Article

Resolving Genotype–Phenotype Discrepancies of the Kidd Blood Group System Using Long-Read Nanopore Sequencing

Morgan Gueuning, Gian Andri Thun, Nadine Trost, Linda Schneider, Sonja Sigurdardottir, Charlotte Engström, Naemi Larbes, Yvonne Merki, Beat M. Frey, Christoph Gassner et al.

Special Issue

Advances in Molecular Diagnostics of Transfusion Medicine

Edited by
Dr. Tracey Madgett

Article

Resolving Genotype–Phenotype Discrepancies of the Kidd Blood Group System Using Long-Read Nanopore Sequencing

Morgan Gueuning [1,*], Gian Andri Thun [1], Nadine Trost [2], Linda Schneider [2], Sonja Sigurdardottir [2], Charlotte Engström [3], Naemi Larbes [3], Yvonne Merki [2], Beat M. Frey [1,2,3], Christoph Gassner [4], Stefan Meyer [2,†] and Maja P. Mattle-Greminger [1,†]

[1]   Department of Research and Development, Blood Transfusion Service Zurich, Swiss Red Cross, Rütistrasse 19, 8952 Schlieren, Switzerland

[2]   Department of Molecular Diagnostics and Cytometry, Blood Transfusion Service Zurich, Swiss Red Cross, 8952 Schlieren, Switzerland

[3]   Department of Immunohematology, Blood Transfusion Service Zurich, Swiss Red Cross, 8952 Schlieren, Switzerland; c.engstroem@zhbsd.ch (C.E.)

[4]   Institute of Translational Medicine, Private University in the Principality of Liechtenstein, 9495 Triesen, Liechtenstein; christoph.gassner@ufl.li

[*]   Correspondence: m.gueuning@zhbsd.ch; Tel.: +41-58-272-52-41

[†]   These authors contributed equally to this work.

Abstract: Due to substantial improvements in read accuracy, third-generation long-read sequencing holds great potential in blood group diagnostics, particularly in cases where traditional genotyping or sequencing techniques, primarily targeting exons, fail to explain serological phenotypes. In this study, we employed Oxford Nanopore sequencing to resolve all genotype–phenotype discrepancies in the Kidd blood group system (JK, encoded by *SLC14A1*) observed over seven years of routine high-throughput donor genotyping using a mass spectrometry-based platform at the Blood Transfusion Service, Zurich. Discrepant results from standard serological typing and donor genotyping were confirmed using commercial PCR-SSP kits. To resolve discrepancies, we amplified the entire coding region of *SLC14A1* (~24 kb, exons 3 to 10) in two overlapping long-range PCRs in all samples. Amplicons were barcoded and sequenced on a MinION flow cell. Sanger sequencing and bridge-PCRs were used to confirm findings. Among 11,972 donors with both serological and genotype data available for the Kidd system, we identified 10 cases with unexplained conflicting results. Five were linked to known weak and null alleles caused by variants not included in the routine donor genotyping. In two cases, we identified novel null alleles on the *JK*01* (Gly40Asp; c.119G>A) and *JK*02* (Gly242Glu; c.725G>A) haplotypes, respectively. Remarkably, the remaining three cases were associated with a yet unknown deletion of ~5 kb spanning exons 9–10 of the *JK*01* allele, which other molecular methods had failed to detect. Overall, nanopore sequencing demonstrated reliable and accurate performance for detecting both single-nucleotide and structural variants. It possesses the potential to become a robust tool in the molecular diagnostic portfolio, particularly for addressing challenging structural variants such as hybrid genes, deletions and duplications.

Keywords: Kidd blood group system; structural variant; Oxford nanopore sequencing; haplotype; long-read sequencing; high-throughput genotyping

1. Introduction

The latest advancements in third-generation long-read sequencing (TGS) technologies offer notable advantages, including the capacity to elucidate extensive haplotypes and characterize genomic regions that pose challenges for Sanger sequencing or next-generation short-read sequencing [1]. In the specific field of transfusion medicine, TGS complements traditional approaches used for the blood group assessment of samples with complex or discordant serological and genetic results [2–6]. Such cases include serological reactions

exhibiting unexpected weak agglutination or null phenotypes, often resulting from rare or unknown genetic variants that are not typed in routine genotyping workflows [7].

Traditionally, Sanger sequencing of the underlying blood group genes has been used to resolve such elusive cases. However, it has several limitations. A major drawback is the lack of phase information, i.e., the inability to reconstruct haplotypes, due to overlapping signals from maternal and paternal alleles. This makes, for instance, functional interpretation challenging since potentially causative variants identified cannot be assigned to the respective blood group allele background. Another limitation is that indels and structural variants (SVs) cause a shift in signals and thus hamper the readability of the sequences. The length restriction of Sanger sequencing is another limitation, which usually results in neglecting the investigation of non-coding regions of a gene, including the promoter region, enhancer elements and transcription factor binding sites. The limited length of the sequenced region also hinders the detection of SVs. Large deletions, for example, may simply go unnoticed with Sanger sequencing in case of an unaffected second allele in the background, which gets sequenced (allelic dropout). Some of these limitations would be mitigated when using short-read sequencing technologies, but the short read length still hampers full-gene haplotype reconstruction and the analysis of SVs, alongside the lack of scalability to single-sample diagnostics.

The long-read sequencing technology by Oxford Nanopore Technologies (ONTs) represents a promising solution to overcome the limitations of the aforementioned sequencing approaches. It allows sequencing long reads of blood group genes at the kilobase (kb) scale in a low throughput setting at high quality [8]. This greatly facilitates haplotype reconstruction and the inclusion of introns and proximal regulatory elements [5,9]. Moreover, it lowers the risk of misinterpreting loss of heterozygosity caused by allelic dropout as the long reads are much more likely to cover the respective breakpoint sites. Previous constraints of ONT sequencing, in particular the low single-read accuracy and the need for advanced bioinformatics skills [10–12], have been greatly alleviated by the latest developments.

In this study, we propose a cost-effective and reliable approach based on ONT sequencing to resolve the Kidd blood group (JK) in samples with discordant serological and genotypic findings. We used the JK system as a proof-of-concept due to a combination of its clinical relevance and its low antigen diversity, enabling straightforward phenotyping. In fact, this blood group system only comprises three antigens: the antithetical JK1 and JK2 antigens, also known as Jk(a) and Jk(b), as well as the high-prevalence antigen JK3 [13]. Antibody formation by alloimmunization is particularly common in patients with sickle cell disease [14] and JK antibodies are a cause of delayed hemolytic transfusion reactions as well as hemolytic diseases of the newborn [15]. The *SLC14A1* gene, spanning over ~30 kb on chromosome 18, consists of 10 exons and translates into a 389-amino acid glycoprotein. The International Society of Blood Transfusion (ISBT) currently recognizes over 60 different *JK* alleles, of which most represent null alleles caused by missense or nonsense single-nucleotide variants (SNVs) [13].

