REGULAR ARTICLE

blood advances

Proof-of-principle: nanopore adaptive sampling enables full blood group genome analysis and resolution of hybrid alleles

Morgan Gueuning,[1] Gian Andri Thun,[1] Samuel Koller,[1,2] Sonja Sigurdardottir,[2] Nadine Trost,[2] Lukas Wagner,[3] Beate Mayer,[3] Charlotte Engström,[4] Maja Patricia Mattle-Greminger,[1] and Stefan Meyer[1,2]

[1]Department of Research and Development and [2]Department of Molecular Diagnostics and Cytometry, Blood Transfusion Service Zurich, Swiss Red Cross, Schlieren, Switzerland; [3]Institute for Transfusion Medicine, Charité Universitätsmedizin Berlin, Berlin, Germany; and [4]Department of Immunohematology, Blood Transfusion Service Zurich, Swiss Red Cross, Schlieren, Switzerland

Key Points

- AS enables full-gene blood group analysis and accurate haplotype generation.
- Novel and hybrid alleles were resolved, including the first publicly available full haplotypes of *GYP*401.02*, *RHD*03N.01*, and *RHD*01EL.44.*

Adaptive sampling (AS), a computational enrichment method developed for Oxford Nanopore Technologies sequencing platforms, offers a promising advance in molecular blood group diagnostics. By leveraging long-read sequencing, AS has the potential to accurately resolve complex structural variants in the RH and MNS blood group systems, while characterizing the entire blood group genome through a simple, fast and locus-adjustable protocol. As proof-of-principal, we evaluated the performance of AS using 5 samples with suspected complex variants in the RH and MNS systems, unresolved by standard immunohematological methods. Samples were sequenced on a PromethION P2 Solo with up to 2 samples per flowcell, generating 37.0 to 52.4 gigabases of data with mean on-target coverages of 18.9× to 53.4×, allowing reliable variant detection. Hybrid alleles were characterized using a de novo assembly approach, whereas variants in nonrecombinant regions were analyzed using both a custom in-house and the EPI2ME reference-based workflow. With reference to field-specific allele collections, 10% to 15% of detected alleles contained novel nonsynonymous single-nucleotide variants (SNVs) or unreported exonic SNV combinations. All suspected hybrid alleles were successfully assembled and identified as *GYP*401.02*, *RHD*03N.01*, and *RHD*01EL.44*, representing, to our knowledge, the first fully characterized haplotypes for these variants publicly available. Overall, AS showed significant potential for advancing blood group genomics by enabling high-resolution, full-gene analysis. Its ability to support high-throughput donor genotyping and precise patient-donor matching may reduce the risk of alloimmunization and delayed hemolytic transfusion reactions, particularly in patients who receive chronic transfusions. These findings highlight AS as a powerful tool for both research and clinical applications in transfusion medicine.

Submitted 15 July 2025; accepted 15 October 2025; prepublished online on *Blood Advances* First Edition 10 November 2025. https://doi.org/10.1182/bloodadvances.2025017463.

All haplotype sequences can be obtained from the National Center for Biotechnology Information GenBank sequence database (accession numbers: PX148158-PX148644). A detailed list of sequence accession numbers is provided in supplemental Table 11.

Any other data are available from the corresponding author, Morgan Gueuning (m.gueuning@zhbsd.ch), on request.

The full-text version of this article contains a data supplement.

Introduction

Adaptive sampling (AS) represents a considerable innovation in the realm of high-throughput sequencing. This unique feature compatible with Oxford Nanopore Technologies (ONT) sequencing platforms enables enrichment of user-defined genomic regions through a computational approach.[1,2] As DNA strands pass through nanopores, they are immediately mapped against user-defined reference sequences. If a strand does not align with a target region, it is ejected from the nanopore through a reversal of the ionic current, thereby freeing the pore for a new strand. By dynamically selecting genomic regions to be sequenced, AS optimizes sequencing resources and accelerates read acquisition (ie, sequencing depth) in regions of interests (ROI). In addition to its versatility, the key advantage of this approach is that it eliminates the need for labor-intensive and costly presequencing DNA-enrichment procedures.

AS has already demonstrated considerable potential in diverse clinical applications, including infectious disease detection,[3-5] rare genetic disorder diagnostics,[6,7] and oncology.[8-10] In blood group genetics, AS holds promise for addressing 2 major challenges that have proven difficult to resolve using conventional molecular methods.

First, AS allows targeting all genes relevant for blood group antigen prediction in a resource-efficient manner, while adhering to ethical guidelines by limiting sequencing to blood group loci. Although clinical transfusion practice typically focuses on a limited subset of blood group systems, comprehensive antigen profiling, encompassing the full erythrocytic antigen profile, is especially advantageous for individuals requiring chronic transfusions. Such patients with conditions such as sickle cell disease or thalassemia are at high risk of alloimmunization, that is developing antibodies against mismatched blood group antigens from donor blood.[11-13] On the donor side, high-throughput antigen profiling based on array technology has been proposed as most comprehensive[14,15] and implemented in some large blood transfusion centers.[16] However, such approaches have limitations, in particular for the patient side, because they only rely on known genetic variants, and lack speed and flexibility for single-sample analysis. By enabling high-resolution, fully phased blood group allele determination for all blood groups, AS offers the potential to extend and facilitate donor-recipient matching, better allocate matched blood units, and improve transfusion success rates.[17] This approach is particularly beneficial in populations with diverse genetic backgrounds and, as a consequence, elevated chance of loss of common or formation of uncommon antigens, in which cases identifying antigen-matched donors is especially challenging.[18]

The second challenge that could be addressed by AS is the accurate characterization of alleles in the RH and MNS blood group systems notorious for harboring complex genetic rearrangements due to the underlying paralogous gene systems (*RHD/RHCE* and glycophorin-encoding genes *GYPA/GYPB/GYPE*, respectively). Although resolution of hybrid alleles is rarely required for routine transfusion decisions,[19] their accurate detection remains crucial to avoid delayed hemolytic transfusion reactions.[20-22] Most current methods indirectly infer hybrid alleles by relying on selected exonic variants. AS, however, provides 2 advantages. First, introns are also sequenced allowing breakpoint determination. Second, long reads are adequate for reliable de novo assembly, which allows to bypass error-prone variant calling against reference sequences that considerably deviate from the hybrid alleles. Well-characterized hybrid reference sequences produced by such approaches will become increasingly valuable as high-throughput molecular typing expands in blood group genetics.[14,15,23-27]

In this proof-of-principal paper, we examine the potential of AS in blood group genomics, focusing on its ability to resolve hybrid alleles while accessing the entire blood group genome. We outline the advantages but also discuss current limitations hindering its broader adoption and highlight the necessary advancements in blood group-specific bioinformatics tools and reference data sets to further optimize its use.

Methods

Preanalytics and sample selection

We analyzed 5 samples (S01-S05) with suspected complex variants unresolved by standard molecular or serological methods. Four were blood donors (S01, S02, S04, and S05) from the Zurich blood bank with suspected rearrangements in the RH or MNS systems, and 1 (S03) was a young patient with sickle cell disease from Charité Hospital, Berlin. Written informed consent for molecular blood group analysis was obtained from all donors, whereas it was not required for the patient under German national regulations for blood group genotyping. Ethics committee/institutional review board of Federal Office of Public Health Switzerland waived ethical approval for this work (license no. Lab-070028).

All samples underwent standard serological testing and genotyping based on polymerase chain reaction (PCR) with sequence-specific primers. Each case showed inconsistencies between phenotype and genotype or harbored unexplained antibodies. In short, S01 showed a discrepancy between RhCE phenotype and genotype, which could not be resolved by exon sequencing. S02 had weak MNS2 antigen expression due to the likely presence of a *GYP*401* (*GYP*Sch*) allele. S03, with anti–e-like antibodies, showed an unreported *RHCE* single-nucleotide variant (SNV) combination, not resolvable into haplotypes. Finally, S04 and S05 were suspected to harbor *RHD-CE-D* hybrid alleles based on matrix-assisted laser desorption/ionization time-of-flight (MALDI-TOF) mass spectrometry (MS) allelic ratio analyses. A detailed description of preanalytical methods and results is given in supplemental Information Section 1.1 and supplemental Table 1, respectively.

