Journal of Translational Medicine

RESEARCH

Open Access

An *RHD* gene variant, common in southern Chinese Han, involves a short 688 bp deletion and a long 21.8 kb insertion in the *RHD* exon 10 non-coding region, affecting the accuracy of red cell genotyping

Jizhi Wen[1,2], Fei He[3], Barbera Veldhuisen[4], Chunyan Mo[1,2], Ling Wei[1,2], Shuangshuang Jia[1,2], C. Ellen. van der Schoot[4], Willy Albert Flegel[5*] and Yanli Ji[1,2*]

Abstract

Background Accurate typing of the D antigen is critical to provide guidance for blood transfusion and safe management of Rh-incompatible pregnancies. *RHD* genotyping assays have been widely adopted to improve D variant detection, especially for common *RHD* alleles with missense variants and *RHD-CE-D* hybrids. Few structural variants of the *RHD* gene are known to involve large fragments of non-RH sequences, despite causing phenotype-genotype discrepancies.

Methods We had previously detected reduced copy numbers of *RHD* exon 10 in Chinese Han. Their *RHD* genes were further analyzed using long-read sequencing of *RHD* and whole-genome long-read sequencing using PacBio platform. The population frequency and practical relevance were evaluated, such as effects on D antigen expression and accuracy of routine *RHD* genotyping assays.

Results A novel structural variant of the *RHD* gene, dubbed *RHD-CE-TMEM50A-D*, had distinct features: a short deletion of the non-coding region of *RHD* exon 10 (688 bp del), a large insertion (21.8 kb ins) involving the inversion of non-coding region of *RHCE* exon 10 (174 bp) and exon 7 to intron 2 of *TMEM50A* gene (21,648 bp), and several single nucleotide substitutions. The large delins fragment was firstly found in 9 individuals, further representing 4.1% of random D+ donors (40/982), 2.9% of D−C+/E+ individuals who carry the non-functional *RHD*01N.04* (6/205), and 3.4% of weak/partial D individuals (7/207) who carry *RHD*DFR2* and *RHD*DVI.3*, but 0% in a different cohort of Asian-type DEL individuals (0/205). The large delins fragment located in non-coding region of *RHD* exon 10 of four different alleles did not significantly change the D antigen expression but caused inconclusive results in several routine *RHD* genotyping assays.

*Correspondence:
Willy Albert Flegel
waf@nih.gov
Yanli Ji
jiyanli2013@163.com

Full list of author information is available at the end of the article

Conclusions The novel *RHD-CE-TMEM50A-D* allele represented approximately 1 in 49 chromosomes among the southern Chinese Han population. The variations of these *RHD* alleles involve the replacement of some *RHD*-specific sequences by the corresponding *RHCE*-specific sequences in the non-coding region of *RHD* exon 10, commonly used for *RHD*-specific primers design in genotyping assay. Hence, red cell genotyping assays, if applied to individuals of East Asian heritage, should recognize this relatively common allele to avoid *RHD* allele dropout.

Keywords Deletion-insertion (delins) variation, Red cell genotyping, Immunohematology, Anti-D

Introduction

The Rh blood group system (ISBT 004) is one of the most complex and clinically significant blood group systems. So far, a total of 56 antigens have been identified in the Rh system, which are encoded by the *RHD* and *RHCE* genes, 2 highly homologous and closely linked genes encompassing 10 exons, respectively [1]. The 3' ends of the *RHD* and *RHCE* genes face each other on the short arm of chromosome 1 and are separated by an approximately 30 kb DNA segment containing the *TMEM50A* gene of 7 exons (formerly referred to as *SMP1*) [2]. The open reading frames (ORF) of the *RHD* and *TMEM50A* genes share identical orientation, while the ORF of the *RHCE* gene has the opposite direction. The *RHD* gene, approximately 62.3 kb long, is flanked by 2 highly homologous DNA segments of approximately 9 kb, known as *Rhesus boxes*, upstream and downstream of the *RHD* gene [2].

The D antigen (RH1) of the Rh blood group system is highly immunogenic, and alloimmunization against D antigen may cause hemolytic transfusion reaction (HTR) and severe hemolytic disease of the fetus and newborn (HDFN). Therefore, it is clinically relevant to correctly type blood donors, blood recipients, pregnant women, and fetus to prevent any adverse reaction caused by anti-D immunization.

The routine D typing is performed using serologic hemagglutination tests. In addition to the common RhD positive (D+) and RhD negative (D−) phenotypes, other D variant phenotypes have been described, including D-elute (DEL), weak D and partial D. To date, more than 400 *RHD* gene variants (alleles) have been documented [3] accounting for weak D, partial D, DEL, and D − phenotypes. Serologic D typing can only preliminarily detect a D − or variant phenotype; for most alleles, red cell genotyping assays are required to determine the exact types, if clinically needed.

Based on the definitive *RHD* genotype and previous clinical evidence, appropriate guidance can be provided about clinical transfusion or management of Rh-incompatible pregnancies. For instance, in the US White population, *RHD* genotyping is recommended for the patients with a serologic weak D phenotype [4]. For the patients identified as weak D type 1, 2, and 3 genotypes, D+ red cells transfusion is recommended, and for the pregnant women of these genotypes, Rh immune globulin prophylaxis (RhIG) is unnecessary. In the East Asian population, routine *RHD* genotyping is performed to identify Asian-type DEL in serologically D − patients; patients with Asian-type DEL could safely be transfused with D+ red cells, and pregnant women with Asian-type DEL do not require RhIG [5].

Many molecular techniques based on the polymerase chain reaction (PCR) with hybridization or sequencing, such as quantitative real-time PCR, droplet digital PCR, DNA microarray [6], multiplex ligation-dependent probe amplification (MLPA) [7], and next-generation sequencing [8] have been widely applied for *RHD* genotyping in transfusion as well as noninvasive fetal *RHD* genotyping in perinatal medicine. The primers or probes used in all these assays are designed based on the nucleotide differences between the 2 highly homologous *RHD* and *RHCE* genes. However, there are only 39 to 44 differences in the coding sequences (1,254 bp) of the 10 exons between the 2 genes. Besides, the non-coding regions of *RHD* exon 10 (1,521 bp) and *RHCE* exon 10 (278 bp) are identical for the initial 104 bp; but their sequences further downstream differ at all and, hence, are used to design primers and probes.

Deletion-Insertion (delins) are defined as a sequence change when nucleotides are replaced by other nucleotides, not representing a single nucleotide substitution or an inversion, compared to a reference sequence [9]. By applying long-read sequencing, a few complex *RH* structural variants have been identified, involving large deletions [10, 11] and a large inversion [12]. These *RH* alleles are usually missed by conventional assays for red cell genotyping in routine clinical applications, causing allele dropout and inaccurate *RH* genotyping results. Fortunately, these alleles seem to be rare in populations.