Here, we investigated all unresolved genotype–phenotype discrepant cases collected over seven years of routine high-throughput donor genotyping at Blood Transfusion Service Zurich (Switzerland) using Matrix-Assisted Laser Desorption Ionization–Time-of-Flight Mass Spectrometry (MALDI-TOF MS) [7,16]. The workflow was based on long-range PCR (LR-PCR) amplification of *SLC14A1* and the subsequent barcoded nanopore-sequencing of amplicons. As only the LR-PCR primer design was gene-specific, our workflow can easily be customized to target other blood group genes. Overall, our results demonstrated that ONT sequencing can effectively and accurately resolve genotype–phenotype discrepancies and also those unresolved by Sanger sequencing. This cost-effective approach holds great promise in general for accurate blood group determination in challenging cases.

2. Materials and Methods

A graphical overview of how discrepant samples were identified and processed is provided in Figure 1.
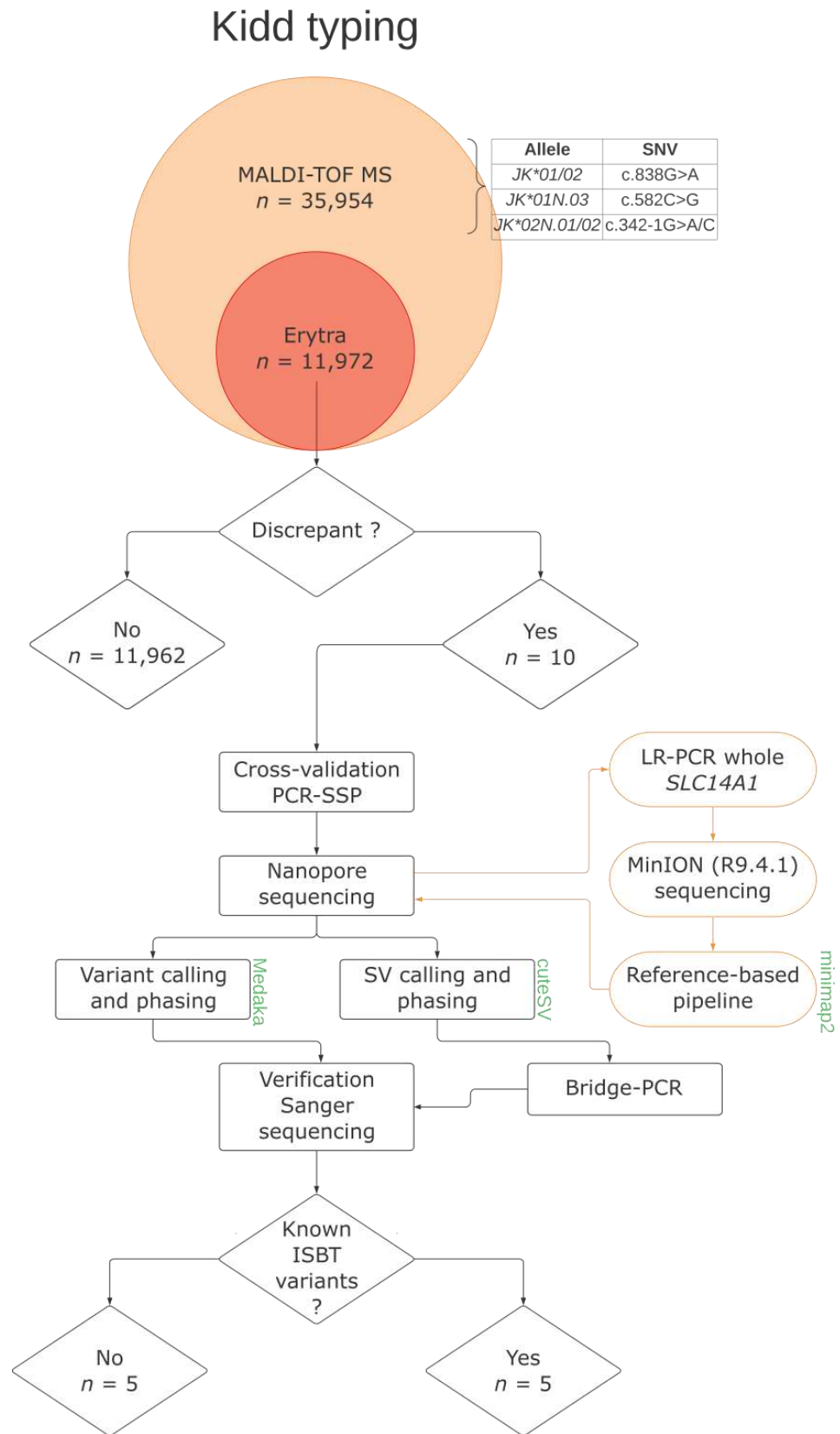
Figure 1. Flowchart depicting identification and processing of genotype–phenotype discrepancies. Main bioinformatics tools are provided in green. MALDI-TOF MS stands for Matrix-Assisted Laser Desorption Ionization–Time-of-Flight Mass Spectrometry; PCR-SSP for Sequence-Specific-Priming PCR; LR-PCR for Long-Range PCR; SV for Structural Variant; and ISBT for International Society of Blood Transfusion. Erytra is an automated gel cards system used to serologically phenotype donors.

2.1. Routine High-Throughput Donor Geno- and Phenotyping

The genomic DNA of blood donors was extracted on a Chemagic MSM I extraction robot (Chemagen, Perkin Elmer, Waltham, MA, USA). High-throughput genotyping of 46 blood group antigens of 35,954 donor samples collected in the years 2015 to 2021 was performed via MALDI-TOF MS, as described previously [16,17]. Genotyping included *JK* alleles of interest in a Swiss and international donor pool. Specifically, *JK*01* and *JK*02* were discriminated by the SNV c.838G>A. The null alleles *JK*01N.03* (found in Swiss families) [18], *JK*02N.01* (predominantly detected in Asian populations [19]) and *JK*02N.02* were typed using the c.582C>G, c.342-1G>A and c.342-1G>C variants [20], respectively. All coding DNA positions provided are based on reference transcript NM_015865.7.