Nanopore sequencing

Genomic DNA was extracted from EDTA whole blood using the Monarch High Molecular Weight DNA extraction kit (New England Biolabs, Ipswich, MA). The thermal mixer's speed was set to 2000 rpm. To maximize DNA concentration for library preparation, elution was performed in 50 instead of 100 µL elution buffer II. DNA concentrations were quantified using the Nanodrop 2000 spectrophotometer as well as a Qubit fluorometer (Thermo Fisher Scientific, Waltham, MA).

For S01 to S03, libraries were individually constructed without barcoding, following ONT's ligation sequencing genomic DNA V14 protocol for PromethION (version: GDH_9174_v144_revN_10-Nov2022). Because average sequencing coverage with single-sample libraries was ample for reliable variant calling (>35×,

Table 1. Sample-specific sequencing and mapping metrics

Sample	Barcode	Total output* (Gb)	N50 (kb)	Mean read quality	Mean sequence length for read classification during AS (±SD)	Mean coverage (±SD)	
						Inside ROI	Outside ROI
S01	No	51.7	22.0	24.7	371 (±82)	53.4 (±15.0)	15.4 (±8.7)
S02	No	47.3	21.6	22.8	371 (±74)	52.4 (±15.9)	15.3 (±8.0)
S03	No	37.0	18.9	23.6	370 (±71)	38.5 (±11.8)	10.9 (±6.5)
S04	Yes (2)	23.6†	25.0	22.6	361 (±86)	27.1 (±7.6)	6.5 (±4.4)
S05	Yes (2)	26.0†	32.4	22.6	361 (±86)	18.9 (±6.6)	5.0 (±3.4)

Mean coverages have been calculated after basecalling with the fast mode and before excluding reads below a length of 1500 (ie, ejected reads). Read quality is a measure of accuracy in PHRED scores, for example, 20 and 30 translate to 99% and 99.9% correct base calls, respectively.

Gb, gigabases; SD, standard deviation.

*Minimal quality score Q8.

†Total output of 52.4 Gb, including 2.8 Gb that could not be assigned to either S04 or S05.

Table 1), S04 and S05 were multiplexed to reduce sequencing costs using the ONT Native Barcoding Kit 24 V14 protocol (version: NBE_9163_v225_revQ_15Sep2022).

Libraries were sequenced on a PromethION P2 Solo device using FLO-PRO114M flow cells. Those were washed twice (every ~24 hours) with the Flow Cell Wash Kit (EXP-WSH004) before being reloaded with library. For the ROI in AS, we constructed a Fasta file containing 57 sequences extracted from the latest human reference genome (T2T-CHM13v2.0; accession number: GCF_009914755.1, referred as T2T hereafter). These sequences included 64 target genes (51 encoding human erythrocyte antigens, 7 human platelet antigens, 4 human neutrophil antigens, and 2 transcription factors) with the contraction of 13 genes into 6 multigene loci. A buffer region of 50 kilobases (kb) upstream and downstream of the respective genes (loci) was added. The 57 sequences were on average 152 kb long and accounted for 8 680 191 base pairs (bp; 0.28%) of the human genome. A browser extensible data file containing the genomic T2T coordinates is provided as supplemental Table 2.

Computational analyses of Nanopore sequencing data

A schematic overview of computational workflows is provided in Figure 1, whereas complete details on protocols for each workflow are described in supplemental Information Section 1.2.

Briefly, to minimize computational burden, we initially basecalled and demultiplexed (for S04 and S05) reads with the Dorado (version 0.7.1) fast-accuracy model and subsequently mapped reads to T2T with Minimap2[28] (version 2.28). Low-quality and secondary alignments were filtered, and coverage metrics were computed using Qualimap2[29] and mosdepth.[30] Results were plotted using the R packages ggplot2[31] (version 3.5.1) and circlize[32] (version 0.4.16). On-target reads were rebasecalled using the Dorado (version 0.7.1) super-accuracy model and filtered by quality and length (>1500 nucleotides) using SAMtools (version 1.9)[33] and SeqKit[34] (version 2.5.1).

We investigated 3 distinct bioinformatic workflows for generating haplotypes of blood group genes as well as human platelet antigen and human neutrophil antigens genes: (1) an in-house workflow and (2) ONT's EPI2ME Labs workflow, both based on reference-based variant calling for all target genes; and (3) a de novo workflow for genes with suspected hybrid alleles. Workflow 1 involved an in-house pipeline using Clair3 (version 1.0.8)[35] for SNV and indel detection, WhatsHap (version 2.3)[36] for variant phasing, and Sniffles2 (version 2.4)[37] for structural variant (SV) calling. Workflow 2 used EPI2ME Labs standard pipelines "wf-alignment" and "wf-human-variation," focusing solely on SNV and small indel detection with default settings, followed by comparisons with workflow 1. Workflow 3, the de novo assembly approach, used Hifiasm (version 0.19.9)[38] to assemble HERRO-corrected reads.[39] Resulting contigs of the suspected hybrid alleles were evaluated against publically available *RHD* and *RHCE* sequences (supplemental Table 3), as well as *GYP* hybrid sequences[40] to define precise breakpoints.

Variants from workflow 1 were annotated with the blood group allele database from the International Society of Blood Transfusion (ISBT)[40] and compared with in-house MALDI-TOF MS genotype data (supplemental Table 4). These manual ISBT allele predictions were further validated with RBCeq2.[41] Novel coding variants were assessed for deleteriousness using combined annotation dependent depletion, (CADD) version 1.7 scoring,[42] and flanking intronic variants were evaluated for splicing effects using Alamut Visual Plus. Finally, calls pointing to rare variants (minor allele frequency [MAF] <0.01) were verified by Sanger sequencing.

Results

Nanopore sequencing and variant calling

A detailed description of sequencing output is provided in supplemental Information Section 2.1. In brief, PromethION runs produced 37.0 to 52.4 gigabases (Gb) per sample (Q $\geq$ 8), yielding mean on-target coverages between 18.9× and 53.4×, with a mean enrichment factor of 3.6 (Table 1; Figure 2; supplemental Figure 1). Sample-specific coverages per gene are summarized in Figure 2B, highlighting genes defined as outliers. Most genes with very low coverage either pointed to hemizygosity or encompassed homologous regions, complicating unambiguous read mapping. Because coverage across *CD99* was virtually absent in all samples, the gene was excluded from downstream analyses. The adjacent *XG* gene, however, showed expected coverage and remained in the analyzed gene panel.

