**ORIGINAL RESEARCH ARTICLE**

# Resolving genotype-phenotype discrepancies of the Kidd blood group system using long-read nanopore sequencing

**Authors:** M Gueuning[1], GA Thun[1], N Trost[2], L Schneider[2], S Sigurdardottir[2], C Engström[3], N Larbes[3], Y Merki[2], BM Frey[1,2,3], C Gassner[4], S Meyer[2]*, MP Mattle-Greminger[1]*

**Affiliations**

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

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

[3]Department of Immunohematology, Blood Transfusion Service Zurich, Swiss Red Cross, Schlieren, Switzerland

[4]Institute of Translational Medicine, Private University in the Principality of Liechtenstein, Triesen, Liechtenstein

**Correspondence**

Morgan Gueuning, Department of Research and Development, Blood Transfusion Service Zurich, Swiss Red Cross, Rütistrasse 19, 8952 Schlieren, Switzerland, Phone +41 58 272 52 41, E-mail: m.gueuning@zhbsd.ch

*These authors contributed equally to this work

## Abstract

Due to substantial improvement 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 Blood Transfusion Service Zurich. Discrepant results from standard serological typing and donor genotyping were confirmed by 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) haplotype, 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

## 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 to 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 identified potentially causative variants 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. 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, beside the lack of scalability to single sample diagnostics.

The long-read sequencing technology by Oxford Nanopore Technologies (ONT) represents a promising solution to overcome limitations of 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 latest developments.

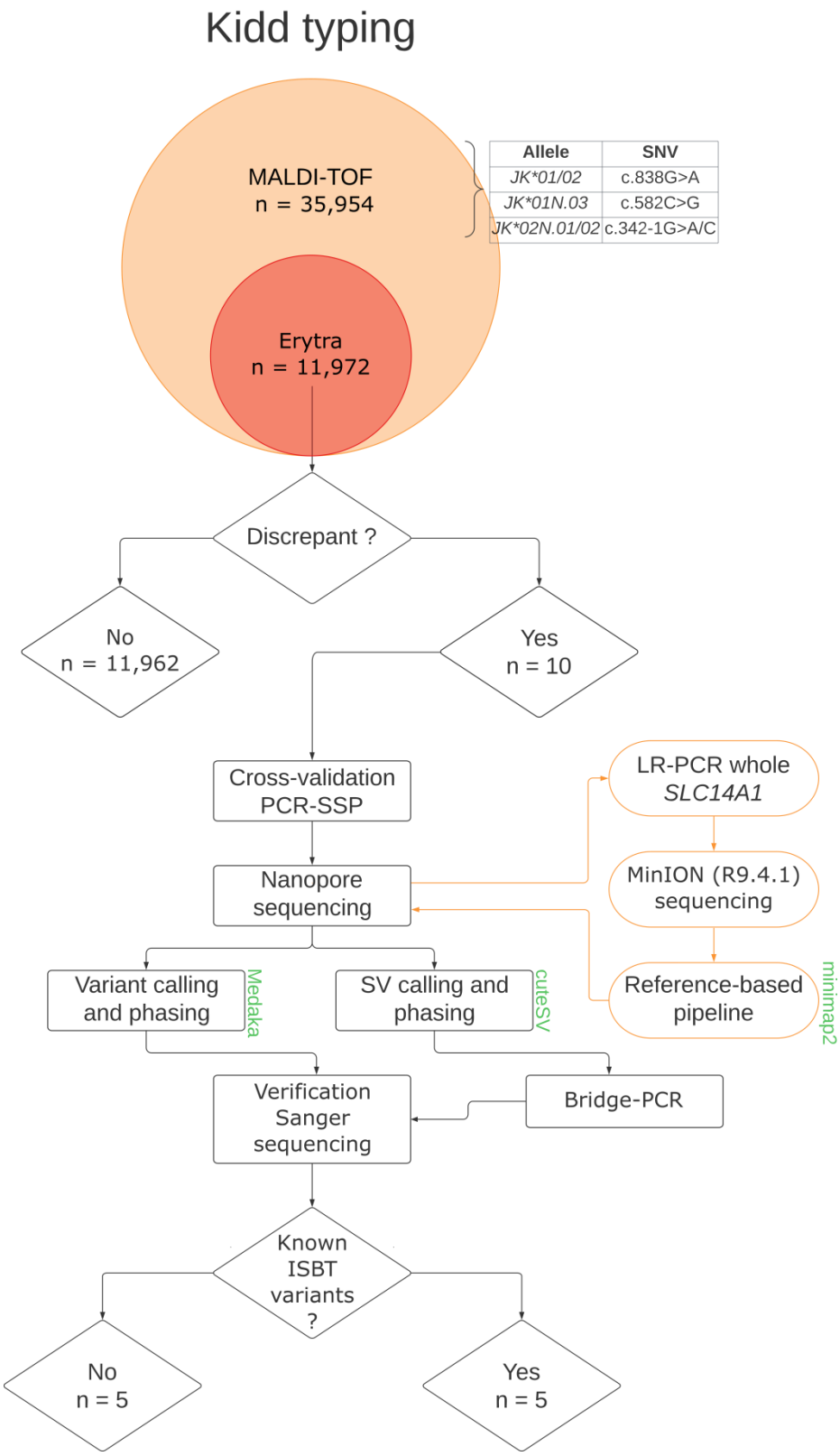
In this study, we propose a cost-effective and reliable approach based on ONT sequencing to resolve the Kidd blood group (JK) in samples with discordant serological and genotypic findings. We used the JK system as a proof-of-concept due to a combination of its clinical relevance and its low antigen diversity, 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 disease 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 *SLC12A1* and 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, also those unresolved by Sanger sequencing. The cost-effective approach holds great promise in general for accurate blood group determination in challenging cases.