Of all donors genotyped using MALDI-TOF MS, 11,972 donors were additionally serologically phenotyped on a follow-up donation using an Erytra automated gel cards system (Grifols, Barcelona, Spain) for seven blood groups, among which JK was present. In the case of genotype–phenotype discrepancy between MALDI-TOF MS genotyping and phenotyping on the Erytra, the sample was analyzed further using other standard serological techniques and commercial PCR-SSP kits (sequence-specific-priming PCR; inno-train, Kronberg, Germany). For this, DNA was manually re-extracted from the follow-up donation using the Nucleon BACC 3 kit (Gen-Probe Life Sciences, San Diego, CA, USA). DNA concentrations were measured using Nanodrop 2000 (Thermo Fisher Scientific, Waltham, MA, USA).

2.2. Nanopore Sequencing of Genotype–Phenotype Discrepancies

In the case of confirmed genotype–phenotype discrepancy, the entire coding region of the *SLC14A1* gene (~24 kb, exon 3 to 10) was amplified and sequenced using ONT. Amplifications were performed with two LR-PCRs. The first fragment (12,981 bp) was amplified using the forward primer 5′-TGCCACTTGAGTGTTTTCATTTGATGCTGC-3′ and reverse primer 5′-CACTATCCCTCCTCCTTTTTGTTCCCAAGC-3′, located in intron 2 and intron 8, respectively (Figure 2a). PCR primers for the second fragment (13,112 bp) were 5′-AAGTGACGTCCCCTCTCTGAGAGCATTAAA-3′ (forward) and 5′-AACATTCTGACAAGTGGCTGGTCCTAGAGA-3′ (reverse), located in intron 8 and ~400 bp downstream of the gene, respectively (Figure 2a). To facilitate subsequent phasing, both LR-PCR fragments were designed to have a large overlap (~1550 bp). PCR amplifications were performed in duplicates using the PrimeSTAR GXL polymerase (Takara Bio, Kusatsu, Japan) and 250 ng of genomic DNA per reaction, following the manufacturer's protocol. To increase amplification success, we added 1 M of Betaine enhancer (VWR, Radnor, PA, USA) per reaction. The PCR's profile followed a 2-step approach with a 10 s denaturation step at 95 °C and a 10 min extension step at 68 °C for 30 cycles. Amplification success was verified on a 0.8% agarose gel stained with GelRed Nucleic Acid Gel (Biotium, Fremont, CA, USA). After verification, PCR replicates were pooled and purified with 1× Agencourt AMPure XP magnetic beads (Beckman Coulter, Brea, CA, USA). Purified PCR products were eluted in 25 μL sterile H_2O and quantified using a dsDNA broad range assay kit on a Qubit fluorometer 3.0 (both Invitrogen, Waltham, MA, USA). Sequencing libraries were constructed following ONT's 'Amplicon barcoding with Native Barcoding' protocol (version: NBA_9102_v109_revF_09Jul2020). As a starting material, we pooled 50 fmol of both purified amplicons for each sample. After completion of the protocol, 44 fmol of the final barcoded library was loaded on a MinION (R9.4.1) flow cell. Sequencing was stopped after no more pores were active (~72 h).
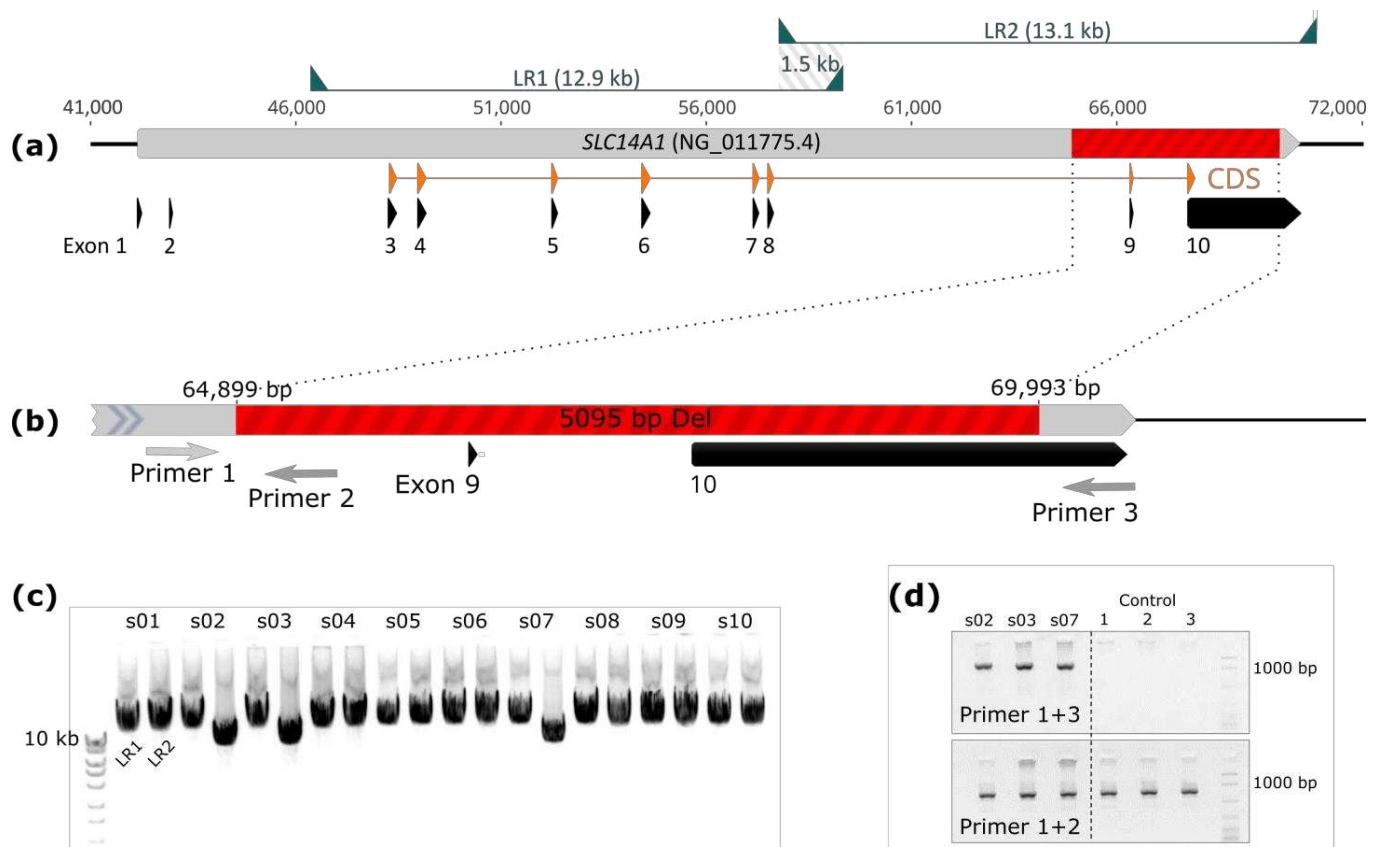
Figure 2. Genetic structure of the *SLC14A1* gene encoding the JK blood group system, and details of the identified large novel deletion. (**a**) *SLC14A1* gene showing exons, coding DNA sequence (CDS), newly identified ~5 kb deletion (red), positions of long-range PCRs (LR1/LR2), as well as overlap between both amplicons. (**b**) Region of ~5 kb deletion with breakpoints and primer positions used for bridge-PCR. (**c**) Gel electrophoresis of long-range PCR products (LR1 and LR2) of all samples. (**d**) Gel electrophoresis of bridge-PCR products for the three samples with novel deletion and wild-type control reactions.