Supplemental Information Section 2.2 summarizes the short variant calling results in detail. Reference-based analysis with

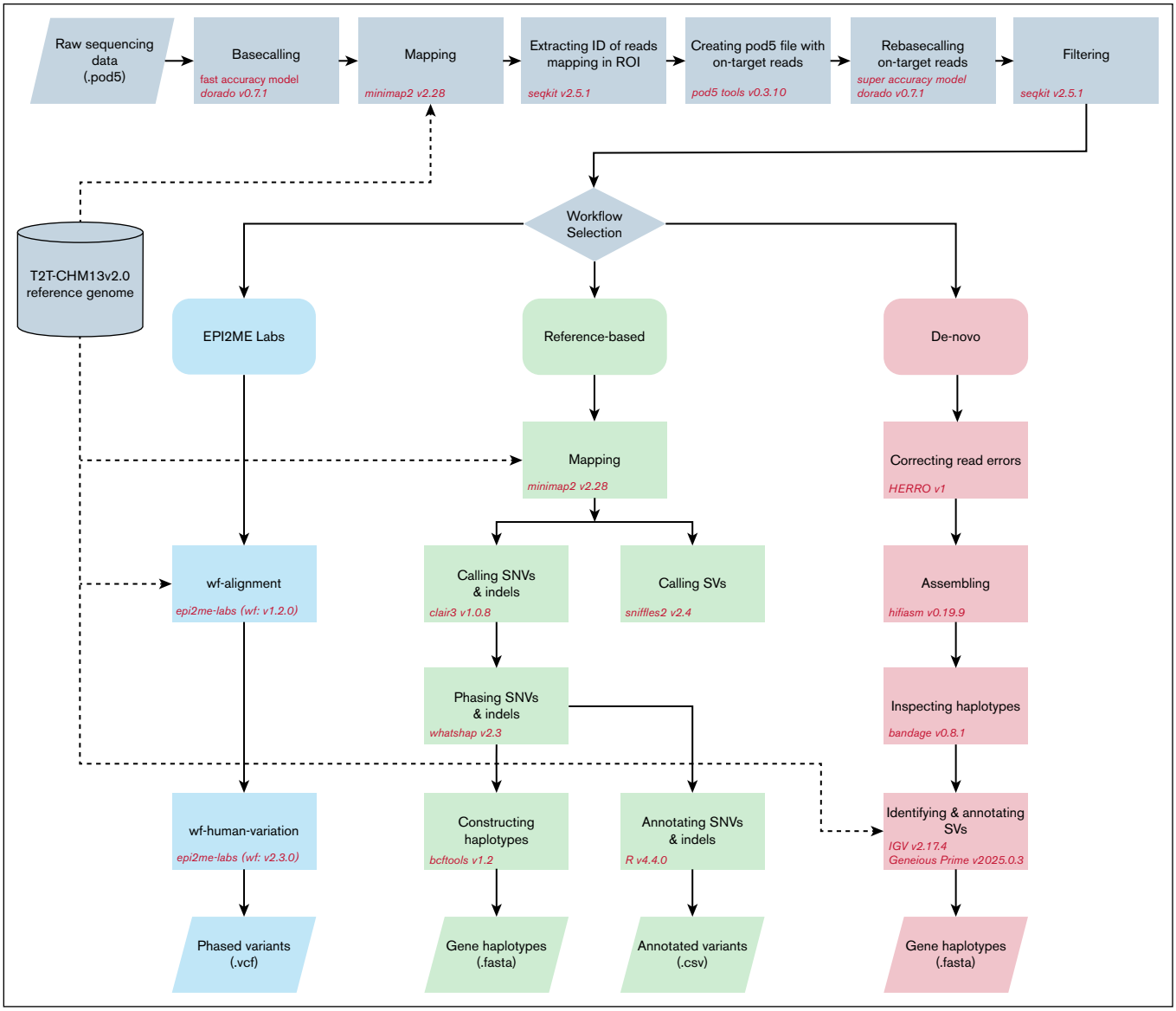

Figure 1. Overview of the bioinformatic workflows applied to ONT sequencing data. A preworkflow is given at the top of the flowchart before splitting up into the 3 workflows EPI2ME Labs (blue), inhouse reference-based (green), and de novo assembled (red). The tools used at each stage of the analysis are indicated alongside their corresponding step. wf, workflow.

workflow 1 identified up to ~6000 variants per sample, with up to 192 variants present in coding or flanking regions (Table 2). After excluding unreliable variant calls (supplemental Information Section 2.3) and restricting analysis to blood group genes with a ISBT reference sequence (n = 47), 56 novel protein-altering alleles were identified (Table 3). Twenty-four alleles were in recently ratified blood group systems (ISBT 037-044), which lack comprehensive allele catalogs. Of the remaining 32, 19 (plus the 8 kb deletion in *RHCE*01*) carried at least 1 potentially deleterious SNV (CADD score of >15). Some variants were recurrent across samples and common in the population (MAF >0.05), or located on null alleles. The remaining 7 alleles contained potentially deleterious variants observed only once, all rare (MAF <0.05), except c.298G>A on *FY*02* in S02. Lists of variants not being part of ISBT allele categories[40] and their predicted deleteriousness are provided in supplemental Tables 5A and 5B. Allele assignments using ISBT nomenclature and RBCeq2-based phenotype predictions are given in supplemental Tables 6 and 7, and are further described in supplemental Information Section 2.4.

Among the identified variants in coding or exon-flanking regions, 78 were rare or of unknown frequency in the general population. After excluding those present in the *GYP*401* hybrid allele (n = 5) or genomic regions challenging to analyze (*C4A*/*C4B*, *CR1*, and *FCGR3B*, n = 47), all remaining 26 variants (25 SNVs and 1 indel) across 18 genes were confirmed by Sanger sequencing (supplemental Information Section 2.5; supplemental Table 8).