In this study, we applied long-read sequencing methods and analyzed the individuals, previously found to lack a signal for *RHD* exon 10 [13, 14]. A novel *RHD* allele, dubbed *RHD-CE-TMEM50A-D*, was identified being prevalent in the southern Chinese Han population. This common *RHD* allele involved a deletion and long insertion of non-*RH* sequences (delins) located in the non-coding region of *RHD* exon 10. Three novel alleles *RHD*01N.04-CE-TMEM50A-D*, *RHD*DFR2-CE-TMEM50A-D*, and *RHD*DVI.3-CE-TMEM50A-D* also involved the same delins fragment but were rare even in East Asia.

Materials and methods

Study participants

A total of 12 samples with reduced copy number of *RHD* exon 10, including 10 D+ samples with a deduced *RHD*D(1–9)/01* genotype and 2 partial D (DFR) samples with a deduced *RHD*D-CE(4)-D(5–9)/01N.01* genotype, were identified in our previous studies by MLPA analysis [13, 14] in the reference laboratory of the Guangzhou Blood Center, Southern region of China. At that time, we predicted that all or part of *RHD* exon 10 might be deleted or be replaced by the equivalent sequences from the *RHCE* gene [13, 14]. DNA samples from 7 D+ individuals and 2 DFR individuals were available to be re-analyzed in this study.

The study was approved by the Ethics Committee of the Guangzhou Blood Center (approval numbers 2020–057 and 2023–011). Written informed consent was obtained from each participant.

Long-read PacBio sequencing of *RHD* gene and data analysis

DNA samples of D+ ($n=7$) and DFR individuals ($n=2$) available were commissioned for long-read sequencing of the *RHD* gene using PacBio Sequel II platform (Pacific Biosciences; Menlo Park, CA, USA) first based on Single Molecule Real Time (SMRT) sequencing technology (Haorui Genomics; Xi'an, China).

In brief, 7 pairs of primers for long-range PCR were designed to cover the entire region of the *RHD* gene to detect the common *RHD* variants including the single nucleotide variants (SNVs) and in-frame insertions/deletions, and some *RHD-CE-D* hybrid alleles. The length of the amplicons was approximately 11.5 to 13.7 kb (Fig. 1A), in which more than 1 kb fragment overlapped between adjacent amplicons to facilitate assembly of long-range PCR fragments and obtain the complete sequence of 2 *RHD* alleles. *RHD*-specific primers were designed based on the nucleotide differences between the 2 highly homologous *RHD* and *RHCE* genes. Because the D – phenotype in all populations commonly results from the complete deletion of the *RHD* gene (*RHD*01N.01* allele) [7, 13], a special pair of primers was designed to detect *RHD*01N.01*(Fig. 1B). In addition, 3 primer pairs for long-range PCR were designed to cover the complete *TMEM50A* gene sequence (Fig. 1C). Multiplex long-range PCR, long-read PacBio sequencing and data analysis were conducted according to the manufacturer's instructions (Method S1 in the Supplementary Material).

Whole-genome long-read sequencing and data analysis

Among the 9 samples, 1 D+ and 1 DFR individual with the samples of adequate DNA quality were chosen for whole-genome long-read sequencing (PacBio Sequel II System based on SMRT technology; Haorui Genomics). The SMRTbell libraries were prepared from genomic

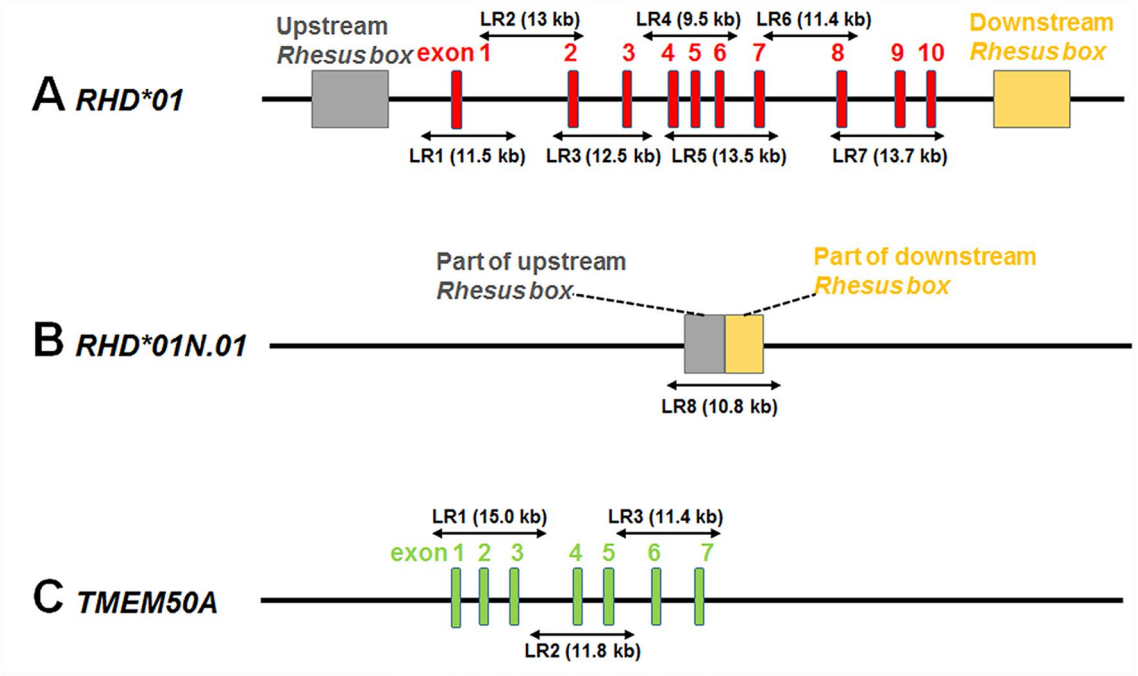

Fig. 1 The long-range amplification of the *RHD* and *TMEM50A* genes. (**A**) The coding sequence (CDS) and upstream and downstream *Rhesus Boxes* of *RHD* gene, position and length of seven long-range PCRs (LR1 to LR7) as well as the overlap between the adjacent amplicons, were showed. (**B**) The position and length of the eighth LR PCRs (LR8) of the *RHD* gene to detect the *RHD*01N.01* allele (*RHD* gene deletion). (**C**) The CDS of the *TMEM50A* gene, position and length of three LR PCRs (LR1 to LR3), as well as the overlap between the adjacent amplicons, were showed

DNA (Procedure and Checklist – 20 kb Template Preparation Using BluePippin Size-Selection System; PacBio) [15].