2.3. Nanopore Bioinformatics Processing

Raw nanopore reads were demultiplexed and basecalled using ONT's *Guppy* (v.4.4.2). After demultiplexing, reads in FASTQ format were filtered based on expected length and observed read length distributions. More specifically, we filtered out all reads shorter than 12,500 bp and longer than 13,600 bp. For three samples (s02, s03 and s07) harboring a second shorter peak in read length distribution (~8000 bp; Figure 2c), we additionally kept all reads having a length between 7500 bp and 8500 bp. After size selection, *PoreChop* (v.0.2.3) was used to trim the remaining adapter and barcode sequences.

To reduce computational burden, reads were randomly downsampled to 1000 reads per fragment using *seqkit* (v.0.15). To differentiate between fragment 1 and 2, which were very similar in read length, all size-selected reads were first mapped to the *SLC14A1* reference sequence (NG_011775.4) using *minimap2* (v.2.17) and separated according to their mapped genomic location. For the three samples showing the secondary shorter peak, mapping showed that reads responsible for this peak partially corresponded to fragment 2. Therefore, we further downsampled to 500 reads for both the long and short fragment 2. Finally, uniformly downsampled datasets were remapped to NG_011775.4 using *minimap2*, allowing the presence of supplementary alignments (i.e., split-read mapping). Mapped reads were sorted, indexed and converted to BAM format using *samtools* (v.1.15).

Variant calling and phasing of called variants were performed using *Medaka* (v.1.2.3). As a consequence of the high coverage, the threshold quality for calling indels and SNVs (default: Q9 and Q8, respectively) was increased to Q10. The presence of SVs was investi-

gated with *cuteSV* (v.1.0.11). Command options were set as recommended for nanopore data. *BCFtools* (v.1.9) was finally used to output phased sequences in the FASTA format.

2.4. Variant Confirmation via Sanger Sequencing and Bridge-PCR

Nanopore sequencing results were confirmed using Sanger sequencing. Specific PCR primers for each newly detected variant were designed and PCRs were carried out using AmpliTaq DNA polymerase (Life Technologies, Carlsbad, CA, USA). Primers and PCR protocols are available on request. Amplicon purification and Sanger sequencing were outsourced to an external company (Microsynth AG, Balgach, Switzerland).

Finally, a bridge-PCR assay was designed to confirm a large deletion found with nanopore sequencing in three samples (Figure 2b). One primer pair was designed to amplify over the deletion using primers located at the flanking regions, with the forward primer 5′-AAAAGCAACTTTGAATTGGAG-3′ being located upstream of the 5′ breakpoint ("Primer 1" in Figure 2b) and the reverse primer 5′-AATTCTTCTTTCACACGTTCC-3′ downstream of the 3′ breakpoint ("Primer 3" in Figure 2b). Another reverse primer 5′-GGGATGTGGCAGACATGGTGG-3′ ("Primer 2" in Figure 2b) was designed to amplify the wild type allele, enabling an assessment of the heterozygosity status of the deletion (Figure 2d). This reverse primer was designed to bind within the region of the deletion and amplify with the upstream primer targeting the 5′ breakpoint. For both primer pairs, three wild type controls were used (Figure 2d).

3. Results

3.1. Observed Genotype–Phenotype Discrepancies

The congruence of JK phenotypes deduced from genotypes of seven years of high-throughput donor screening and serological typing was very high. In detail, of the 35,954 donors genotyped via MALDI-TOF MS, 35,937 (99.95%) passed quality control and were assigned to *JK*01* and/or *JK*02* alleles using the SNV c.838G>A. Among them, 50.43% (*n* = 18,124) were *JK*01/02* heterozygous, 26.30% (*n* = 9452) *JK*01/01* homozygous and 23.27% (*n* = 8361) *JK*02/02* homozygous, as expected under Hardy–Weinberg equilibrium (*p* > 0.05). Out of 11,972 donors with available serological data, we observed seven samples where the null allele causing variants included in the MALDI-TOF MS module explained their null phenotype (*n* = 4 for *JK*01N.03* and *n* = 3 for *JK*02N.01*). There were only 10 discrepant cases (0.08%) in which MALDI-TOF MS genotyping did not agree with the observed phenotype. Phenotypes and genotypes for all 10 discrepant cases were confirmed via additional serological analyses and PCR-SSP.