## Material 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.




**Routine high-throughput donor geno- and phenotyping**

108 Genomic DNA of blood donors was extracted on a Chemagic MSM I extraction robot (Chemagen,
109 Perkin Elmer). High-throughput genotyping of 46 blood group antigens of 35,954 donor samples
110 collected in the years 2015 to 2021 was performed by MALDI-TOF MS as described previously [16,17].
111 Genotyping included *JK* alleles of interest in a Swiss and international donor pool. Specifically, *JK*01*
112 and *JK*02* were discriminated by the SNV c.838G>A, the null alleles *JK*01N.03* (found in Swiss families)
113 [18], *JK*02N.01* (predominantly detected in Asian populations [19]), and *JK*02N.02* were typed using
114 the c.582C>G, c.342-1G>A, and c.342-1G>C variants [20], respectively. All coding DNA positions
115 provided are based on reference transcript NM_015865.7.

116 Of all donors genotyped by MALDI-TOF MS, 11,972 donors were additionally serologically phenotyped
117 on a follow-up donation using Erytra automated system (Grifols) for seven blood groups, among which
118 was JK. In case of genotype-phenotype discrepancy between MALDI-TOF MS genotyping and
119 phenotyping on the Erytra, the sample was analyzed further using other standard serological
120 techniques and commercial PCR-SSP kits (sequence-specific-priming PCR; inno-train). For this, DNA
121 was manually re-extracted from the follow-up donation using the Nucleon BACC 3 kit (Gen-Probe Life
122 Sciences). DNA concentrations were measured using Nanodrop 2000 (Thermo Fisher Scientific).

123 **Nanopore sequencing of genotype-phenotype discrepancies**

124 In case of confirmed genotype-phenotype discrepancy, the entire coding region of the *SLC14A1* gene
125 (~24 kb, exon 3 to 10) was amplified and sequenced using ONT. Amplifications were done with two
126 LR-PCRs. The first fragment (12,981 bp) was amplified using the forward primer 5'-
127 TGCCACTTGAGTGTTTTCATTTGATGCTGC-3' and reverse primer 5'-
128 CACTATCCCTCCTCCTTTTTGTTCCCAAGC-3'. PCR primers for the second fragment (13,112 bp) were 5'-
129 AAGTGACGTCCCCTCTCTGAGAGCATTAAA-3' (forward) and 5'-
130 AACATTCTGACAAGTGGCTGGTCCTAGAGA-3' (reverse). To facilitate subsequent phasing, both LR-PCR
131 fragments were designed to have a large overlap (~1,550 bp). PCR amplifications were done in
132 duplicates using the PrimerSTAR GXL polymerase (Takara Bio) and 250 ng of genomic DNA per
133 reaction, following the manufacturer's protocol. To increase amplification success, we added 1 M of
134 Betaine enhancer (VWR) per reaction. The PCR's profile followed a 2-step approach with a 10 second
135 denaturation step at 95°C and a 10 minute extension step at 68°C for 30 cycles. Amplification success
136 was verified on a 0.8% agarose gel stained with GelRed Nucleic Acid Gel (Biotium). After verification,
137 PCR replicates were pooled and purified with 1x Agencourt AMPure XP magnetic beads (Beckman
138 Coulter). Purified PCR products were eluted in 25 µl sterile $H_2O$ and quantified using a dsDNA broad
139 range assay kit on a Qubit fluorometer 3.0 (both Invitrogen). Sequencing libraries were constructed

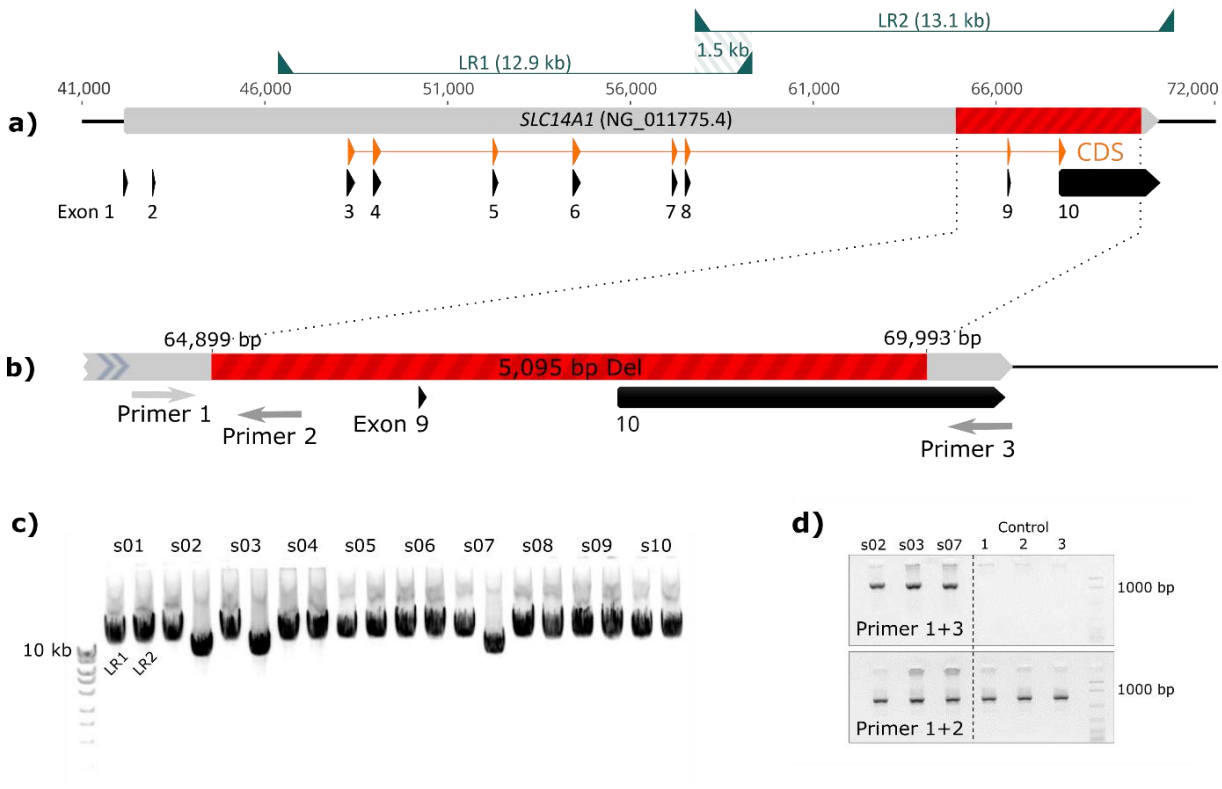
140 following ONT's 'Amplicon barcoding with Native Barcoding' protocol (version:
141 NBA_9102_v109_revF_09Jul2020). As starting material, we pooled 50 fmol of both purified amplicons
142 for each sample. After completion of the protocol, 44 fmol of the final barcoded library was loaded on
143 a MinION (R9.4.1) flow cell. Sequencing was stopped after no more pores were active (~72 h).

144 **Nanopore bioinformatics processing**

145 Raw nanopore reads were demultiplexed and basecalled using ONT's *Guppy* (v.4.4.2). After
146 demultiplexing, reads in FASTQ format were filtered based on expected length and observed read
147 length distributions. More specifically, we filtered out all reads shorter than 12,500 bp and longer than
148 13,600 bp. For three samples (s02, s03 and s07) harboring a second shorter peak in read length
149 distribution (~8,000 bp; Figure 2c) we additionally kept all reads having a length between 7,500 bp
150 and 8,500 bp. After size selection, *PoreChop* (v.0.2.3) was used to trim remaining adapter and barcode
151 sequences.

152

153 **Figure 2. Genetic structure of the *SLC14A1* gene encoding the JK blood group system, and details on**
154 **the identified large novel deletion.** (a) *SLC14A1* gene showing exons, coding DNA sequence (CDS),
155 newly identified ~5 kb deletion (red), positions of long-range PCRs (LR1/LR2), as well as overlap
156 between both amplicons. (b) Region of ~5 kb deletion with breakpoints and primer positions used for
157 bridge-PCR. (c) Gel electrophoresis of long-range PCR products (LR1 and LR2) of all samples. (d) Gel
158 electrophoresis of bridge-PCR products for the three samples with the novel deletion and wild-type
159 control reactions.




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 was performed using *Medaka* (v.1.2.3). As a consequence of the high coverage, threshold quality for calling indels and SNVs (default: Q9 and Q8, respectively) was increased to Q10. Presence of SVs was investigated 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 FASTA format.

**Variant confirmation by Sanger sequencing and bridge-PCR**

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

181 Finally, a bridge-PCR assay was designed to confirm a large deletion found with nanopore sequencing
182 in three samples. One primer pair was designed to amplify over the deletion using primers located at
183 the flanking regions, with the forward primer being located upstream of the 5' breakpoint and the
184 reverse primer downstream of the 3' breakpoint. Another reverse primer was designed to amplify the
185 wild type allele, enabling assessment of the heterozygosity status of the deletion. This reverse primer
186 was designed to bind within the region of the deletion and amplify with the upstream primer targeting
187 the 5' breakpoint. For both primer pairs, three wild type controls were used.

188 **Results**

189 **Observed genotype-phenotype discrepancies**

190 Congruence of JK phenotypes deduced from genotypes of seven years of high-throughput donor
191 screening and serological typing was very high. In detail, of the 35,954 donors genotyped by MALDI-
192 TOF MS, 35,937 (99.95%) passed quality control and were assigned to *JK*01* and/or *JK*02* alleles using
193 the SNV c.838G>A. Among them, 50.43% (n = 18,124) were *JK*01/02* heterozygous, 26.30% (n = 9,452)
194 *JK*01/01* homozygous, and 23.27% (n = 8,361) *JK*02/02* homozygous, as expected under Hardy-
195 Weinberg equilibrium (P>0.05). Out of 11,972 donors with available serological data, we observed
196 seven samples where the null allele causing variants included in the MALDI-TOF MS module explained
197 their null phenotype (n = 4 for *JK*01N.03* and n = 3 for *JK*02N.01*). There were only 10 discrepant
198 cases (0.08%) in which MALDI-TOF MS genotyping did not agree with the observed phenotype.
199 Phenotypes and genotypes in all of them were confirmed by additional serological analyses and PCR-
200 SSP.

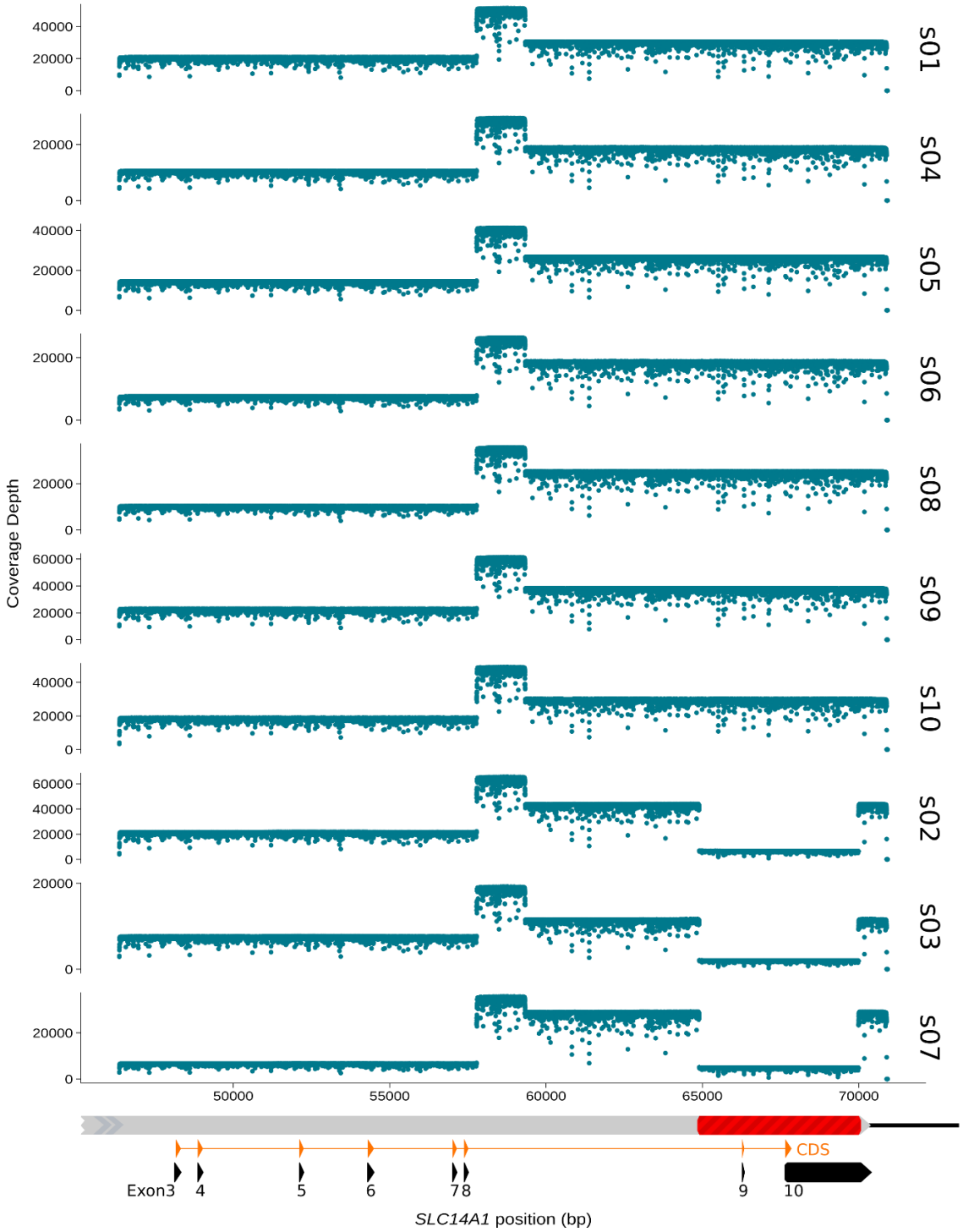
201 **Nanopore sequencing output**

202 We obtained very high coverage for each amplified fragment in all 10 samples with discordant serology
203 and genotype (Figure 3). Overall, the nanopore sequencing run produced ~1 million reads (≥ Q7) with
204 a median PHRED score of 12.7. After demultiplexing and filtering reads according to fragment sizes,
205 mean number (± standard deviation) of reads per barcode (i.e. sample) was 41,640 (± 15,182).

206 Mapping of filtered reads against the *SLC14A1* reference sequence showed a drop in coverage over a
207 length of 5 kb in the second fragment for three samples (Figure 3). For these same samples (s02, s03
208 and s07), the gel electrophoresis for verifying PCR amplification successes prior to sequencing had also
209 already revealed the presence of shorter fragments in the second LR-PCR (Figure 2c).

210

211 **Figure 3. Nanopore sequencing coverage by nucleotide position for each sample.** Coverages were
212 computed by mapping length-filtered reads to *SLC14A1* reference sequence (NG_011775.4) using
213 minimap2. For readability, samples showing a drop in coverage in LR2 were grouped together.
214 Graphical representation of mapping location on *SLC14A1* gene is given underneath the coverage
215 plots. Coverage scales are not fixed.


Coverage Depth
s01
s04
s05
s06
s08
s09
s10
s02
s03
s07
50000
55000
60000
65000
70000
CDS
Exon3
4
5
6
7
8
9
10
SLC14A1 position (bp)

216

217

218 **Resolving discrepant cases**

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

228

229 **Table 1. Summary of serology, genotyping, and nanopore sequencing for all ten genotype-**
230 **phenotype discrepancy cases.** Discrepant antigens, based on comparison of observed with deduced
231 phenotype, as well as causative alleles are highlighted in red.

| 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*† |

232 § Novel alleles not yet described. † Allele harbors at least one additional silent SNV not reported in
233 corresponding ISBT allele.

Three samples did not show any promising candidates when calling SNVs and short indels. However, SV calling identified a large 5,095 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 exactness of these breakpoint positions (Figure 2d). In all three samples, the deletion was located on a *JK*01* allele (c.838G), and donors were *JK*01*/*JK*02* heterozygotes.

## 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, greatly facilitating the identification of the underlying genetic basis for the observed genotype-phenotype discrepancies and directly producing 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). 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 *JK01/02* genotyping. Over seven years of MALDI-TOF MS high-throughput donor genotyping, only two known weak alleles were flagged as suggestively causative for discrepant

265 cases, i.e. resulting in the absence of observable levels of agglutination with respective anti-Jk
266 antibodies. One was the aforementioned variant c.742G>A, defining the *JK*01W.05* allele. This allele
267 is currently listed as 'weak', but has already been reported to cause null phenotypes by others [24].
268 The other variant was c.130G>A, defining the *JK*02W.03* allele. This variant is a common SNV (MAF >
269 0.05; rs2298720), which also defines the most frequent weak allele on the *JK*01* background
270 (*JK*01W.01*) along with several other weak alleles [13]. Since we did not observe further discrepancies
271 pointing to weak alleles harboring this variant, there is little evidence that expression levels of
272 *JK*02W.03* commonly fall below the detection threshold, pointing to an exceptional case reported
273 here. Without phenotyping methods offering higher sensitivity, however, we cannot conclusively
274 determine whether expression levels caused by these two presumably weak alleles were just below
275 our detection threshold or if those two alleles indeed caused true null phenotypes in our samples.
276 Efforts for in-depth phenotyping using absorption/elution techniques are ongoing.

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

286 Our most remarkable finding in this study was the discovery of a large structural variant that was
287 underlying three of the observed genotype-phenotype discrepancies. Our nanopore sequencing
288 strategy revealed a large deletion of approximately 5 kb spanning from exon 9 to 10 on the *JK*01*
289 allele. This deletion represents, to the best of our knowledge, the largest SV ever documented within
290 the JK blood group system. Currently, the Kidd ISBT allele table only encompasses one deletion of ~1.6
291 kb defining the *JK*01N.01* allele. First reported in Tunisian women [27], it has been sporadically
292 reported in several other populations [18,28]. This deletion (c.1_341del) spans over exons 3 and 4 of
293 the *SLC14A1* gene [27] and results in the lack of translation of the Kidd protein as the start codon is
294 located in exon 3. Considering the complete absence of exon 9 and a portion of exon 10 in alleles
295 carrying our newly discovered large deletion, it is highly probable that the resulting protein is also not
296 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 [29] 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 7,500 and 22,600 SVs per genome were observed [30,31]. Additionally, SVs are responsible for an estimated ~30% of rare heterozygous (MAF < 1%) gene inactivation events per individual [30]. 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 harbouring 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.

## Author Contributions

Conceptualization, M.G., G.A.T., S.M. and M.P.M.-G.; methodology, M.G., G.A.T., N.T., L.S., S.S., N.L. and Y.M.; software, M.G. and G.A.T.; validation, M.G., G.A.T., N.T., L.S., S.M. and M.P.M.-G.; formal analysis, M.G. and G.A.T.; resources, C.E., B.M.F., C.G. and S.M.; writing – original draft preparation, M.G.; writing – review & editing, M.G., G.A.T. and M.P.M.-G.; visualization, M.G.; supervision, S.M. and M.P.M.-G. All authors read and approved the final manuscript.

## Funding Information

This research received no external funding.

## Institutional Review Board Statement

Not applicable. (According to the Swiss cantonal and national legislation, molecular blood group analyses are no subject to ethical authorization).

## Informed Consent Statement

General informed consent was obtained from all subjects involved in the study.

## Data Availability Statement

The identified 5,095 bp deletion (NG_011775.4:g.64898_69992del; NM_015865.7:c.947-1453_*2066del) on the *JK*01* haplotype has been submitted to ClinVar (accession number: VCV001202625.1). Full-length haplotype sequences of all sequenced alleles, covering the entire coding region of SLC14A1, have been submitted to GenBank (accession numbers: PP034563-PP034582).

## Conflict of Interest

C.G. acts as a consultant to Inno-Train GmbH, Kronberg im Taunus, Germany. C.G. holds the European and US patents P3545102 and US20190316189 on the “Determination of the genotype underlying the S-s-U- phenotype of the MNSs blood group system”.

All other authors declare no conflict of interest.

340

341 **References**


342 1. Nurk, S.; Koren, S.; Rhie, A.; Rautiainen, M.; Bzikadze, A. V.; