genomes. Fichou et al. employed the LRS technology to investigate and characterize the predominant haplotypes in the three primary alleles (*01, *02, and *02N.01) of the ACKR1 gene, which encodes the Duffy blood group antigens. They identified a subset of reference alleles/haplotypes in the ACKR1 gene through LRS analysis [20]. This approach, utilizing LR haplotype analysis, has been previously demonstrated to be effective [35,36]. In our study, we also utilized LRS to examine the core region of the MNS blood group system. The LRS technology effectively distinguished the M, N, S, and s blood groups based on specific genetic variations in *GYPA* (*GYPA's* c.59, c.71, c.72) and *GYPB* (*GYPB's* c.143), which aligned with traditional serological methods. This suggests that the LRS method is capable of accurately identifying the predominant genotypes of MNS. Additionally, our study also identified a rare GYP.Mur subtype (GYP B1-136-B*137-212-A213-229-B230-366) using the LRS technology, which was also reported by Wei et al. using the MLPA technique [10]. These findings demonstrate the comparable accuracy and reliability of the LRS technology. Furthermore, the investigation also detected 3 haplotypes containing GYP.Mur, 3 haplotypes containing the St^a antigen, 1 haplotype containing the *GYP(B-E-B)* hybrid in *GYPB* intron 3, and 1 allele containing the *GYP(A-E-A)* hybrid in *GYPA* intron 3 through the examination of random donor samples. Although hybrid glycophorins such as *GYP*Vw (A-B-A), GYP*Bun (B-A-B), GYP*Hil (A-B), GYP*Zan* (A-A, del exon 3) were not included or validated in this study, it is anticipated that this method can effectively discriminate and identify all MNS antigens. Furthermore, the 7 kb regions of the first exon of *GYPA*, *GYPB*, and *GYPE*, devoid of any exon variants, contain a limited amount of informative data. Consequently, in order to minimize sequencing expenses, subsequent extensive clinical RBC antigen typing on a large scale focuses solely on sequencing the core region of *GYPA* and *GYPB*. Additionally, given the restricted sample size of this retrospective study, a comprehensive examination of the benefits of MNS blood typing using LRS necessitates the implementation of a large-scale, multi-center prospective cohort study in future research endeavors.

Conclusion

This study aims to develop and implement a comprehensive analysis platform for the identification of MNS Blood Group using LRS techniques on a large scale. The findings of this study demonstrate that the utilization of LRS technology holds promise for the standardization of MNS genotyping in future research endeavors. By employing a comprehensive sequence of the core region of *GYPA* and *GYPB*, the LRS technology exhibits the capability to accurately and reliably determine all MNS genotypes in a single assay, presenting a novel avenue for MNS blood group typing. The implementation of LRS technology has facilitated the identification of novel alleles and gene variants, thereby contributing new perspectives on the genetic mechanisms underlying glycophorin conversion and unequal crossing-over.

Please cite this article as: H. Xu, X. Su, Q. Zuo et al., The Research of a Large-Scale Analysis Platform for MNS Blood Group Identification Based on Long-Read Sequencing, Transfusion Medicine Reviews, https://doi.org/10.1016/j.tmrv.2024.150836

Ethics Approval and Consent to Participate

The experimental procedures were conducted in line with the principles of the World Health Organization and the Declaration of Helsinki. The study was approved by the Medical Ethics Committee of Shaanxi Blood Center, in accordance with guidelines of the Nation Health and Medical Research Council of China.

Author Contributions

Hua Xu and Qinqin Zuo conceived and designed the study; Liangzi Zhang and Xiaoyue Chu performed the study and analyzed data; Hua Xu and Xiaomin Su wrote the manuscript; Hua Xu, Liangzi Zhang, and Qinqin Zuo critically reviewed and edited the manuscript; and all authors interpreted the data and approved the final draft.

Declaration of competing interest

The authors declare no conflict of interest.

Funding

This work was supported by the Xi'an Science and Technology Plan Project (21YXYJ0085, 22YXYJ0068) and Shaanxi Provincial Health Research Fund (2022D050).

Acknowledgments

The authors thank the anonymous reviewers for carefully reviewing the manuscript. The study was approved by the Ethics Committee of the Shaanxi Provincial Blood Center.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.tmrv.2024.150836.

References

[1] Lopez GH, Hyland CA, Flower RL. Glycophorins and the MNS blood group system: a narrative review. Ann Blood 2021;6(39):1–16.
[2] Blumenfeld OO, Huang CH. Molecular genetics of glycophorin MNS variants. Transfus Clin Biol 1997;4(4):357–65.
[3] Dahr W. Miltenberger subsystem of the MNSs blood group system: review and outlook. Vox Sanguin 1992;62(3):129–35.
[4] Hassan SN, Ponnuraj KT, Mohamad S, Hassan R, Rahman WSWA. Molecular detection of glycophorins A and B variant phenotypes and their clinical relevance. Transfus Med Rev 2019;33(2):118–24.
[5] Willemetz A, Nataf J, Thonier V, Peyrard T, Arnaud L. Gene conversion events between *GYPB* and *GYPE* abolish expression of the S and s blood group antigens. Vox Sang 2015;108(4):410–16.
[6] Chandanayingyong D, Pejrachandra S. Studies on the Miltenberger complex frequency in Thailand and family studies. Vox Sang 1975;28(2):152–5.
[7] Yang CA, Lin JA, Chang CW, Wu KH, Yeh SP, Ho CM, et al. Selection of GP. Mur antigen-negative RBC for blood recipients with anti-"Mi^a" records decreases transfusion reaction rates in Taiwan. Transfus Med 2016;26(5):349–54.
[8] Huynh NT, Ford DS, Duyen TT, Huong MT. Jk and Mi.III phenotype frequencies in North Vietnam. Immunohematology 2003;19(2):57–8.
[9] Prathiba R, Lopez CG, Usin FM. The prevalence of GP Mur and anti-"Mi^a" in a tertiary hospital in Peninsula Malaysia. Malays J Pathol 2002;24(2):95–8.
[10] Wei L, Shan ZG, Flower RL, Wang Z, Wen JZ, Luo GP, et al. The distribution of MNS hybrid glycophorins with Mur antigen expression in Chinese donors including identification of a novel GYP. Bun allele. Vox Sang 2016;111(3):308–14.
[11] Mallari RA, Chan A, Powers RJ, Pandipati S, Bensing KM, Biese D, et al. Fetal inheritance of GP*Mur causing severe HDFN in an unrecognized case of maternal alloimmunization. Transfusion (Paris) 2020;60(4):870–4.
[12] Li S, Wei L, Fang Q, Ji Y, Gao Y. Successful prenatal management of two foetuses affected by antibodies against GP.Mur with prenatal genotyping analysis and a literature review. Blood Transfus 2021;19(2):135–43.
[13] Daniels G. MNS blood group system. Human blood groups. Oxford: Blackwell Science; 2012. p. 125–9.
[14] Palacajornsuk P, Nathalang O, Tantimavanich S, Bejrachandra S, Reid ME. Detection of MNS hybrid molecules in the Thai population using PCR-SSP technique. Transfus Med 2007;17(3):169–74.
[15] Wei L, Sun AN, Wen JZ, Wang Z, Li Q, Liao YT, et al. Molecular genetic analysis of Mi^a-positive hybrid glycophorins revealed two novel alleles of GP. Vw and multiple variant transcripts of GYPB existing in both the homozygous GP. Mur and wild-type GPB individuals. Transfusion (Paris) 2021;61(8):2477–86.
[16] Wei L, Lopez GH, Zhang Y, Wen JZ, Wang Z, Fu YS, et al. Genotyping analysis of MNS blood group GP (B-A-B) hybrid glycophorins in the Chinese Southern Han population using a high-resolution melting assay. Transfusion (Paris) 2018;58(7):1763–71.
[17] Wei L, Ji YL, Mo CY, Zhang RQ, Zhao Y, Luo H, et al. Anti-Mur screening and Mur antigen genotyping of voluntary blood donors in Guangzhou. Nan Fang Yi Ke Da Xue Xue Bao 2012;32(12):1833–5.
[18] Wheeler M, Frazar C, Lannert K, Fletcher SN, Huston H, Harris S, et al. Prediction of MNS blood group antigens using next generation sequencing. Blood 2016;128(22):1458.
[19] Zhang Z, An HH, Vege S, Hu TS, Zhang SP, 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(1):180–91.
[20] 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(1):23–31.
[21] Mohamed S, Muna I. Characterisation of rh and other blood group systems amongst the Maldivian blood donors. Med J Malaysia 2013;68(5):393–6.
[22] Chen JY, Huang HH, Yu SY, Wu SJ, Kannagi RJ, Khoo KH. Concerted mass apectrometry-laned glycomic approach for preciaion mapping of aulfo sialylated N-glycane on human peripheral blood mononu-clear cells and lymphocytes. Glycobiology 2018;28(1):9–20.
[23] Fu XL, Cai XQ, Zhao XD, Weng AH, Xia L. Gene frequency of MNS and Mur blood group system in Li ethnic group in Hainan province: distribution and comparison with other ethnic groups. Chin J Public Health 2022;38(6):775–8.
[24] Yasuda H, Ohto H, Nollet KE, Kawabata K, Saito S, Yagi Y, et al. Hemolytic disease of the fetus and newborn with late-onset anemia due to anti-M: a case report and review of the Japanese literature. Transfus Med Rev 2014;28(1):1–6.
[25] Sun AN, Duan SB, Yi F, Liao YT, Ding SH, Zeng JF, et al. The investigation and application of Mur blood group antigen and antibody. Chin J Blood Transfusion 2017;30(6):586–9.
[26] Lin X, Rubio G, Patel J, Banerjee S, Yang JC. Hybrid glycophorin and red blood cell antigen genotyping in Asian American type O blood donors with Mi^a phenotype. Transfusion (Paris) 2019;59(12):3767–75.
[27] Mallari RA, Chan A, Powers RJ, Pandipati S, Denomme GA. Fetal inheritance of GP*Mur causing severe HDFN in an unrecognized case of maternal alloimmunization. Transfusion (Paris) 2020;60(4):870–4.
[28] Borys S, Hohendorff J, Frankfurter C, Kiec-Wilk B, Malecki MT. Negative pressure wound therapy use in diabetic foot syndrome—from mechanisms of action to clinical practice. Eur J Clin Invest 2019;49(4):e13067.
[29] Reid ME, Rios M, Powell VI, Charles-Pierre D, Malavade V. DNA from blood samples can be used to genotype patients who have recently received a transfusion. Transfusion (Paris) 2000;40(1):48–53.
[30] Castilho L, Rios M, Bianco C, Pellegrino J, Alberto FL, Saad STO, et al. DNA-based typing of blood groups for the management of multiply-transfused sickle cell disease patients. Transfusion (Paris) 2002;42(2):232–8.
[31] Sell AM, Visentainer JE. Blood grouping based on PCR methods and agarose gel electrophoresis. Methods Mol Biol 2015;1310:37–49.
[32] Prager M. Molecular genetic blood group typing by the use of PCR-SSP technique. Transfusion (Paris) 2007;47(1 suppl):54S–59S.
[33] Nathalang O, Intharanut K, Siriphanthong K, Nathalang S, Kupatawintu P. Duffy blood group genotyping in Thai blood donors. Ann Lab Med 2015;35(6):618–23.
[34] Liu Z, Liu M, Mercado T, Illoh O, Davey R. Extended blood group molecular typing and next-generation sequencing. Transfus Med Rev 2014;28(4):177–86.
[35] Schmid P, Ravenell KR, Sheldon SL, Flegel WA. DARC alleles and Duffy phenotypes in African Americans. Transfusion (Paris) 2012;52(6):1260–7.
[36] Yin Q, Srivastava K, Gebremedhin A, Makuria AT, Flegel WA. Long-range haplotype analysis of the malaria parasite receptor gene ACKR1 in an East-African population. Hum Genome Var 2018;5(1):26.