3.2. Nanopore Sequencing Output

We obtained very high coverage for each amplified fragment in all 10 samples with discordant serology and genotype (Figure 3). Overall, the nanopore sequencing run produced ~1 million reads (≥Q7) with a median PHRED score of 12.7. After demultiplexing and filtering reads according to fragment sizes, the mean number (±standard deviation) of reads per barcode (i.e., sample) was 41,640 (±15,182). The mapping of filtered reads against the *SLC14A1* reference sequence showed a drop in coverage over a length of 5 kb in the second fragment for three samples (Figure 3). For these same samples (s02, s03 and s07), the gel electrophoresis for verifying PCR amplification successes prior to sequencing had also already revealed the presence of shorter fragments in the second LR-PCR (Figure 2c).
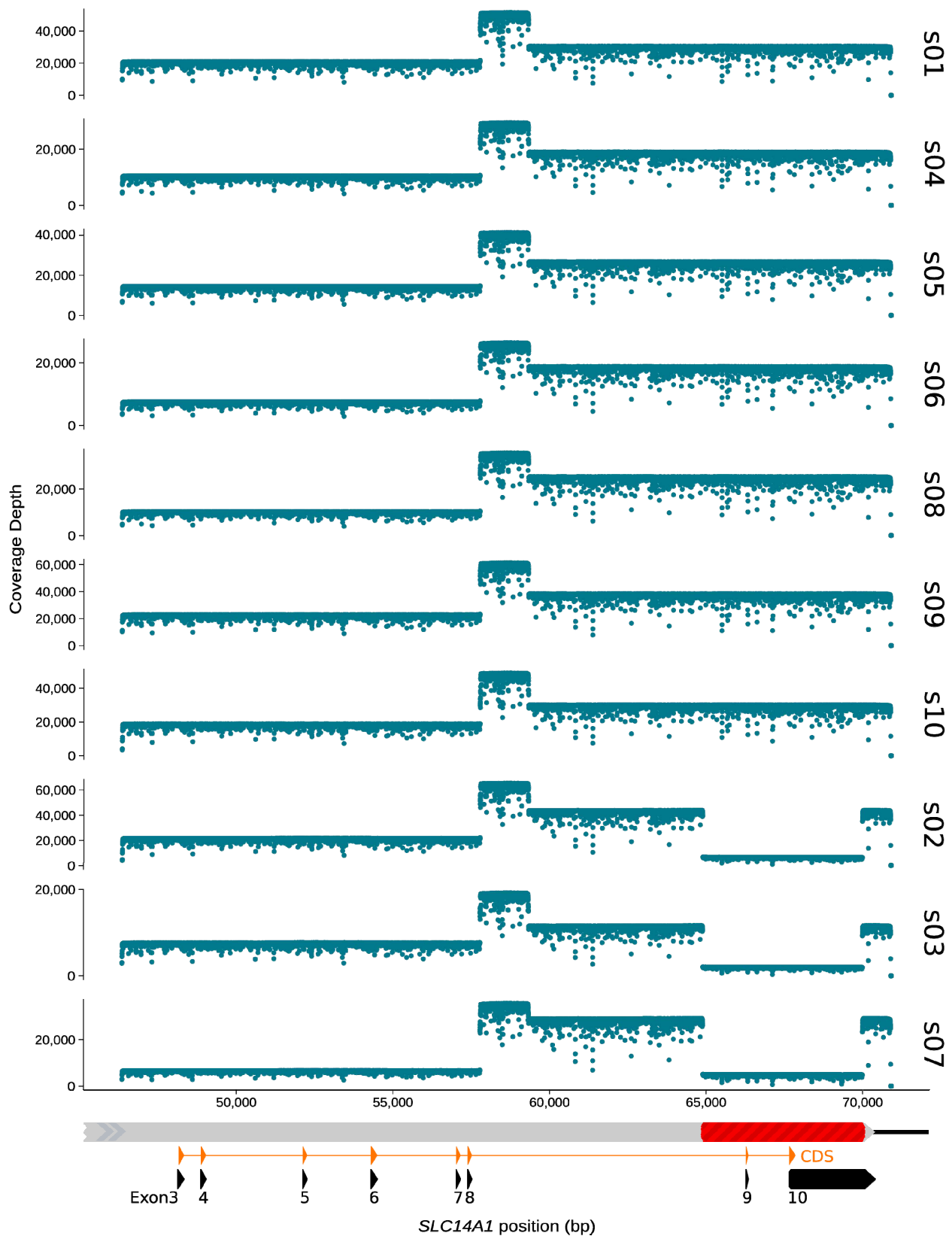
Figure 3. Nanopore sequencing coverage by nucleotide position for each sample. Coverages were computed by mapping length-filtered reads to *SLC14A1* reference sequence (NG_011775.4) using minimap2. For readability, samples showing a drop in coverage in LR2 were grouped together. Graphical representation of mapping location on *SLC14A1* gene is given underneath the coverage plots. Coverage scales are not fixed.

3.3. Resolving Discrepant Cases

A summary of the serology and genetic characterization of the 10 discrepant cases is provided in Table 1 and Supplementary Table S1. In seven cases, we detected SNVs in coding regions that could explain the observed phenotype. In five donors, these variants had previously been reported causing weak (*JK*01W.05*, *JK*02W.03*) and null alleles (*JK*02N.06/08/09*), respectively. In the other two donors, we identified exonic SNVs that had so far not been linked to Kidd phenotypes. In one case, the SNV was located in exon 3 (c.119G>A; NG_011775.4:g.48415G>A) on a *JK*01* allele, altering the protein sequence close to the N-terminus (Gly40Asp). The other novel SNV was located in exon 7 (c.725G>A, NG_011775.4:g.57200G>A) on a *JK*02* allele and corresponded to a missense mutation changing the codon 242 from glycine to glutamic acid (Gly242Glu). Sanger sequencing confirmed the presence of all those SNVs.

Table 1. Summary of serology, genotyping and nanopore sequencing for all ten genotype–phenotype discrepancy cases. Discrepant antigens, based on a comparison of observed with deduced phenotype, as well as causative alleles are highlighted in red. A comprehensive table, including all exonic nucleotide changes, corresponding amino acid modifications and GenBank accession numbers, is provided as Supplementary Material (Table S1).

Sample	Observed Serology	Deduced Phenotype Based on MALDI-TOF MS Genotyping	Nanopore Sequencing	
			Haplotype 1	Haplotype 2
Known weak and null alleles				
s01	Jk(a+^{weak} b−)	Jk(a+b+)	*JK*01W.06*	*JK*02N.08* [†]
s04	Jk(a+b−)	Jk(a+b+)	*JK*01*	*JK*02W.03*
s05	Jk(a−b+)	Jk(a+b+)	*JK*01W.05*	*JK*02* [†]
s06	Jk(a+b−)	Jk(a+b+)	*JK*01*	*JK*02N.06* [†]
s08	Jk(a+^{weak} b−)	Jk(a+b+)	*JK*01W.06*	*JK*02N.09* [†]
Novel null alleles				
s09	Jk(a+b−)	Jk(a+b+)	*JK*01*	*JK*02(c.725G>A)Null* [§†]
s10	Jk(a−b+)	Jk(a+b+)	*JK*01(c.119G>A)Null* [§]	*JK*02* [†]
Novel structural variant				
s02	Jk(a−b+)	Jk(a+b+)	*JK*01(Ex9_10del)Null* [§]	*JK*02* [†]
s03	Jk(a−b+)	Jk(a+b+)	*JK*01(Ex9_10del)Null* [§]	*JK*02* [†]
s07	Jk(a−b+)	Jk(a+b+)	*JK*01(Ex9_10del)Null* [§]	*JK*02* [†]

[§] Novel alleles not yet described. [†] Allele harbors at least one additional synonymous SNV not reported in corresponding ISBT allele (see Supplementary Table S1).