Briefly, a total amount of 15 μg genomic DNA was sheared by g-TUBEs (Covaris, Massachusetts, USA). Genomic DNA fragments were purified and concentrated (0.45 × AMPure PB magnetic beads; PacBio), and subsequently damage-repaired with ExoVII, end-repaired with End Repair Mix, and ligated with a blunt adaptor. After removing failed ligation fragments with exonuclease (ExoIII and ExoVII), the ligation products were purified (0.45 × AMPure PB magnetic beads), and size-selected (BluePippin Size-Selection System; Sage Science, Beverly, USA). Genomic DNA fragments were further damage-repaired, bead-purified, measured (Agilent 2100 Bioanalyzer; Agilent Technologies, Santa Clara, USA), and used as SMRTbell templates. These SMRTbell templates were annealed with the primers and bound to DNA polymerase (DNA/Polymerase kit; PacBio Biosciences) and magnetic beads, and sequenced (Sequel II platform, PacBio, a licensed medical device approved by the National Medical Products Administration (NMPA) of China, Registration No. 20,233,456,789) with a run time of 30 hours.

Raw sequencing data generated by the Pacific Sequel II platform were processed using the SMRTlink v10.1.0 software following the quality control protocol, by removing low-quality reads and adapters to generate subreads. Regions with coverage below 10x were masked prior to assembly. The subreads were then processed to produce high quality circular consensus sequencing (CCS) using the CCS software. The CCS reads were aligned to the human genome reference version GRCh38 with pbmm2 alignments [16]. Genome-wide coverage was analyzed by Integrative Genomics Viewer [17]. Haplotype-resolved assembly was performed by hifiasm and overlap base pairs between the CCS reads.

Confirmation of 2 breakpoints of the *RHD-CE-TMEM50A-D* allele by Sanger sequencing

To confirm the breakpoints identified by whole-genome long-read sequencing, the 5' and 3' breakpoint regions were amplified and subjected to Sanger sequencing. To confirm the 5' breakpoint, 2 pairs of primers were synthesized in a nested 2-step PCR: 1F/*RHD*-LR6F-intron 8F: 5'-ATACATTCCATCCAGAACTGTTCACC-3'; 1R/*TMEM50A*-intron 6F: 5'-CCCGCTAGGTCTTGTTCCTAC-3'; 2F/*RHD*-exon 10-FL: 5'-GTTGGATTTTAAGCAAAAGCATCCAAGAA-3'; and 2R/*TMEM50A*-exon 7F-SP: 5'-ATATTCCGTGGTCAAAATTCTTCCTCACTAT-3'. To confirm the 3'-breakpoint, a pair of primers was synthesized: 3F/*TMEM50A*-sr45-exon 3R [2]: 5'-CAGCTGCATCTATGATAATCCACC-3'; and 3R/*RHD*-3′ UTR2 [18]: 5'-CAAAGAAAATTGGCAAAACAAGATTT-3'. The PCR conditions were detailed (Method S2 in the Supplementary Material).

Bioinformatics analysis of the sequences of two breakpoint regions

To explore the potential formation mechanism of the novel variation (*RHD-CE-TMEM50A-D*) in the *RHD* gene, the nucleotide sequences of 300 bp upstream and downstream from the 5' breakpoint and 3' breakpoint were analyzed for repetitive elements using an online bioinformatics tool (RepeatMasker) [19].

Screening of the novel *RHD-CE-TMEM50A-D* allele in southern Chinese Han

Given that the D+ blood type (99.6%) is predominant in the Chinese Han population, while D– (0.3%) [20] and weak/partial D (0.014%) [14], and DEL (0.07%) are relatively rare, we randomly collected D+ samples to screen for the novel *RHD-CE-TMEM50A-D* allele. For the D–, weak/partial D, and DEL types, which are extremely rare in random samples, we used the samples collected from previous studies to explore their distribution. The cohort selection criteria were as follows:

For D+ population, peripheral blood samples of 982 randomly selected D+ Chinese Han blood donors were collected at the Guangzhou Blood Center.

For D – population, in Chinese Han, the dominant genetic basis of the "true" D – phenotype is the deletion of the entire *RHD* gene (*RHD*01N.01/01N.01*, approximate 70% of D – and 0.21% in common Chinese Han population) associated with *RHCE*ce.01* [13], so, these D–c–e – individuals with deletion of *RHD* gene are unlikely to carry this *RHD-CE-TMEM50A-D* allele. In difference, many non-functional *RHD* alleles including the common *RHD*01N.03*, *RHD*01N.04* and *RHD*01N.16* are associated with C+/E+ phenotypes (approximate 30% of D – and 0.09% in common Chinese Han population) [13]. Hence, for D – samples, 205 D–C+/E+ pregnant women, but excluding Asian-type DEL, were collected in the reference lab of Guangzhou Blood Center.

For weak/partial D samples, 207 Chinese Han specimens, having the serologic typing and *RHD* genotyping results in previous study [21], were collected for screening.

For Chinese DEL samples, majority of them (99.6%, 539/541) were Asian-type DEL [5]. Therefore, 205 Chinese pregnant women with DEL phenotype and known Asian-type DEL genotype were collected in the reference lab of Guangzhou Blood Center for screening.

Genomic DNAs were extracted from peripheral blood samples (QuickGene DNA whole blood kit S with QuickGene-Mini80 equipment; Fujifilm, Tokyo, Japan). The pair of sequence-specific primers, 3F/*TMEM50A*-sr45-exon 3R [2] and 3R/*RHD*-3′ UTR2 [18] (Method

S2 and Figure S1 in the Supplementary Material), previously established to confirm the 3' breakpoint region, was used to screen the carriers for the novel variation. A 20 µL PCR mixture consisted of 30 ng of genomic DNA, 10 µL 2×GoTaq Colorless Master Mix (Promega, Madison, USA), and 0.3 µM forward and reverse primers. The PCR conditions were: 5 minutes at 94 °C; 35 cycles of 30 seconds at 94 °C, 30 seconds at 52 °C, 3 minutes at 72 °C; final extension at 72 °C for 5 minutes. Carriers of the *RHD-CE-TMEM50A-D* allele would produce an amplicon of 3,053 bp length, while other individual cannot produce any amplicon because the primers are oriented in the same direction.

Evaluation of the impact of the novel *RHD-CE-TMEM50A-D* alleles on genotyping results of several commonly used *RHD* assays

We evaluated the performance of several brands of *RHD* genotyping analysis that were routinely used in China, including Sanger-sequencing [7], *RH*-MLPA (MRC-Holland; Amsterdam, the Netherlands) [13], and PCR-SSP assay (RBC-Ready Gene CDE; inno-train Diagnostik, Kronberg, Germany), that are the assays not licensed. Five samples known to carry the *RHD-CE-TMEM50A-D* were used for evaluation of the impact for accuracy of these assays.