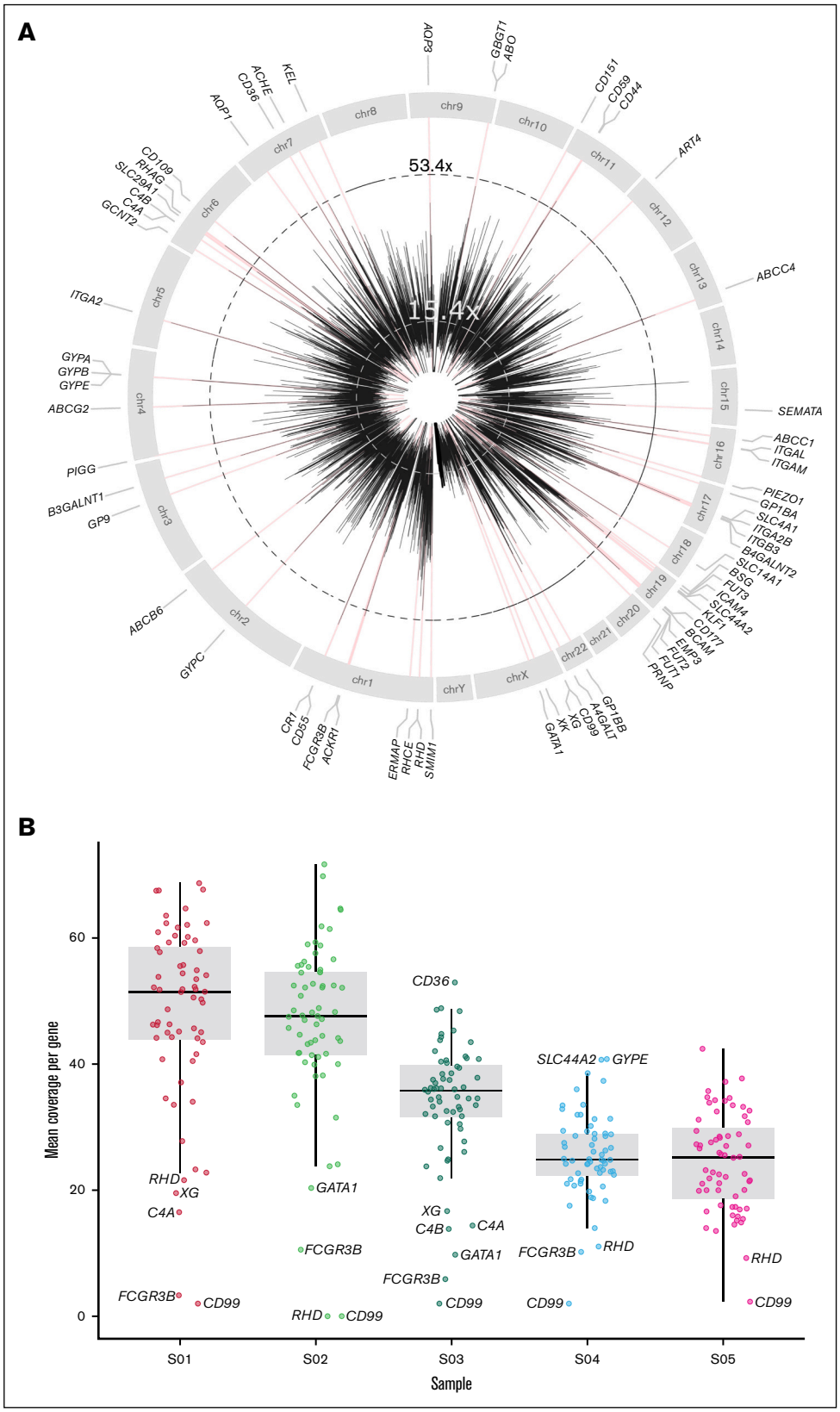

Figure 2. Coverage of on-target and off-target regions across samples. (A) Circos plot showing read coverage across the genome for S01. Genes of interest are indicated according to chromosomal positions around the outer circle (N = 64). Mean coverages across the genome (100 kb windows) are shown by black lines; mean coverages in ROI are further highlighted in red. Circles of dashed lines represent mean read coverage within (black) and outside (white) ROI. Corresponding circos plots for S02 to S05 are shown in supplemental Figure 1. (B) Sample-stratified box plots calculated from mean coverages per targeted gene. Genes with coverage identified as outliers or below 10× are labeled by their names.

Table 2. Sample-specific variant calling results

Sample	Total variants	Variants in coding exons and intronic flanking regions (±10 bp) of ROI (Q10)					
		N	Frameshift	Nonsense	Nonsynonymous	Synonymous	Flanking
S01	3778	130 **(102)**	1	1	59 **(46)**	50 **(38)**	19 **(16)**
S02	4004	136 **(108)**	1	1	68 **(50)**	45 **(40)**	21 **(16)**
S03	5985	172 **(140)**	1	2	63 **(49)**	81 **(69)**	25 **(19)**
S04	4439	174 **(100)**	3 **(1)**	1	98 **(45)**	56 **(41)**	16 **(12)**
S05	4309	192 **(120)**	2 **(1)**	2	110 **(57)**	61 **(44)**	17 **(16)**

The total number of variants did not include SVs and excluded any variant calls in *CD99* (XG) because of low coverage. Number of variants in coding exons and splice site regions (column 3) were further divided according to their consequences (columns 4-8). Bold numbers in brackets represent variants with high reliability. This excludes variants called in *GYPA* and *GYPB* for S02 and in *RHD* and *RHCE* for S04 and S05 owing to the presence of hybrid alleles. Due to ambiguous read alignments to *C4A* and *C4B*, as well as to repetitive regions within *CR1* and the *FCGR2*/*FCGR3* gene cluster locus, variants in those 4 genes were also excluded.

Short variant calls from workflows 1 and 2 showed very high concordance. Restricting comparisons to Q ≥ 10 variants in coding regions yielded identical calls for S01 and S05 and differences of 2 (S04) or 3 variants (S02, S03) in the other samples. All these discrepancies were confined to genes with unreliable variant calls (*GYPA*, *C4A*/*C4B*, and *FCGR3B*).

SV analysis in workflow 1 identified between 18 and 33 insertions or deletions per sample, with comparable median lengths of insertions (110-330 bp) and deletions (109-210 bp), supplemental Table 9. Two deletions (rs71151470 in *ITGAM* and the diagnostically relevant rs1557633501 in *RHCE*, associated with RH4 expression) were shared among all samples. Sniffles2 failed to identify the large deletions causing *RHD***01N.01* (ie, the loss of the entire *RHD* gene), present in all the samples. After reducing SV calls to those overlapping coding regions, a single deletion (8636 bp, NC_060925.1:g.25199922–25208558) spanning *RHCE* introns 8 to 9 in S01 remained and was confirmed by gap-PCR (supplemental Information Section 1.2.2).

De novo assembly for the 57 ROI produced between 103 and 156 primary contigs per sample (supplemental Table 10). The N50 values ranged from 136 kb (S02) to 223 kb (S01), indicating high assembly contiguity as target loci were on average 152 kb. Concentrating on the hybrid alleles in the MNS (S02) and RH (S04, S05) systems, they were all successfully resolved by this method (see "Sample-specific blood group alleles"). Consensus sequences from workflow 1 and de novo contigs for *RHD* and *RHCE* showed very high similarity (99.81%-99.97%, N = 10). Most discrepancies occurred in homopolymer/repetitive regions, leaving 0 to 9 SNVs per sample elsewhere. One exonic error (an insertion in *RHD* exon 7 in the reference-based consensus) was determined by read inspection. Both workflows consistently identified known *RHCE* intronic insertions (109 bp, 288 bp, 653 bp) and an additional 979-bp *RHCE*-specific insertion in intron 9 of *RHD***01EL.44*.

Sample-specific blood group alleles

Here, we focus on the most relevant blood group alleles identified in each sample, specifically those not fully resolved by alternative technologies. Summaries of additional findings are provided in supplemental Information Section 2.6. For hybrid alleles, we used workflow 3, given the high likelihood of ambiguous read mapping and the frequent failure of SV callers to detect such recombination events.

S01 was the only sample in which the SV calling tool detected an SV spanning coding exons. The aforementioned >8 kb deletion spanning from *RHCE* intron 8 to 9 was novel and represented the largest deletion reported in the *RHCE* gene to date. The deletion was called on the haplotype carrying the *RHCE*01* allele explaining the original phenotype-genotype discrepancy, that is, the absent RH4 expression. Moreover, this deletion was not only called in the reference-based workflows but also identified through

Table 3. Sample-specific summary of assigned blood group alleles

Sample	Assigned alleles*	Reference-matching alleles	Other known alleles	Novel alleles with only synonymous variants	Novel alleles with nonsynonymous variants
S01	97	53	9	20 **(19)**	15 **(10)**
S02	97	49 **(48)**	9	23	16 **(11)**
S03	97	47	9	22	19 **(13)**
S04	100	55	13	15	17 **(11)**
S05	97	50	10	21 **(20)**	16 **(11)**

Included were 45 blood group systems and 2 transcription factors (GATA1, KLF1) recognized by ISBT (as of May 2024), but the last ratified system, CD36, had to be excluded because no official reference sequence had yet been defined. Due to low coverages across samples, *CD99* allele determination (XG) was also excluded. Reference-matching alleles are those, with respect to coding variation, identical to the ISBT-defined reference allele for each blood group gene (eg, *ABO*A1.01* for *ABO*). Results of the pathogenicity analysis of variants in novel alleles are provided in supplemental Table 5. Due to ambiguous read alignments to *C4A* and *C4B*, as well as to repetitive regions within *CR1*, allele assignment for these genes is less reliable and number of alleles excluding those genes are given in brackets and highlighted in bold.

*Differences in number of assigned alleles are due to genes located on the X chromosome (ie, *XK*, *XG*, and *GATA1*).

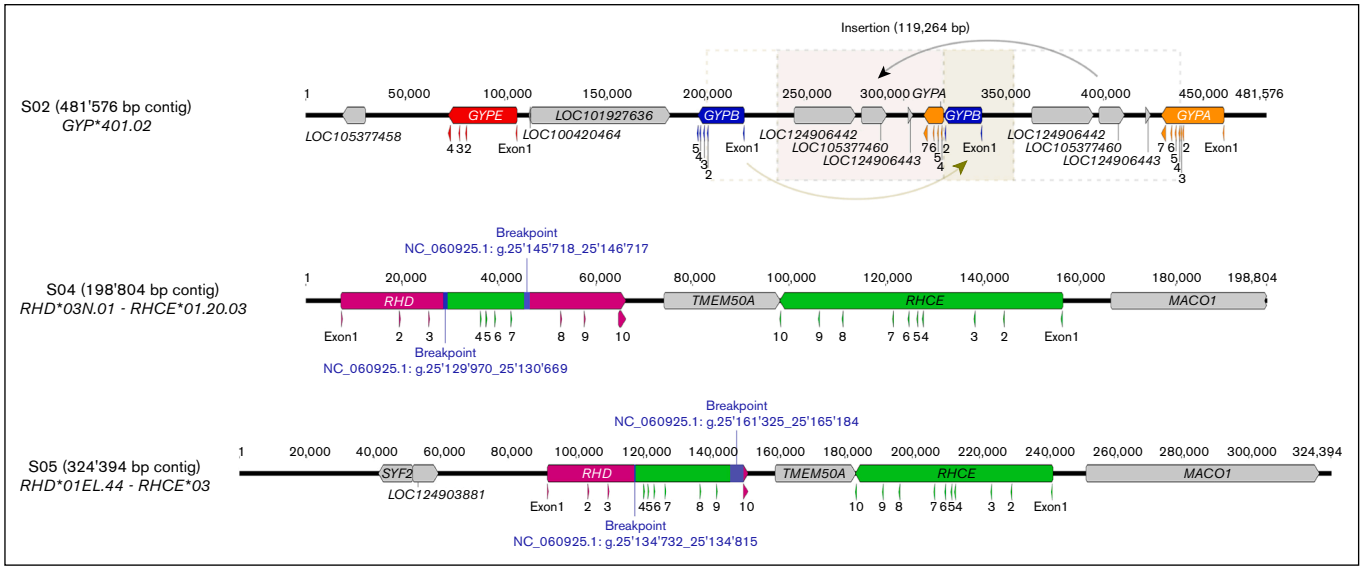

Figure 3. Overview of contigs containing hybrid alleles. Contigs were annotated using official ISBT blood group reference annotations. Genes of uncertain functions are depicted by LOC numbers. LOC, uncharacterized locus.

de novo assembly (workflow 3), in which it was present within a 254.7 kb contig, spanning the entire *RH* locus composed of *RHD***01N.01* (ie, absence of *RHD*) and *RHCE***01*. For the 67 variants typed by MALDI-TOF MS across 21 blood group genes, 100% concordance was observed (supplemental Table 4).

For S02, the de novo workflow assembled a 481.6 kb contig, encompassing *GYPA*, *GYPB*, and *GYPE*, alongside a 119 264 bp insertion, Figure 3. This insertion contained a hybrid gene formed between *GYPB* (exons 1, 2, and pseudoexon 3) and *GYPA* (exons 4-7), corresponding to the *GYP***401* allele types, which explained the weak MNS2 detection without presence of a *GYPA***02* allele. Mapping to *GYPA* and *GYPB* reference sequences allowed narrowing down the recombination site within the insertion to ~180 bp. Comparison with all described 10 *GYP***401* sequences[40] showed identical breakpoint region to *GYP***401.02*, also called *GYP***Sch(type B)*, with only 2 divergent nucleotides in the ~940 bp intronic region between pseudoexon 3 and exon 4.[43,44] Both were in the part originating from *GYPA*. The first, at position 696, was an unknown and therefore questionable SNV only present in *GYP***401.01* and *02* reference sequences. The second divergent SNV was a common *GYPA* variant (rs115171794, position 776). Although the breakpoint region of our assembled allele was also in agreement with *GYP***401.10*, no continuous sequence for that allele[44] was available. Our assembly also revealed the presence of *GYPA***01* and *GYPB***04.06* alleles as part of the *GYP***401.02*. Among 29 variants typed by MALDI-TOF MS, 1 discordance (rs7682260) arose due to the *GYP***401.02* hybrid allele, miscalled as *GYPA***01* homozygous in MALDI-TOF data (supplemental Table 4).

For S03, the young patient with sickle cell disease, the exonic *RHCE* variants initially detected by Sanger sequencing could not be phased into distinct haplotypes (supplemental Table 1). Workflow 1 phased variants c.48G>C, c.105C>T, and c.1025C>T to the same haplotype, forming a new ISBT allele on the basis of *RHCE***01.02.01* with additional c.105C>T (p.Asp35=), which is not expected to further change phenotype (ie, weak/partial RH4 and RH5). On the other haplotype, c.254G>C was the only variant called and defines *RHCE***01.06.01*, which results in a weak/partial RH5. Hence, both alleles supported compromised RH5 expression, in line with the detected anti–e-like antibody in the patient before transfusion. With respect to MNS, we detected the null allele *GYPB***03N.04* in combination with synonymous c.96T>C and c.102A>G. This allele included the exon 5 skipping c.270+5G>T variant (rs139511876),[45] which was reassuringly predicted to weaken the splice site between exon and intron 5 (supplemental Table 5B). The MNS:-3,w5 phenotype predicted for this allele was serologically masked by *GYPB***03* on the other haplotype, but a minimal MNS6 expression (He+) could be expected (not determined).[46] The 29 variants called by MALDI-TOF MS across 18 blood group genes were all congruent with those found in Nanopore sequencing data (supplemental Table 4).

For S04, we identified the suspected *RHD-CE-D* hybrid allele located on a 198.8 kb–long contig (Figure 3). The allele comprised exons 1 to 3 and 8 to 10 from *RHD*, separated by *RHCE* exons 4 to 7. Additionally, this contig exhibited characteristics of *RHD***03N.01* (c.186T in exon 2 as well as c.410T and c.455C in exon 3), which explained the weak RH2 positivity in the absence of any *RHCE***02* allele in *cis* or *trans* (RH10 and RH20 were not serologically investigated). The hybrid allele was found in *cis* with *RHCE***01.20.03*, defined by c.48C, c.733G, and c.1006T positions, the latter 2 also found in exons 5 and 7 within the hybrid, and in *trans* with *RHCE***01*. Mapping 25 full or partial *RHD* and *RHCE* sequences (supplemental Table 3) to our contig narrowed down potential breakpoints to regions of a few hundred base pairs in introns 3 (NC_060925.1:g.25129970-25130669) and 7 (NC_060925.1:g.25145718-25146717). The identified breakpoints were consistent with those previously described.[47] The 29 variants called by MALDI-TOF MS were all in agreement with those detected in the Nanopore data (supplemental Table 4).

Finally, we detected an *RHD-CE-D* hybrid allele in S05, located on a 324.4 kb contig (Figure 3). The part with *RHD* origin comprised exons 1 to 3, followed by exons 4 to 9 from *RHCE*, and exon 10 from *RHD,* characteristic of *RHD**01EL.44*. The hybrid allele was in *cis* with *RHCE*03*. Analysis of breakpoints using the same strategy as for S04 identified a homologous region in intron 3 in which recombination likely occurred. We could narrow down this 5′ breakpoint region to <100 bp (NC_060925.1:g.25134732-25134815), which closely resembled the one previously reported.[48] In intron 9, the breakpoint region was much larger (~4 kb) due to identity of the 2 genes (NC_060925.1:g.25161325-25165184). Although classified as *RHD**01EL*, causing a Del (D-elute) phenotype, we could neither detect any RH1 antigen with adsorption-elution (performed on 2 independent samples) nor with flow cytometry techniques. Also, for S05, all the available blood group variants called by MALDI-TOF MS (N = 29) were concordant with the calls from Nanopore data (supplemental Table 4).

Discussion

Resolving hybrid alleles

Using a de novo approach based on assembling reads that were previously corrected for ONT-specific errors, we successfully constructed haplotype contigs ranging from 199 to 482 kb, spanning the entire *RH* or *GYP* locus with their respective hybrid alleles. Regarding the *GYP* hybrid allele, this is, to our knowledge, the first reported *GYP**401* allele sequence spanning the entire locus. Exact breakpoint determination for the insertion proved infeasible due to extensive sequence identity. We could, however, identify the same ~180 bp region as in type 02 (and 10) for the recombination spot between *GYPB* and *GYPA*. For the 2 investigated *RH* hybrid alleles, *RHD**03N.01* and *RHD**01EL.44*, breakpoint regions were comparable with those previously reported,[47,48] but refinement was limited to several kilobases due to complete sequence identity across intron 9 of *RHD* and *RHCE*. Notably, both adsorption-elution and flow cytometric testing yielded negative results for this *DEL* allele, which was different from previous in-house and external experience.[49] Although differences in the 5′ breakpoint could be responsible for this finding, Del phenotypes may also be misclassified as RhD-negative when expressing a partial phenotype.[50]

Hybrid alleles have engaged the blood group genetics community for a long time. Approaches based on short-read sequencing of exons had some success inferring hybrids based on exon coverage ratios.[51-54] Furthermore, short-read data from whole-genome sequencing studies or based on long-range PCR amplicon sequencing have been used for full-gene haplotyping of hybrid alleles, but the interpretation of coverage ratios and phasing remained delicate.[55-58] With the advent of accurate long-read sequencing, unambiguous read mapping in repetitive regions became feasible. Nevertheless, SV callers developed for long-read data still struggle with translocations in paralogous gene systems. This challenge was evident in our study, in which reference-based SV calling tools not only failed at calling translocations (hybrid alleles) but also missed the deletions associated with *RHD**01N.01* in all samples.

As a consequence of the limited success of reference-based approaches, assembly-based methods have emerged. Here, we used a specific combination of Hifiasm and HERRO, which was critical for achieving high-quality assemblies. In fact, using alternative assemblers or omission of the read correction tool resulted in shorter and fragmented contigs. Other assembly-based methods have suggested locus-specific solutions. For instance, a capture-based enrichment of *RH* sequences followed by PacBio long-read sequencing has been proposed,[59] but assembly success was dependent on high sample heterogeneity owing to modest read length (2-3 kb). Chinook, a tool recently introduced by ONT, also relies on capturing target reads followed by a locus-specific assembly, but is currently only designed to resolve SVs in the *CYP2D6/D7/D8* gene-pseudogene system. In future, replacing single reference genomes by pangenomes could also represent a solution for better resolving hybrid alleles because it alleviates the dependency on the representativeness of the chosen reference genome (supplemental Information Section 3.1).[60]

Resolving the entire blood group genome

In this study, we sequenced 56 loci including 48 blood group and 2 transcription factor genes. Such targeted sequencing approach offers significant advantages over high-throughput genotyping methods for blood group determination,[15,61-64] particular for rare variants, structural modifications, or haplotype phasing. Previous targeted exome sequencing[52] captured coding variants but left phasing and haplotypes incomplete. Our AS long-read sequencing approach solved these issues, producing full genes and fully phased haplotypes. Although current cost (~$600 per sample) limits routine use, this method could be beneficial for cases involving unexplained discrepancies in the RH or MNS system; for patients who receive frequent transfusions, for whom comprehensive antigen determination is crucial; or for urgent clinical scenarios, in which rapid turnaround is critical. For instance, in a young patient with sickle cell disease (S03), we achieved complete allelic resolution within 2 to 3 days, including phased *RHCE* alleles, clarifying antibody findings and enabling precise donor matching. Establishing a full blood group genome early in life may reduce alloimmunization risk. However, to maximize the benefits of such comprehensive blood group resolution, equivalent methods to provide extended genotype data for large and diverse panels of donors are essential.[16] Integrating AS-based sequencing with scalable genotyping approaches could become key to improving transfusion safety and advancing personalized blood matching strategies.

Full-gene haplotypes also provide comprehensive genetic information in the often overlooked intronic and gene-flanking regions, along with their associated regulatory elements.[65-67] Although the regulation of many blood group genes remains poorly understood, a growing body of literature indicates that variants in these regions can dictate differential gene expression.[68-70] Currently, only ~130 intronic variants, primarily located in splice sites, are documented in the ISBT tables.[40] In only 5 samples, we identified 2 unreported variants predicted to form alternative splice sites in *ERMAP* (Scianna blood group system) and *B4GALNT2* (SID blood group system). Furthermore, we found many new alleles with solely synonymous variants (Table 3). As not expected to alter phenotypes, they are not systematically reported and classified into separate allelic categories by the ISBT, but their handling is inconsistent (eg, *JK**01W.01* vs *JK**01W.06*).[40]

Many novel alleles were detected in Chido/Rodgers (*C4A/B*) and Knops (*CR1*), in which reference-based calling was unreliable

(supplemental Information Section 2.3). Variant detection in *C4A/B*, *CR1*, and *FCGR3B* would likely benefit from assembly-based strategies, although this was beyond the scope of this study.

We detected 56 novel alleles based on nonsynonymous variation (Table 3), that is, 1 in 8 alleles harbored novel coding SNVs or unreported SNV combinations. Notably, some of these were shared among samples and almost half of these novel alleles were found in genes of recently ratified blood groups with relatively low clinical significance and still modest allele collections. Excluding recently ratified genes, 19 alleles carried coding SNVs with CADD score of >15. Many were recurrent, associated to null alleles, or common in the population (MAF >0.05), suggesting limited impact. Six SNVs remain candidates for novel antigen-modifying alleles, although confirmation requires functional investigation.

The accelerating discovery of novel alleles presents new challenges for the curation and classification of blood group alleles. Moreover, the ability to resolve full-length haplotypes as shown in this study may result in a new standard, requiring updates to existing nomenclature guidelines, including a revision of what constitutes a unique blood group allele. A classification system based solely on exonic and splice-site variants is no longer sufficient to reflect the rapid accumulation of haplotype-resolved sequences covering entire genes.[56,57,66,67,71]

Practical recommendations for workflow selection

Our findings indicate that although reference-based approaches are generally sufficient for blood group genotyping, they may miss clinically important SVs, such as the *RHD**01N.01* deletion or *GYP* hybrid alleles. This underscores the value of complementing reference-based analysis with de novo assembly in loci prone to recombination, ensuring accurate detection of alleles with potential transfusion implications. Integrating both strategies not only strengthens analytical robustness but also minimizes the risk of misclassification that could affect antigen prediction. For novel or low-coverage variants, orthogonal confirmation remains essential to safeguard diagnostic accuracy.

Alternatives and outlook

Comprehensive long-read whole-genome sequencing can resolve complex hybrid alleles in blood group genetics,[72] but high costs and extensive ethical constraints limit its use, making targeted enrichment necessary. We explored AS as a practical alternative because it combines parallel targeting of many genes with preserving of native read length. Alternative approaches are either suitable to only tackle few genes (eg, long-range PCR[66] and CRISPR–associated protein 9–guided adapter ligation[73]) or have limitations with respect to read length (eg, hybridization-based long-read capturing[74]). These limitations hinder the resolution of hybrid alleles, no matter if a reference- or assembly-based approach is chosen. Because enrichment is purely computational, AS also offers rapid turnaround (~2 hours from DNA extraction to sequencing) and flexibility to adjust targeted regions. The main hindrance for broader adoption remains its relatively high per-sample costs primarily due to suboptimal enrichment efficiency. Improvements in sequencing chemistry should boost yield and multiplexing, lowering costs. Technical barriers, for example the need of setting up own pipelines, have largely disappeared since tools such as EPI2ME Labs became available, whose accuracy and ease of use were validated in our analysis. A key remaining challenge is translating variant calls into blood group alleles, because most tools are optimized for short reads,[75,76] but new software such as RBCeq2[41] shows promise in addressing this gap. In conclusion, Nanopore sequencing of the blood group genome, enriched with AS, proved highly suitable to elucidate allelic composition in most blood group systems. Moreover, complex hybrid alleles in the RH or MNS systems were successfully resolved by opting for an assembly-based workflow.

Acknowledgment

This work was financially supported by the Blood Transfusion Service Zurich, Swiss Red Cross (Switzerland).

Authorship

Contribution: M.G., G.A.T., M.P.M.-G., and S.M. initiated the study and contributed ideas; M.G., G.A.T., M.P.M.-G., and S.M. designed the study and experiments; S.S., N.T., L.W., B.M., and C.E. contributed samples and provided preanalysis data; M.G., G.A.T., and S.K. performed experiments and analyzed data; M.G. and G.A.T. wrote the manuscript; and all authors commented on the manuscript and approved the final version.

Conflict-of-interest disclosure: The authors declare no competing financial interests.

ORCID profiles: M.G., 0000-0001-8574-9640; G.A.T., 0000-0003-4436-3455; S.K., 0000-0003-0965-0539; B.M., 0000-0002-6238-8405.

Correspondence: Morgan Gueuning, Department of Research and Development, Blood Transfusion Service Zurich, Swiss Red Cross, Rütistrasse 19, 8952 Schlieren, Switzerland; email: m.gueuning@zhbsd.ch.

References

1. Loose M, Malla S, Stout M. Real-time selective sequencing using nanopore technology. *Nat Methods*. 2016;13(9):751-754.
2. Payne A, Holmes N, Clarke T, Munro R, Debebe BJ, Loose M. Readfish enables targeted nanopore sequencing of gigabase-sized genomes. *Nat Biotechnol*. 2021;39(4):442-450.
3. Lin Y, Dai Y, Liu Y, et al. Rapid PCR-based nanopore adaptive sequencing improves sensitivity and timeliness of viral clinical detection and genome surveillance. *Front Microbiol*. 2022;13:929241.
4. Liu P-Y, Wu H-C, Li Y-L, et al. Comprehensive pathogen identification and antimicrobial resistance prediction from positive blood cultures using nanopore sequencing technology. *Genome Med*. 2024;16(1):141.

5. Ulrich J-U, Epping L, Pilz T, et al. Nanopore adaptive sampling effectively enriches bacterial plasmids. *mSystems*. 2024;9(3):e0094523.
6. Miller DE, Hanna P, Galey M, et al. Targeted long-read sequencing identifies a retrotransposon insertion as a cause of altered GNAS exon A/B methylation in a family with autosomal dominant pseudohypoparathyroidism type 1b (PHP1B). *J Bone Miner Res*. 2022;37(9):1711-1719.
7. Miyatake S, Koshimizu E, Fujita A, et al. Rapid and comprehensive diagnostic method for repeat expansion diseases using nanopore sequencing. *NPJ Genom Med*. 2022;7(1):62.
8. Geyer J, Opoku KB, Lin J, et al. Real-time genomic characterization of pediatric acute leukemia using adaptive sampling. *Leukemia*. 2025;39(5):1069-1077.
9. Chevrier S, Richard C, Mille M, Bertrand D, Boidot R. Nanopore adaptive sampling accurately detects nucleotide variants and improves the characterization of large-scale rearrangement for the diagnosis of cancer predisposition. *Clin Transl Med*. 2025;15(1):e70138.
10. Nakamura W, Hirata M, Oda S, et al. Assessing the efficacy of target adaptive sampling long-read sequencing through hereditary cancer patient genomes. *NPJ Genom Med*. 2024;9(1):11.
11. da Cunha Gomes EG, Machado LAF, de Oliveira LC, Neto JFN. The erythrocyte alloimmunisation in patients with sickle cell anaemia: a systematic review. *Transfus Med*. 2019;29(3):149-161.
12. Linder GE, Chou ST. Red cell transfusion and alloimmunization in sickle cell disease. *Haematologica*. 2021;106(7):1805-1815.
13. Shah FT, Sayani F, Trompeter S, Drasar E, Piga A. Challenges of blood transfusions in β-thalassemia. *Blood Rev*. 2019;37:100588.
14. Flegel WA, Gottschall JL, Denomme GA. Implementing mass-scale red cell genotyping at a blood center. *Transfusion*. 2015;55(11):2610-2609.
15. Gleadall NS, Veldhuisen B, Gollub J, et al. Development and validation of a universal blood donor genotyping platform: a multinational prospective study. *Blood Adv*. 2020;4(15):3495-3506.
16. Gleadall NS, Koets L, Shamardina O, et al. Array genotyping of transfusion relevant blood cell antigens in 6946 ancestrally diverse subjects. *Blood*. 2025;146(12):1511-1524.
17. Schonewille H, Honohan Á, van der Watering LMG, et al. Incidence of alloantibody formation after ABO-D or extended matched red blood cell transfusions: a randomized trial (MATCH study). *Transfusion*. 2016;56(2):311-320.
18. Pirenne F, Pondarré C. Alloimmunization and hyperhemolysis in sickle cell disease. *Hematol Am Soc Hematol Educ Progr*. 2023;2023(1):653-659.
19. Fichou Y. Hybrids and microconversions in RH genes: investigation and implication in transfusion therapy. *Ann Blood*. 2023;8(Ii):1-21.
20. Sippert E, Fujita CR, Machado D, et al. Variant RH alleles and Rh immunisation in patients with sickle cell disease. *Blood Transfus*. 2015;13(1):72-77.
21. Fasano RM, Chou ST. Red blood cell antigen genotyping for sickle cell disease, thalassemia, and other transfusion complications. *Transfus Med Rev*. 2016;30(4):197-201.
22. Chou ST, Flanagan JM, Vege S, et al. Whole-exome sequencing for RH genotyping and alloimmunization risk in children with sickle cell anemia. *Blood Adv*. 2017;1(18):1414-1422.
23. 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):177-186.
24. 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(4):418-424.
25. Fürst D, Tsamadou C, Neuchel C, Schrezenmeier H, Mytilineos J, Weinstock C. Next-generation sequencing technologies in blood group typing. *Transfus Med Hemother*. 2020;47(1):4-13.
26. Vorholt SM, Hamker N, Sparka H, et al. High-throughput screening of blood donors for twelve human platelet antigen systems using next-generation sequencing reveals detection of rare polymorphisms and two novel protein-changing variants. *Transfus Med Hemother*. 2020;47(1):33-44.
27. Lane WJ. Recent advances in blood group genotyping. *Ann Blood*. 2021;6(5):31.
28. Li H. Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics*. 2018;34(18):3094-3100.
29. Okonechnikov K, Conesa A, García-Alcalde F. Qualimap 2: advanced multi-sample quality control for high-throughput sequencing data. *Bioinformatics*. 2016;32(2):292-294.
30. Pedersen BS, Quinlan AR. Mosdepth: quick coverage calculation for genomes and exomes. *Bioinformatics*. 2018;34(5):867-868.
31. Wilkinson L. ggplot2: elegant graphics for data analysis by WICKHAM, H. *Biometrics*. 2011;67(2):678-679.
32. Gu Z, Gu L, Eils R, Schlesner M, Brors B. *circlize* implements and enhances circular visualization in R. *Bioinformatics*. 2014;30(19):2811-2812.
33. Danecek P, Bonfield JK, Liddle J, et al. Twelve years of SAMtools and BCFtools. *Gigascience*. 2021;10(2):giab008.
34. Shen W, Le S, Li Y, Hu F. SeqKit: a cross-platform and ultrafast toolkit for FASTA/Q file manipulation. *PLoS One*. 2016;11(10):e0163962.
35. Zheng Z, Li S, Su J, Leung AW-S, Lam T-W, Luo R. Symphonizing pileup and full-alignment for deep learning-based long-read variant calling. *Nat Comput Sci*. 2022;2(12):797-803.
36. Patterson M, Marschall T, Pisanti N, et al. WhatsHap: weighted haplotype assembly for future-generation sequencing reads. *J Comput Biol*. 2015;22(6):498-509.
37. Smolka M, Paulin LF, Grochowski CM, et al. Detection of mosaic and population-level structural variants with Sniffles2. *Nat Biotechnol*. 2024;42(10):1571-1580.

38. Cheng H, Concepcion GT, Feng X, Zhang H, Li H. Haplotype-resolved de novo assembly using phased assembly graphs with hifiasm. *Nat Methods*. 2021;18(2):170-175.

39. Stanojevic D, Lin D, Nurk S, Florez De Sessions P, Sikic M. Telomere-to-telomere phased genome assembly using HERRO-corrected simplex nanopore reads. *bioRxiv*. Preprint posted online 21 May 2024. https://doi.org/10.1101/2024.05.18.594796

40. *Table of Blood Group Systems*. The International Society of Blood; 2024. Accessed 15 May 2025. https://www.isbtweb.org/isbt-working-parties/rcibgt/blood-group-allele-tables.html

41. McIntyre L, Sowah S, Perros A, Hyland C, Flower R, Roulis E. RBCeq2. GITHUB. Accessed 28 April 2025. https://github.com/limcintyre/RBCeq2

42. Rentzsch P, Witten D, Cooper GM, Shendure J, Kircher M. CADD: predicting the deleteriousness of variants throughout the human genome. *Nucleic Acids Res*. 2019;47(D1):D886-D894.

43. Meyer S, Vollmert C, Trost N, et al. MNSs genotyping by MALDI-TOF MS shows high concordance with serology, allows gene copy number testing and reveals new St(a) alleles. *Br J Haematol*. 2016;174(4):624-636.

44. Watanabe-Okochi N, Tsuneyama H, Isa K, et al. Novel hybrid genes and a splice site mutation encoding the Sta antigen among Japanese blood donors. *Vox Sang*. 2020;115(8):756-766.

45. Lapadat R, Anani WQ, Bensing KM, et al. A pair of S-silencing single nucleotide variants cis-linked on GYPB. *Transfusion*. 2021;61(4):E34-E36.

46. Daniels G. Human Blood Groups. Wiley; 2013.

47. Flegel WA, Wagner FF. Two molecular polymorphisms to detect the (C)ce(s) type 1 haplotype. *Blood Transfus*. 2014;12(1):136-137.

48. Srivastava K, Bueno MU, Flegel WA. Breakpoint regions of an RHD-CE(4-9)-D allele and a rare JK allele in a Pacific Islander individual. *Blood Transfus*. 2024;22(3):189-197.

49. Lilić M, Jaklin G, Gojčeta K, Raos M, Golubić-Ćepulić B. Uncommon RHD variants and an unconventional RHCE hybrid allele with D epitope expression in blood donors from northwestern Croatia. *Transfusion*. 2025;65(7):1319-1327.

50. Wagner FF. Serology and molecular biology of DEL: a narrative review. *Ann Blood*. 2023;8:28.

51. Dezan MR, Ribeiro IH, Oliveira VB, 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.

52. Schoeman EM, Lopez GH, McGowan EC, 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(4):1078-1088.

53. Wheeler MM, Lannert KW, Huston H, et al. Genomic characterization of the RH locus detects complex and novel structural variation in multi-ethnic cohorts. *Genet Med*. 2019;21(2):477-486.

54. Stef M, Fennell K, Apraiz I, et al. RH genotyping by nonspecific quantitative next-generation sequencing. *Transfusion*. 2020;60(11):2691-2701.

55. Halls JBL, Vege S, Simmons DP, et al. Overcoming the challenges of interpreting complex and uncommon RH alleles from whole genomes. *Vox Sang*. 2020;115(8):790-801.

56. Tounsi WA, Madgett TE, Avent ND. Complete RHD next-generation sequencing: establishment of reference RHD alleles. *Blood Adv*. 2018;2(20):2713-2723.

57. Tounsi WA, Halawani AJ, Sillence KA, Kiernan M, Avent ND, Madgett TE. RHCE genotyping using next generation sequencing: allele specific reference sequences. *Transfusion*. 2025;65(2):363-374.

58. Chang T-C, Haupfear KM, Yu J, et al. A novel algorithm comprehensively characterizes human RH genes using whole-genome sequencing data. *Blood Adv*. 2020;4(18):4347-4357.

59. Zhang Z, An HH, Vege S, 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):180-191.

60. Liao W-W, Asri M, Ebler J, et al. A draft human pangenome reference. *Nature*. 2023;617(7960):312-324.

61. Gassner C, Meyer S, Frey BM, Vollmert C. Matrix-assisted laser desorption/ionisation, time-of-flight mass spectrometry–based blood group genotyping–the alternative approach. *Transfus Med Rev*. 2013;27(1):2-9.

62. Meyer S, Trost N, Frey BM, Gassner C. Platelet antigens using high-throughput MALDI-TOF mass spectrometry. *Methods Mol Biol*. 2015;1310:51-70.

63. Yee MEM, Josephson CD, Winkler AM, et al. Red blood cell minor antigen mismatches during chronic transfusion therapy for sickle cell anemia. *Transfusion*. 2017;57(11):2738-2746.

64. Guo Y, Busch MP, Seielstad M, et al. Development and evaluation of a transfusion medicine genome wide genotyping array. *Transfusion*. 2019;59(1):101-111.

65. Thun GA, Gueuning M, Sigurdardottir S, et al. Novel regulatory variant in ABO intronic RUNX1 binding site inducing A(3) phenotype. *Vox Sang*. 2024;119(4):377-382.

66. Gueuning M, Thun GA, Wittig M, et al. Haplotype sequence collection of ABO blood group alleles by long-read sequencing reveals putative A1 -diagnostic variants. *Blood Adv*. 2023;7(6):878-892.

67. Ying Y, Zhang J, Hong X, et al. Comprehensive annotation of complete ABO alleles and resolution of ABO variants by an improved full-length ABO haplotype sequencing. *Clin Chem*. 2025;71(4):510-519.

68. Sano R, Nakajima T, Takahashi K, et al. Expression of ABO blood-group genes is dependent upon an erythroid cell-specific regulatory element that is deleted in persons with the B m phenotype. *Blood*. 2012;119(22):5301-5310.

69. Wu PC, McGowan EC, Lee YQ, Ghosh S, Hansson J, Olsson ML. Epigenetic dissection of human blood group genes reveals regulatory elements and detailed characteristics of KEL and four other loci. *Transfusion*. 2024;64(6):1083-1096.

70. McGowan EC, Wu PC, Hellberg Å, Lopez GH, Hyland CA, Olsson ML. A bioinformatically initiated approach to evaluate GATA1 regulatory regions in samples with weak D, Del, or D– phenotypes despite normal RHD exons. *Transfus Med Hemother*. 2024;51(4):252-264.

71. Gueuning M, Thun GA, Trost N, et al. Resolving genotype–phenotype discrepancies of the Kidd blood group system using long-read nanopore sequencing. *Biomedicines*. 2024;12(1):225.

72. Li M, Wang L, Li A, et al. Integrated analyses reveal unexpected complex inversion and recombination in RH genes. *Blood Adv*. 2024;8(12):3154-3165.

73. Gilpatrick T, Lee I, Graham JE, et al. Targeted nanopore sequencing with Cas9-guided adapter ligation. *Nat Biotechnol*. 2020;38(4):433-438.

74. Steiert TA, Fuß J, Juzenas S, et al. High-throughput method for the hybridisation-based targeted enrichment of long genomic fragments for PacBio third-generation sequencing. *NAR Genom Bioinform*. 2022;4(3):lqac051.

75. Lane WJ, Vege S, Mah HH, et al. Automated typing of red blood cell and platelet antigens from whole exome sequences. *Transfusion*. 2019;59(10):3253-3263.

76. Hyvärinen K, Haimila K, Moslemi C, et al. A machine-learning method for biobank-scale genetic prediction of blood group antigens. *PLOS Comput Biol*. 2024;20(3):e1011977.