Three samples did not show any promising candidates when calling SNVs and short indels. However, SV calling identified a large 5095 bp deletion in all three samples (Figure 2a). Using NG_011775.4 as reference, *cuteSV* suggested breakpoints located before position 64,898 bp in intron 8 and after position 69,992 bp in exon 10 (NG_011775.4:g.64898_69992del; NM_015865.7:c.947-1453_*2066del, Figure 2b). Bridge-PCR followed by Sanger sequencing confirmed the exactness of these breakpoint positions (Figure 2d). In all three samples, the deletion was located on a *JK*01* allele (c.838G) which could be inferred serologically since donors were *JK*01*/*JK*02* heterozygotes and phenotypically Jk(a−b+). Additionally, we could genetically phase the deletion to the c.838G allele.

4. Discussion

By employing a combination of LR-PCR and long-read nanopore sequencing, we successfully resolved all genotype–phenotype discrepancies in the Kidd blood group system observed over seven years of routine high-throughput donor genotyping. Our developed protocol allowed sequencing the complete coding region of the *SLC14A1* gene, including introns, as haplotypes. This greatly facilitated the identification of the underlying genetic basis for the observed genotype–phenotype discrepancies and produced high-quality haplotype reference sequences for newly described *JK* alleles.

The majority of *JK* weak and null alleles listed in the ISBT tables are known to occur at very low frequencies across different populations [4,21,22], supporting the low number

of discrepancies identified in our study. Specifically, nanopore sequencing allowed us to identify the following known SNVs: c.130G>A (rs2298720), c.191G>A (rs114362217), c.742G>A (rs763095261), c.871T>C (rs78242949) and c.956C>T (rs565898944). The SNVs could be unambiguously phased to the respective *JK*01/02* background allele, thus resulting in *JK*02W.03*, *JK*02N.09*, *JK*01W.05*, *JK*02N.06* and *JK*02N.08* alleles, respectively (Table 1; Supplementary Table S1). All these variants show minor allele frequencies (MAFs) in large-scale sequencing projects [23] that are in close agreement with the exceeding rarity observed in our study (MAFs < 0.001), with the exception of rs2298720, which is a common SNV (see discussion below). It is important to acknowledge that these variants (alleles) were only detected in *JK*01/02* heterozygotes of our donor population since discordant serology would not have appeared in homozygotes due to the masking effect of the wildtype allele.

Our workflow did not aim to detect variants that determine weak alleles, given that the transfusion recommendations for carriers of weak JK phenotypes are not different from those with normally expressed antigens. Consequently, weak agglutination levels were not considered as discrepant as long as in agreement with *JK*01/02* genotyping. After over seven years of MALDI-TOF MS high-throughput donor genotyping, only two known weak alleles have been flagged as suggestively causative for discrepant cases, i.e., resulting in the absence of observable levels of agglutination with respective anti-Jk antibodies. One was the aforementioned variant c.742G>A, defining the *JK*01W.05* allele. This allele is currently listed as 'weak' but has already been reported to cause null phenotypes by others [24]. The other variant was c.130G>A, defining the *JK*02W.03* allele in the presence of the synonymous c.588A>G, or *JK*02W.04* in its absence. This variant is a common SNV (MAF > 0.05; rs2298720), which also defines the most frequent weak allele on the *JK*01* background (*JK*01W.01*) along with several other weak alleles [13]. Since we did not observe further discrepancies pointing to weak alleles harboring this variant, there is little evidence that expression levels of *JK*02W.03* commonly fall below the detection threshold, pointing to an exceptional case reported here. However, without phenotyping methods offering higher sensitivity, we cannot conclusively determine whether expression levels caused by these two presumably weak alleles were just below our detection threshold or if those two alleles indeed caused true null phenotypes in our samples. Efforts for in-depth phenotyping using absorption/elution techniques are ongoing.

Among the remaining cases that displayed genotype–phenotype discrepancies in our study, two were found with exonic SNVs likely defining new null alleles not yet listed by the ISBT [13]. One of these variants, c.725G>A (rs1197896884), is located in exon 7 and has been previously reported in two Europeans in the gnomAD [25] sequence collection of more than 125,000 sequenced exomes and whole genomes. The other variant, c.119G>A, is entirely novel. Notably, a singleton with a different base change (G>C) at the same position, resulting in a different amino acid change (Gly40Ala), has been found in the gnomAD [25] sequence collection (rs1253002394), suggesting a tri-allelic SNV. Furthermore, it is noteworthy that a yet another change in the same amino acid (Gly40Ser, c.118G>A, rs145283450) is known to lead to the presence of null alleles (*JK*01N.17*, *JK*02N.21*) [13,26].

Interestingly, all sequences of *JK*02* alleles examined in this study exhibited an additional synonymous SNV, c.588A>G (p.Pro196Pro). In the current JK blood group ISBT table (v8.1), only three alleles (*JK*01W.06*, *JK*02W.03* and *JK*01N.20*) incorporate this variant, alongside SNVs causing alterations in the amino acid chain. Of these, *JK*01W.06* and *JK*02W.03* possess sister alleles lacking c.588A>G (*JK*01W.01* and *JK*02W.04*, respectively). Notably, in Europeans [27], the linkage disequilibrium between the SNV discriminating *JK*01/02* (rs1058396; c.838G>A) and c.588A>G (rs2298718) is high ($R^2 = 0.72$), with the c.838A allele always carrying c.588G, i.e., a complete absence of the A-A haplotype (D' = 1). Consequently, one would anticipate a quasi-total prevalence of c.588A>G on *JK*02* alleles. Since this variation does not affect the protein sequence, it may have often been overlooked during the submission of novel alleles.

Our most remarkable finding in this study was the discovery of a large structural variant that was responsible for three of the observed genotype–phenotype discrepancies.

Our nanopore sequencing strategy revealed a large deletion of approximately 5 kb spanning from exon 9 to 10 on the *JK*01* allele of three unrelated donors. This deletion represents, to the best of our knowledge, the largest SV ever documented within the JK blood group system. Currently, the Kidd ISBT allele table only encompasses one deletion of ~1.6 kb defining the *JK*01N.01* allele. First reported in Tunisian women [28], it has been sporadically reported in several other populations [18,29]. This deletion (c.1_341del) spans over exons 3 and 4 of the *SLC14A1* gene [28] and results in the lack of translation of the Kidd protein as the start codon is located in exon 3. Considering the complete absence of exon 9 and a portion of exon 10 in alleles carrying our newly discovered large deletion, it is highly probable that the resulting protein is also not functional. However, as in the case of the two aforementioned SNVs causing *JK*02W.03* and *JK*01W.05*, additional in-depth analyses using adsorption–elution or flow cytometric experiments [30] will be required to validate a true null phenotype of the three samples harboring the large deletion.

The apparent lack of described SVs in the Kidd system extends to almost all blood groups [13]. With the continuing accumulation of high-coverage whole genome sequencing datasets and improved algorithms for SV discovery, it has become clear that SVs appear to be more common in the human genome than previously thought. Depending on the population targeted and the sequencing technology used, between 7500 and 22,600 SVs per genome were observed [31,32]. Additionally, SVs are responsible for an estimated ~30% of rare heterozygous (MAF < 1%) gene inactivation events per individual [31]. Therefore, the extremely rare reports of SVs in blood group genes could simply be linked to the hitherto lack of adequate tools for their detection.

This study serves as a compelling example for illustrating the power of nanopore sequencing in resolving complex SVs that would have remained elusive using conventional sequencing methods. Indeed, using Sanger sequencing alone, the sequencing results for exon 9 and 10 of the three samples harboring the deletion would have appeared as homozygous since only the wildtype allele would have been amplified and sequenced. In contrast, nanopore sequencing enabled us to precisely identify the breakpoints, which were subsequently verified using bridge-PCR.

With the continuing reduction in cost and the increase in accessible and easy-to-use bioinformatic tools, we forecast that long-read sequencing technologies will be more and more frequently used to overcome challenging diagnostic cases in blood group genetics.

References

1. Nurk, S.; Koren, S.; Rhie, A.; Rautiainen, M.; Bzikadze, A.V.; Mikheenko, A.; Vollger, M.R.; Altemose, N.; Uralsky, L.; Gershman, A.; et al. The Complete Sequence of a Human Genome. *Science* **2022**, *376*, 44–53. [CrossRef] [PubMed]

2. Lane, W.J.; Gleadall, N.S.; Aeschlimann, J.; Vege, S.; Sanchis-Juan, A.; Stephens, J.; Cone Sullivan, J.; Mah, H.H.; Aguad, M.; Smeland-Wagman, R.; et al. Multiple GYPB Gene Deletions Associated with the U−Phenotype in Those of African Ancestry. *Transfusion* **2020**, *60*, 1294–1307. [CrossRef] [PubMed]

3. Zhang, Z.; An, H.H.; Vege, S.; Hu, T.; Zhang, S.; Mosbruger, T.; Jayaraman, P.; Monos, D.; Westhoff, C.M.; Chou, S.T. Accurate Long-Read Sequencing Allows Assembly of the Duplicated RHD and RHCE Genes Harboring Variants Relevant to Blood Transfusion. *Am. J. Hum. Genet.* **2022**, *109*, 180–191. [CrossRef]

4. Montemayor, C.; Simone, A.; Long, J.; Montemayor, O.; Delvadia, B.; Rivera, R.; Lewis, K.L.; Shahsavari, S.; Gandla, D.; Dura, K.; et al. An Open-source Python Library for Detection of Known and Novel Kell, Duffy and Kidd Variants from Exome Sequencing. *Vox Sang.* **2021**, *116*, 451–463. [CrossRef] [PubMed]

5. Thun, G.A.; Gueuning, M.; Sigurdardottir, S.; Meyer, E.; Gourri, E.; Schneider, L.; Merki, Y.; Trost, N.; Neuenschwander, K.; Engström, C.; et al. Novel Regulatory Variant in ABO Intronic RUNX1 Binding Site Inducing A3 Phenotype. *bioRxiv* **2023**. [CrossRef] [PubMed]

6. Isa, K.; Takada, S.; Takeda, H.; Tsuneyama, H.; Ogasawara, K.; Takahashi, D.; Miyazaki, T.; Miyata, S.; Satake, M. Two New *JK* Silencing Alleles Identified by Single Molecule Sequencing with 20-Kb Long-reads. *Transfusion* **2023**, *63*, 1441–1446. [CrossRef] [PubMed]

7. Meyer, S.; Vollmert, C.; Trost, N.; Brönnimann, C.; Gottschalk, J.; Buser, A.; Frey, B.M.; Gassner, C. High-Throughput Kell, Kidd, and Duffy Matrix-Assisted Laser Desorption/Ionization, Time-of-Flight Mass Spectrometry-Based Blood Group Genotyping of 4000 Donors Shows Close to Full Concordance with Serotyping and Detects New Alleles. *Transfusion* **2014**, *54*, 3198–3207. [CrossRef] [PubMed]

8. Thun, G.A.; Gueuning, M.; Mattle-Greminger, M. Long-Read Sequencing in Blood Group Genetics. *Transfus. Med. Hemotherapy* **2023**, *50*, 184–197. [CrossRef]

9. Gueuning, M.; Thun, G.A.; Wittig, M.; Galati, A.-L.; Meyer, S.; Trost, N.; Gourri, E.; Fuss, J.; Sigurdardottir, S.; Merki, Y.; et al. Haplotype Sequence Collection of ABO Blood Group Alleles by Long-Read Sequencing Reveals Putative A1-Diagnostic Variants. *Blood Adv.* **2023**, *7*, 878–892. [CrossRef]

10. Wang, Y.; Zhao, Y.; Bollas, A.; Wang, Y.; Au, K.F. Nanopore Sequencing Technology, Bioinformatics and Applications. *Nat. Biotechnol.* **2021**, *39*, 1348–1365. [CrossRef]

11. Makałowski, W.; Shabardina, V. Bioinformatics of Nanopore Sequencing. *J. Hum. Genet.* **2020**, *65*, 61–67. [CrossRef] [PubMed]

12. Olson, N.D.; Wagner, J.; McDaniel, J.; Stephens, S.H.; Westreich, S.T.; Prasanna, A.G.; Johanson, E.; Boja, E.; Maier, E.J.; Serang, O.; et al. PrecisionFDA Truth Challenge V2: Calling Variants from Short and Long Reads in Difficult-to-Map Regions. *Cell Genomics* **2022**, *2*, 100129. [CrossRef] [PubMed]

13. International Society of Blood Transfusion. Available online: https://www.isbtweb.org/isbt-working-parties/rcibgt.html (accessed on 10 November 2023).

14. da Cunha Gomes, E.G.; Machado, L.A.F.; de Oliveira, L.C.; Neto, J.F.N. The Erythrocyte Alloimmunisation in Patients with Sickle Cell Anaemia: A Systematic Review. *Transfus. Med.* **2019**, *29*, 149–161. [CrossRef]

15. Dean, L. Chapter10: The Kidd Blood Group. In *Blood Groups and Red Cell Antigens*; Bethesda National Library of Medicine: Bethesda, MD, USA, 2005.

16. Gassner, C.; Meyer, S.; Frey, B.M.; Vollmert, C. Matrix-Assisted Laser Desorption/Ionisation, Time-of-Flight Mass Spectrometry–Based Blood Group Genotyping—The Alternative Approach. *Transfus. Med. Rev.* **2013**, *27*, 2–9. [CrossRef] [PubMed]

17. Meyer, S.; Trost, N.; Frey, B.M.; Gassner, C. *Parallel Donor Genotyping for 46 Selected Blood Group and 4 Human Platelet Antigens Using High-Throughput MALDI-TOF Mass Spectrometry BT-Molecular Typing of Blood Cell Antigens*; Bugert, P., Ed.; Springer: New York, NY, USA, 2015; pp. 51–70; ISBN 978-1-4939-2690-9.

18. Irshaid, N.M.; Eicher, N.I.; Hustinx, H.; Poole, J.; Olsson, M.L. Novel Alleles at the JK Blood Group Locus Explain the Absence of the Erythrocyte Urea Transporter in European Families. *Br. J. Haematol.* **2002**, *116*, 445–453. [CrossRef] [PubMed]

19. Irshaid, N.M.; Henry, S.M.; Olsson, M.L. Genomic Characterization of the Kidd Blood Group Gene: Different Molecular Basis of the Jk(a-b-) Phenotype in Polynesians and Finns. *Transfusion* **2000**, *40*, 69–74. [CrossRef] [PubMed]

20. Hamilton, J.R. Kidd Blood Group System: A Review. *Immunohematology* **2015**, *31*, 29–35. [CrossRef]

21. Dinardo, C.L.; Oliveira, T.G.M.; Kelly, S.; Ashley-Koch, A.; Telen, M.; Schmidt, L.C.; Castilho, S.; Melo, K.; Dezan, M.R.; Wheeler, M.M.; et al. Diversity of Variant Alleles Encoding Kidd, Duffy, and Kell Antigens in Individuals with Sickle Cell Disease Using Whole Genome Sequencing Data from the NHLBI TOPMed Program. *Transfusion* **2021**, *61*, 603–616. [CrossRef]

22. Vorholt, S.M.; Lenz, V.; Just, B.; Enczmann, J.; Fischer, J.C.; Horn, P.A.; Zeiler, T.A.; Balz, V. High-Throughput Next-Generation Sequencing of the Kidd Blood Group: Unexpected Antigen Expression Properties of Four Alleles and Detection of Novel Variants. *Transfus. Med. Hemotherapy* **2023**, *50*, 51–65. [CrossRef]

23. Sherry, S.T.; Ward, M.; Sirotkin, K. DbSNP-Database for Single Nucleotide Polymorphisms and Other Classes of Minor Genetic Variation. *Genome Res.* **1999**, *9*, 677–679. [CrossRef]

24. Gaur, L.; Posadas, J.; Teraaur, G.; Hamura, A.; Gile, P.; Nakaya, S. Molecular Diversity of the JK Null Phenotype. *Vox Sang.* **2010**, *99* (Suppl. S1), 371.

25. Chen, S.; Francioli, L.C.; Goodrich, J.K.; Collins, R.L.; Kanai, M.; Wang, Q.; Alföldi, J.; Watts, N.A.; Vittal, C.; Gauthier, L.D.; et al. A Genome-Wide Mutational Constraint Map Quantified from Variation in 76,156 Human Genomes. *bioRxiv* **2022**. [CrossRef]

26. Ramsey, G.; Sumugod, R.D.; Lindholm, P.F.; Zinni, J.G.; Keller, J.A.; Horn, T.; Keller, M.A. A Caucasian JK*A/JK*B Woman with Jk(A+b-) Red Blood Cells, Anti-Jkb, and a Novel JK*B Allele c.1038delG. *Immunohematology* **2016**, *32*, 91–95. [CrossRef]

27. Auton, A.; Abecasis, G.R.; Altshuler, D.M.; Durbin, R.M.; Abecasis, G.R.; Bentley, D.R.; Chakravarti, A.; Clark, A.G.; Donnelly, P.; Eichler, E.E.; et al. A Global Reference for Human Genetic Variation. *Nature* **2015**, *526*, 68–74. [CrossRef]

28. Lucien, N.; Chiaroni, J.; Cartron, J.-P.; Bailly, P. Partial Deletion in the JK Locus Causing a Jknull Phenotype. *Blood* **2002**, *99*, 1079–1081. [CrossRef]

29. Wester, E.S.; Johnson, S.T.; Copeland, T.; Malde, R.; Lee, E.; Storry, J.R.; Olsson, M.L. Erythroid Urea Transporter Deficiency Due to Novel JKnull Alleles. *Transfusion* **2008**, *48*, 365–372. [CrossRef]

30. Liwski, R.; Clarke, G.; Cheng, C.; Abidi, S.S.R.; Abidi, S.R.; Quinn, J.G. Validation of a Flow-cytometry-based Red Blood Cell Antigen Phenotyping Method. *Vox Sang.* **2023**, *118*, 207–216. [CrossRef]

31. Collins, R.L.; Brand, H.; Karczewski, K.J.; Zhao, X.; Alföldi, J.; Francioli, L.C.; Khera, A.V.; Lowther, C.; Gauthier, L.D.; Wang, H.; et al. A Structural Variation Reference for Medical and Population Genetics. *Nature* **2020**, *581*, 444–451. [CrossRef]

32. Beyter, D.; Ingimundardottir, H.; Oddsson, A.; Eggertsson, H.P.; Bjornsson, E.; Jonsson, H.; Atlason, B.A.; Kristmundsdottir, S.; Mehringer, S.; Hardarson, M.T.; et al. Long-Read Sequencing of 3622 Icelanders Provides Insight into the Role of Structural Variants in Human Diseases and Other Traits. *Nat. Genet.* **2021**, *53*, 779–786. [CrossRef] [PubMed]