Results

Complex structural variation recognized by long-read sequencing

To determine the exact molecular basis of 7 D+ and the 2 DFR samples lacking 1 expected *RHD* exon 10 signal by MLPA previously [13, 14], long-read sequencing of the *RHD* gene was performed using 7 pairs of primers for long-range PCR covering the entire *RHD* gene (Fig. 1A). Amplification of 6 *RHD* fragments was successful but failed for the last 7th fragment that was straddling exons 8 to 10 of the *RHD* gene. Long-range PCR and PacBio sequencing for 3 amplicons covering the *TMEM50A* gene (Fig. 1C) revealed the expected 2 copies of the 1st *TMEM50A* amplicon but showed unexpected 3 copies of the 2nd and 3rd *TMEM50A* amplicons. Such results were incompatible with any known structure at the *RHD* gene locus.

Identification of a short deletion and large insertion in the non-coding region of the *RHD* exon 10 by whole-genome long-read sequencing

We further characterized 1 D+ (#304) and 1 DFR (Du104) sample by whole-genome long-read sequencing (PacBio Sequel II system). The average sequencing depths of the D+ and DFR sample across the whole genome were 14.12x and 15.20x, respectively. For the target regions (including *RHD*, *TMEM50A*, and *RHCE* genes), the D+ and DFR sample achieved coverage of 99.05% and 96.2% at average depths of ≥30.2x and ≥12.8x, respectively. The CCS read coverage revealed the 3 copies of the *TMEM50A* intron 2 to exon 7 in both samples (Fig. 2). Within the non-coding region of *RHD* exon 10, two segments showed copy number variation (CNV) in sequencing coverage. The reads spanning the breakpoint of these depth changes were randomly selected. Then, Using the "Supplementary Reads Diagram" function in Integrative Genomics Viewer (IGV), it was demonstrated that the 5' and 3' ends of these reads aligned separately to two distinct genomic regions, with opposing orientations, consistent with a duplication and inversion event (Fig. 3).

Finally, the assembled whole-genome sequence confirmed a complex insertion-deletion (delins) event that occurred in the non-coding region of *RHD* exon 10. For

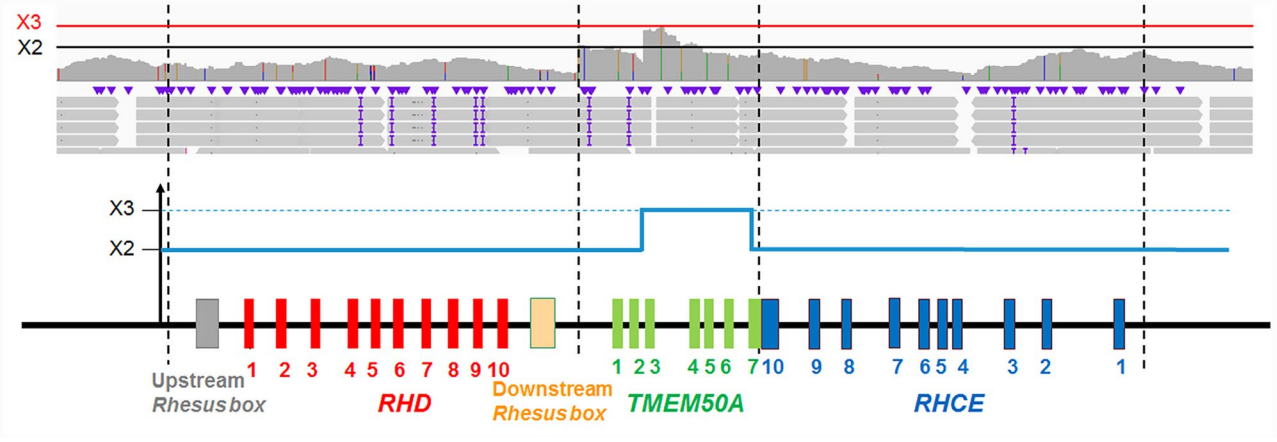

Fig. 2 Circular consensus sequencing read coverage identified in the individuals carrying the novel *RHD-CE-TMEM50A-D* allele. The 10 exons of *RHD* gene, 10 exons of *RHCE* gene, and 7 exons of *TMEM50A* gene were shown as red, blue, and green bars, respectively. The analysis of circular consensus sequencing read coverage revealed three copy numbers of part of the *TMEM50A* gene (intron 2 to exon 7). Two copies and three copies were indicated as X 2 and X 3, respectively

Fig. 3 Duplication and inversion visualized via PacBio whole-genome sequencing in IGV located in *RHD* exon 10. Within the non-coding region of *RHD* exon10, two segments showed copy number variation (CNV) in sequencing coverage. A read spanning the breakpoint of these depth changes was randomly selected, which was shown in red. Using the "Supplementary reads diagram" function in Integrative Genomics viewer (IGV), it was demonstrated that the 5' and 3' ends of this read aligned separately to two distinct genomic regions, with opposing orientations, consistent with a duplication and inversion event

the D+ and DFR individuals, the reference wildtype *RHD* and mutant DFR (*RHD-CE(4)-D)* sequences up to the coding region of exon 10 were identified. But in the non-coding region of *RHD* exon 10 of both allele, a same short 688 bp deletion accompanied with a large 21.8 kb fragment insertion involving the inversion sequences of part of the non-coding region of *RHCE* exon 10 (174 bp) and *TMEM50A* gene (21,648 bp, from intron 2 to exon 7), and several SNVs in the non-coding region of *RHD* exon 10 were detected.

The formal allele designation was NC_000001.11:g.25329029_25329715delins25340601_25362422inv;25329723G > A;25329728A > G;25329730T > C;25329756_25329757delinsTG (Figure S1 in the Supplementary Material). The 22,912 bp nucleotide sequence coving the delins 21.8 kb fragment has been deposited to GenBank (accession number OR661235).

Confirmation of the 2 breakpoint regions of *RHD-CE-TMEM50A-D* allele

The nucleotide sequences of 2 breakpoints obtained by the whole-genome long-read sequencing were further confirmed through amplification using the developed PCR assay and subsequent Sanger sequencing in all 7 D+ and 2 DFR individuals carrying the novel *RHD-CE-TMEM50A-D* allele (Figs. 4F and 4G).

Bioinformatics analysis of repetitive elements in the breakpoint regions

An online bioinformatics tool (RepeatMasker) revealed that the 3' breakpoint junction was located within an *AluSx* sequence in the non-coding region of *RHD* exon 10 (Figure S1 in the Supplementary Material). No such structural element was found at 5' breakpoint junction (not shown).

Prevalence of *RHD-CE-TMEM50A-D* in the southern Chinese Han population

We utilized our PCR that had been developed to confirm the 3' breakpoint region to explore the *RHD-CE-TMEM50A-D* allele prevalence in several group of people with different D phenotypes. As a result, the same delins fragment involving *CE-TMEM50A* was heterozygously identified in 4.1% (40/982) of random D+ donors, 2.9% (6/205) of D−C+/E+ pregnant women who carried the non-functional *RHD*01N.04* allele, 3.4% (7/207) of weak D and partial D individuals, but 0% (0/205) among a different cohort of pregnant women with Asian-type DEL. Hence, 4 different *RHD* alleles carrying the same delins fragment were identified (Fig. 4). And the *RHD*01-CE-TMEM50A-D* allele frequency was 2.0% (95% confidence interval (CI) of 1.4%-2.7%) among D+; the *RHD*01N.04-CE-TMEM50A-D* allele was 1.5% (95% CI 0.3%-2.6%) among D−C+E+; *RHD*DFR2-CE-TMEM50A-D* and

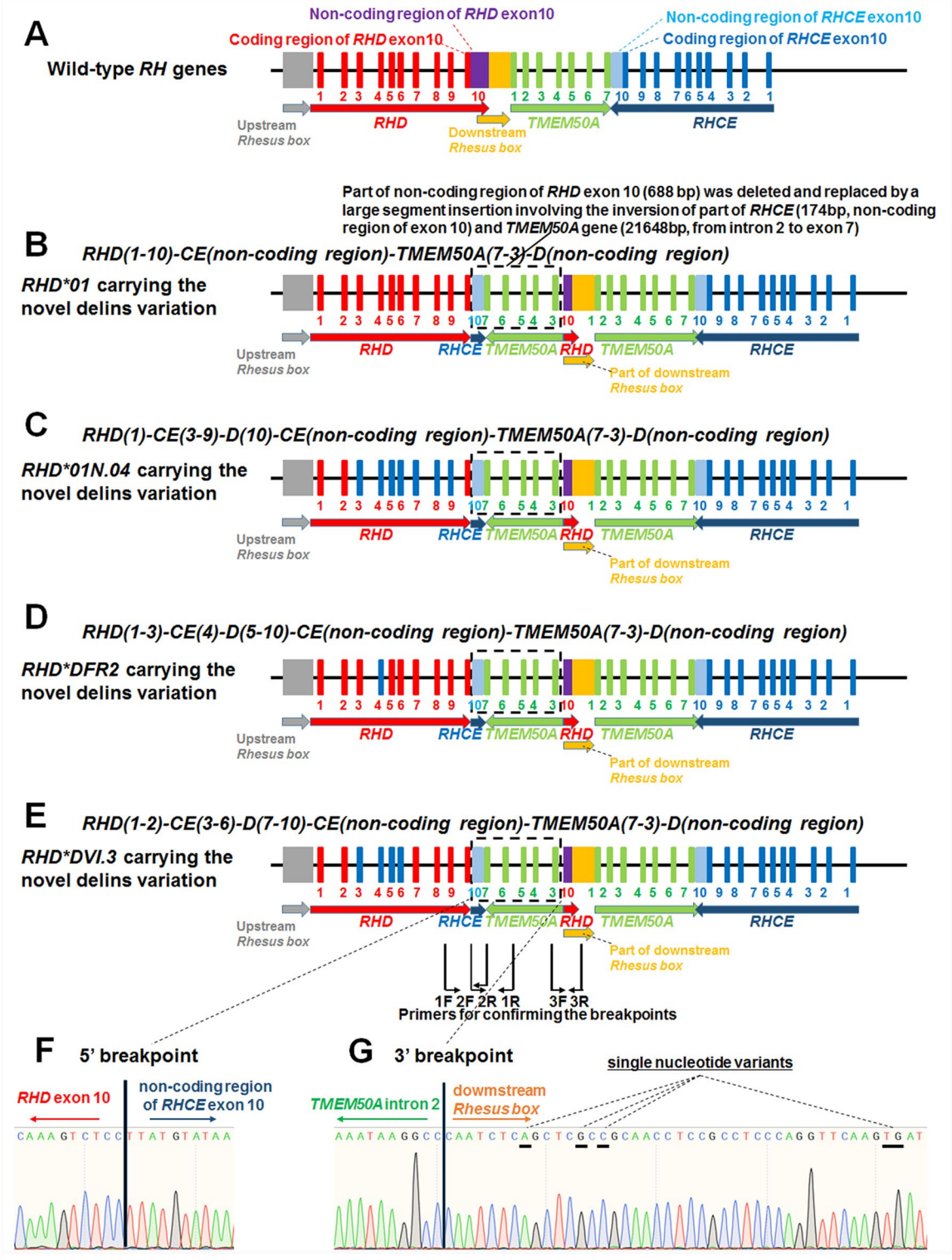

Fig. 4 Genomic structures of the wild-type and four mutant *RH* genes carrying the novel delins fragment, (**A**) Wild-type *RH* genes (shown above) allele. (**B**, **C**, **D**, and **E**) four different types of mutant *RH* genes carrying the novel delins fragment in the non-coding region of *RHD* exon 10 (shown below): a short deletion of non-coding region of *RHD* exon 10 (688 bp) accompanied with a large fragment insertion (21.8 kb) involving the inversion of non-coding region of *RHCE* exon 10 (174 bp) and the exon 7 to intron 2 of *TMEM50A* gene (21,648 bp), and several additional single nucleotide variants. The novel *RHD-CE-TMEM50A-D* alleles resulted in the complete replacement of *RHD* specific non-coding sequence of exon 10 (1,417 bp) by the corresponding *RHCE* specific sequence (174 bp). (**F** and **G**) Sanger sequencing results of 5′ and 3′ breakpoints regions

Table 1 Rh antigens serologic typing and *RHD* genotyping results in 22 Chinese Han individuals carrying the four different types of novel *RHD-CE-TMEM50A-D* alleles

Sample number	Molecular description		Serologic phenotype		Total (n = 22)	GenBank
	***RHD*_ allele 1 ***	***RHD*_allele 2**	**RhD**	**RhCE**		**Accession No**
#304†, MLPA-4, 21, 39, 43, 52, 53, 139	***RHD*01*** with the novel delins *RHD(1–10)-CE(NCR)-TMEM50A(7–3)-D(NCR)*	*RHD*01*	D+	CCee	8	OR661235
G10535, 10,708, 10,794, 10865, 11,208, 11,574	***RHD*01N.04*** with the novel delins *RHD(1–2)-CE(3–9)-D(10)-CE(NCR)-TMEM50A(7–3)-D(NCR)*	*RHD*01N.01*	D–	Ccee (5) CCee (1)	6	Not available‡
Du-104†, 570, 604, 666, 890	***RHD*DFR2*** with the novel delins *RHD(1–3)-CE(4)-D(5–10)-CE(NCR)-TMEM50A(7–3)-D(NCR)*	*RHD*01N.01*	D variant	Ccee	5	PX401959
Du-794, 892	***RHD*DFR2*** with the novel delins *RHD(1–3)-CE(4)-D(5–10)-CE(NCR)-TMEM50A(7–3)-D(NCR)*	*RHD*DEL1*	D variant	CCee	2	
Du-821	***RHD*DVI.3*** with the novel delins *RHD(1–2)-CE(3–6)-D(7–10)-CE(NCR)-TMEM50A(7–3)-D(NCR)*	*RHD*01N.01*	D variant	Ccee	1	PX401958

* The nucleotide sequence of this novel delins fragment (22,912 bp) in the non-coding region of *RHD* exon 10 has been deposited in GenBank (accession number OR661235). The complete nucleotide sequences of *RHD*DFR2* and *RHD*DVI.3* involved the novel delins fragment have been deposited in GenBank (accession number PX401959 and PX401958), respectively

† #304 and Du-104 were analyzed by whole-genome long-read sequencing

‡ The complete nucleotide sequence of *RHD* gene including *RHD*01N.04* with the novel delins fragment was not available

NCR - non-coding region

*RHD*DVI.3-CE-TMEM50A-D* alleles were 1.7% (95% CI 0.4%-2.9%) among weak D and partial D, and 0% (95% CI 0%-0.8%) was in Asian-type DEL cohorts.

Based on the observed carrier frequencies in each cohort and their population proportions in Chinese Han (D + 99.6%; D–C+/E+ 0.09% [20]; weak/partial D 0.014% [14]; and Asian-type DEL 0.07% [22]), the corrected and estimated allele frequency of the novel *RHD-CE-TMEM50A-D* allele approximates 2.03%, corresponding to 1 in 49 chromosomes in the southern Chinese Han individuals.

The delins large fragment in the non-coding region of *RHD* exon 10 showed no marked impact on the original D antigen expression

From the screened samples, we selected 20 samples for further genetic and D phenotypic analysis to investigate whether the genetic large delins fragment located in the non-coding region of *RHD* exon 10 have some impact on the expression of the D antigen (Table 1). Four different haplotypes with *RHD*01*, *RHD*01N.04*, *RHD*DFR2*, and *RHD*DVI.3*, up to *RHD* exon 10 followed by the novel short deletion and 21.8 kb insertion with *CE-TMEM50A* sequences within the non-coding region of *RHD* exon 10 were identified. In the carriers with the *RHD*01N.04-CE-TMEM50A-D/RHD*01N.01* ($n = 6$), *RHD*DFR2-CE-TMEM50A-D/RHD*01N.01* ($n = 5$), and *RHD*DVI.3-CE-TMEM50A-D/RHD*01N.01* ($n = 1$) genotypes, the similar true D-negative and partial D (DFR and DVI) phenotypes were observed respectively, compared with the D-negative, DFR ($n = 3$), DVI ($n = 2$) controls carried *RHD* gene lacking of the novel delins fragment (Additional file: Table S1). The complete sequences of *RHD*DFR2-CE-TMEM50A-D* and *RHD*DVI.3-CE-TMEM50A-D* alleles have also been submitted to GenBank with accession numbers PX401959 and PX401958, respectively.

RHD-CE-TMEM50A-D and the accuracy of *RHD* genotyping

We evaluated the effect of *RHD-CE-TMEM50A-D* on the accuracy of *RHD* genotyping through tested several methods and assays commonly used by laboratories in China, including *RHD* Sanger sequencing, *RH*-MLPA, and PCR-SSP assays. For *RHD* Sanger sequencing and PCR-SSP assays, the amplification and detection of the *RHD* exon 10 failed (Figure S2 in the Supplementary Material). For *RH*-MLPA assays, the signal of 1 copy of *RHD* exon 10 was missing in *RHD-CE-TMEM50A-D* carriers (Table S2 in the Supplementary Material) as described previously [7, 13].

Discussion

Complex structural variant alleles of the *RHD* gene have rarely been reported [10, 11, 18, 23, 24] compared with common missense variants, a few nucleotides insertion or deletions of the *RHD* gene, or *RHD-CE-D* hybrid alleles. Only a few alleles with a large insertion of duplicated *RHD* sequence [24] and several large deletions of *RHD* sequence [10, 11, 18, 23] have been documented. Here we describe a short deletion and large fragment insertion event to form a variant *RH* haplotype, dubbed *RHD-CE-TMEM50A-D*.

In this study, we applied whole-genome long-read sequencing to resolve the ambiguous *RHD* genotyping results found previously [13, 14]. All individuals carried a variant *RH* haplotype caused by complex delins events. The chromosomal structure of the novel *RHD-CE-TMEM50A-D* allele comprised a short deletion of

the non-coding region of *RHD* exon 10 (688 bp) accompanied by a large fragment insertion (21.8 kb, involving the inversion of non-coding region of *RHCE* exon 10 (174 bp) and intron 2 to exon 7 of *TMEM50A* gene (21,648 bp) with several additional SNVs.

Complex structural variants are difficult to identify and characterize by traditional genotyping methods. Our previous MLPA analysis with ability for gene copy number analysis revealed several samples lacking 1 copy of the expected signals for *RHD* exon 10 [13, 14]. Long-read sequencing of the *RHD* and *TMEM50A* genes suggested a novel complex *RHD* structure, while the exact molecular basis still remained elusive. Only whole-genome long-read sequencing was able to clarify the sequence of this complex *RHD* structure. Whole-genome long-read sequencing, although still expensive, is capable of sequencing the single-molecule DNA with an average read length of more than 10 kb, fully enabling the coverage of large structural and/or repetitive sequences, providing a novel approach to solve the genotyping discrepancy in complex cases [25, 26]. Recently, a complex structural variation located among the *RH* genes (*RHCE*Ce(1–2)-D(3–10)-TMEM50A-Ce(10–8)-Ce(3–10)*) associated with the rare D – phenotype has also been resolved by whole-genome long-read sequencing [12].

The novel *RHD-CE-TMEM50A-D* allele was detected in D+, D–C+/E+, and weak/partial D individuals. After adjusting for the population frequencies of D+ (99.6%), D–C+/E+ (0.09%), and D variants (0.014%), the allele's estimated frequency was ascertained as 2.03% in the southern Chinese Han population investigated. In comparison, our retrospective analysis of MLPA data from 3,108 Dutch individuals with an ambiguous D phenotype identified only 1 case (1/3,108) with abnormal reduced copy number of *RHD* exon 10 (unpublished data). Further investigation may reveal its molecular basis. However, preliminary data indicates that *RHD-CE-TMEM50A-D* is rare in the White population from the Netherlands compared with Chinese Han from Guangzhou, Southern China. To determine the frequency of *RHD-CE-TMEM50A-D* in the general Chinese Han population, further studies with larger sample sizes in other regions of China are needed in the future.

Further serologic and genotyping analysis were performed in the novel *RHD-CE-TMEM50A-D* carriers with different D phenotypes, and the results showed that this large delins fragment located in the non-coding region (NCR) of *RHD* exon 10 did not have marked effect on the original D antigen expression. Firstly, in the D+ donors screening, unfortunately, no homozygous carriers of this novel *RHD-CE-TMEM50A-D* were detected (0/982). However, secondly, in the carriers with true D– phenotype and *RHD(1–2)-CE(3–9)-D(10)-CE(NCR)-TMEM50A(7–3)-D(NCR)/01N.01* ($n=6$) genotype, the same true D– phenotype was detected (Table 1 and Table S1 of Additional file). Thirdly, in the DFR and DVI.3 carriers, *RHD(1–3)-CE(4)-D(5–10)-CE(NCR)-TMEM50A(7–3)-D(NCR)/RHD*01N.01* ($n=5$), *RHD(1–2)-CE(3–6)-D(7–10)-CE(NCR)-TMEM50A(7–3)-D(NCR)/RHD*01N.01* ($n=1$) genotypes were obtained. Similar D epitope serologic typing results (using D-Screen) were identified in these carriers compared with the DFR and DVI.3 controls lacking of the novel delins fragment (Additional file: Table S1) [14]. Although the *RHD-CE-TMEM50A-D* allele was not shown to significantly alter D antigen expression, it should result in the replacement of the *RHD* 3'-UTR with the *RHCE* 3'-UTR at the mRNA level. Consequently, this allele may exert subtle regulatory effects on D antigen expression, potentially through miRNA-mediated post-transcriptional regulation.

Bioinformatics analysis for repetitive elements identified an *AluSx* sequence located in the 3' breakpoint junction of *RHD-CE-TMEM50A-D*. Interestingly, the breakpoints of 3 reported large *RHD* deletions [10], as well as hybrid *RHD-CE-D* alleles [27–29] including *RHD*DVI.3* (*RHD-CE(3–6)-D*) and *RHD*DFR* (*RHD-CE(4)-D*) described before, were also located within or very close to *Alu* repeats, which are located in several introns of *RHD* and non-coding region of the exon 10. They may be a hotspot for genomic rearrangements within the *RH* genes.

Based on our assessment of the impact of the novel *RHD-CE-TMEM50A-D* allele on the accuracy of several *RHD* genotyping assays, we propose that any *RHD* genotyping assay employing *RHD*-specific primers or probes targeting the non-coding region of *RHD* exon 10 may yield biased results due to the failure of primers and probes to bind to the *RHD* gene. This region is frequently targeted for the design of specific primers and probes in *RHD* and *RHCE* genotyping assays due to existence of large different sequences between two homologous genes. *RHD-CE-TMEM50A-D* has a complete replacement of *RHD* specific non-coding sequence of exon 10 by the corresponding *RHCE* specific non-coding sequence, thereby hampering the annealing of these *RHD*-specific probes or primers and causing routine assays to fail. This explains the dropout of the *RHD* exon 10 signal in MLPA and PCR-SSP assays and the failure to amplify *RHD* exon 10 in both Sanger and long-range sequencing. Therefore, as genotyping assays continue to develop and be widely applied in transfusion medicine, it is critical to comprehensively consider the genetic diversity across different ethnic populations, including this common *RHD* structural variation distributed in the Southern Chinese population, to avoid allele dropout and improve genotyping accuracy.

Conclusions

The novel *RHD-CE-TMEM50A-D* allele was characterized by a short deletion and a large fragment insertion (delins), which replaced the *RHD*-specific non-coding sequence of exon 10 (1,417 bp) by the corresponding *RHCE* specific non-coding sequence (174 bp). This non-coding exon 10 sequence is often used in *RHD* genotyping. The delins replacement prevents the annealing of *RHD*-specific primers and probes causing allele dropout in many routine clinical assays for red cell genotyping. Importantly, *RHD-CE-TMEM50A-D* is relatively common in the Southern Chinese Han population with an allele frequency of 2.03%. Hence, this novel allele should be considered in the design of red cell genotyping assays to avoid clinically relevant *RHD* genotyping errors if applied to individuals of East Asian background.

Abbreviations

HTR	Hemolytic transfusion reaction
HDFN	Hemolytic disease of the fetus and newborn
DEL	D-eluate
RhIG	Rh immune globulin prophylaxis
MLPA	Multiplex ligation-dependent probe amplification
SMRT	Single molecule real time
SNV	Single nucleotide variant
CCS	Circular consensus sequencing
IGV	Integrative Genomics Viewer
NCR	Non-coding region

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12967-026-08183-1.

Supplementary material 1

Acknowledgements

We thank the individuals and donors for their participation in this study.

Author contributions

JZW performed the experiments and wrote the manuscript; FH, BV, CYM, LW, and SSJ collected samples and analyzed the sequencing data; CES, WAF, and YLJ designed the study; CES edited the manuscript; WAF and YLJ revised the manuscript extensively and wrote the final version.

Funding

This work was supported by Guangdong Basic and Applied Basic Research Foundation (No. 2024A1515012995), the National Natural Science Foundation of China (Grant 82470233), the Key Medical Disciplines and Specialties Program of Guangzhou (2025-2027), and the Intramural Research Program of the NIH Clinical Center (ZIC CL002128). Non-US centers did not receive any funding from the NIH or other U.S. sources for this study.

Data availability

The datasets generated during the current study are not publicly available due to privacy and confidentiality reasons but are available from the corresponding author on reasonable request.

Declarations

Ethics approval and consent to participate

The study was approved by the Ethics Committee of the Guangzhou Blood Center (approval numbers 2020-057 and 2023-011). All participants provided written informed consent prior to participation in accordance with the Declaration of Helsinki.

Consent for publication

All authors read the final version of the manuscript and gave consent for publication

Statement of Disclaimer

This research was supported in part by the Intramural Research Program of the National Institutes of Health (NIH). The contributions of the NIH author are considered Works of the United States Government. The findings and conclusions presented in this paper are those of the authors and do not necessarily reflect the views of the NIH or the U.S. Department of Health and Human Services

Competing interests

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Author details

[1]Institute of Blood Transfusion and Hematology, Guangzhou Blood Center, Guangzhou Medical University, No31, Luyuan Road, Yuexiu District, Guangzhou 510095, China
[2]The Key Medical Laboratory of Guangzhou, No31, Luyuan Road, Yuexiu District, Guangzhou 510095, China
[3]Department of Blood Transfusion, The First Affiliated Hospital of Guangxi Medical University, Nanning, China
[4]Department of Experimental Immunohematology, Sanquin Research and Landsteiner Laboratory, Academic Medical Center, University of Amsterdam, Amsterdam, The Netherlands
[5]Department of Transfusion Medicine, NIH Clinical Center, National Institutes of Health, 9000 Rockville Pike, Bethesda MD, 20892-1184, USA

Received: 5 March 2026 / Accepted: 16 April 2026
Published online: 12 May 2026

References

1. Daniels G. Human blood groups. 3rd. Oxford, UK: Blackwell Scientific; 2013.
2. Wagner FF, Flegel WA. RHD gene deletion occurred in the Rhesus box. Blood. 2000;95(12):3662–68. https://doi.org/10.1182/blood.V95.12.3662.012k12_3662_3668.
3. Wagner FF, Flegel WA. The Rhesus site. Transfus Med Hemother. 2014;41(5):357–63. https://doi.org/10.1159/000366176.
4. Joint statement on phasing-in RHD genotyping for pregnant women and other females of childbearing potential with a serologic weak D phenotype (PDF) is available from. https://www.aabb.org/docs/default-source/default-document-library/positions/statement150722.pdf?sfvrsn=f4f44c30_6. 11 March 2025.
5. Ji Y, Luo Y, Wen J, Sun Y, Jia S, Ou C, et al. Patients with Asian-type DEL can safely be transfused using RhD-positive blood. Blood. 2023;141:2141–50. https://doi.org/10.1182/blood.2022018152.
6. Barriteau CM, Lindholm PF, Hartman K, Sumugod RD, Ramsey G. RHD genotyping to resolve weak and discrepant RhD patient phenotypes. Transfusion. 2022;62(11):2194–99. https://doi.org/10.1111/trf.17145.
7. Haer-Wigman L, Veldhuisen B, Jonkers R, Loden M, Madgett TE, Avent ND, et al. RHD and RHCE variant and zygosity genotyping via multiplex ligation–dependent probe amplification. Transfusion. 2013;53(7):1559–74. https://doi.org/10.1111/j.1537-2995.2012.03919.x.
8. Stef M, Fennell K, Apraiz I, Arteta D, Gonzalez C, Nogues N, et al. RH genotyping by nonspecific quantitative next-generation sequencing. Transfusion. 2020;60(11):2691–701. https://doi.org/10.1111/trf.16034.
9. Hart RK, Fokkema I, DiStefano M, Hastings R, Laros JFJ, Taylor R, et al. HGVS Nomenclature 2024: improvements to community engagement, usability, and computability. Genome Med. 2024;16(1):149. https://doi.org/10.1186/s13073-024-01421-5.
10. Srivastava K, Stiles DA, Wagner FF, Flegel WA. Two large deletions extending beyond either end of the RHD gene and their red cell phenotypes. J Hum Genet. 2018;63(1):27–35. https://doi.org/10.1038/s10038-017-0345-3.
11. Lopez GH, Turner RM, McGowan EC, Schoeman EM, Scott SA, O'Brien H, et al. A DEL phenotype attributed to RHD exon 9 sequence deletion: slipped-strand mispairing and blood group polymorphisms. Transfusion. 2018;58(3):685–91. https://doi.org/10.1111/trf.14439.

12. Li M, Wang L, Li A, Wang B, Yang X, Zhang Y, et al. Integrated analyses reveal unexpected complex inversion and recombination in RH genes. Blood Adv. 2024;8(12):3154–65. https://doi.org/10.1182/bloodadvances.2023012147.
13. Ji YL, Luo H, Wen JZ, Haer-Wigman L, Veldhuisen B, Wei L, et al. RHD genotype and zygosity analysis in the Chinese southern Han D+, D– and D variant donors using the multiplex ligation-dependent probe amplification assay. Vox Sang. 2017;112(7):660–70. https://doi.org/10.1111/vox.12554.
14. Wen J, Jia S, Wang Z, Chen J, Liang Q, Wei L, et al. Molecular and serological analysis of the D variant in the Chinese population and identification of seven novel RHD alleles. Transfusion. 2023;63(2):402–14. https://doi.org/10.1111/trf.17186.
15. Procedure and Checklist - 20 kb Template Preparation using BluePippin™ Size-Selection System. https://www.pacb.com/wp-content/uploads/2015/09/Procedure-Checklist-20-kb-Template-Preparation-Using-BluePippin-Size-Selection.pdf. 11 March 2025.
16. Pbmm2. https://github.com/PacificBiosciences/pbmm2. 11 March 2025.
17. Robinson JT, Thorvaldsdottir H, Winckler W, Guttman M, Lander ES, Getz G, et al. Integrative genomics viewer. Nat Biotechnol. 2011;29(1):24–26. https://doi.org/10.1038/nbt.1754.
18. Fichou Y, Chen JM, Le Marechal C, Jamet D, Dupont I, Chuteau C, et al. Weak D caused by a founder deletion in the RHD gene. Transfusion. 2012;52(11):2348–55. https://doi.org/10.1111/j.1537-2995.2012.03606.x.
19. R. http://repeatmasker.org/. 11 March 2025.
20. Peng CT, Shih MC, Liu TC, Lin IL, Jaung SJ, Chang JG. Molecular basis for the RhD negative phenotype in Chinese. Int J Mol Med. 2003;11:515–21. https://doi.org/10.3892/ijmm.11.4.515.
21. Wen J, Ma X, Jia S, Chen J, Wei L, Ji Y. The need for routine detection of the Asian-type DEL allele in individuals with weak or partial D phenotypes from East and Southeast Asian populations. Vox Sang. 2025;120(7):714–22. https://doi.org/10.1111/vox.70048.
22. Flegel WA, Wagner FF. Del. Blood Transfus. 2020;18:159–62.
23. Richard M, Perreault J, Constanzo-Yanez J, Khalife S, St-Louis M. A new DEL variant caused by exon 8 deletion. Transfusion. 2007;47(5):852–57. https://doi.org/10.1111/j.1537-2995.2007.01199.x.
24. Fichou Y, Parchure D, Gogri H, Gopalkrishnan V, Le Marechal C, Chen JM, et al. Molecular basis of weak D expression in the Indian population and report of a novel, predominant variant RHD allele. Transfusion. 2018;58(6):1540–49. https://doi.org/10.1111/trf.14552.
25. Mizuguchi T, Toyota T, Adachi H, Miyake N, Matsumoto N, Miyatake S. Detecting a long insertion variant in SAMD12 by SMRT sequencing: implications of long-read whole-genome sequencing for repeat expansion diseases. J Hum Genet. 2019;64(3):191–97. https://doi.org/10.1038/s10038-018-0551-7.
26. de la Morena-Barrio B, Stephens J, de la Morena-Barrio Me, Stefanucci L, Padilla J, Minano A, et al. Long-read sequencing identifies the First retrotransposon insertion and resolves structural variants causing antithrombin deficiency. Thromb Haemost. 2022;122(08):1369–78. https://doi.org/10.1055/s-0042-1749345.
27. Granier T, Chiaroni J, Bailly P, Silvy M. First description of a D-CE-D hybrid gene on a weak D type 2 molecular background. Transfusion. 2017;57(5):1248–53. https://doi.org/10.1111/trf.14023.
28. Haire RN, Ohta Y, Strong SJ, Litman RT, Liu Y, Prchal JT, et al. Unusual patterns of exon skipping in Bruton tyrosine kinase are associated with mutations involving the intron 17 3′ splice site. Am J Hum Genet. 1997;60(4):798–807.
29. Matassi G, Cherif-Zahar B, Mouro I, Cartron JP. Characterization of the recombination hot spot involved in the genomic rearrangement leading to the hybrid D-CE-D gene in the D(VI) phenotype. Am J Hum Genet. 1997;60(4):808–17.

Publisher's Note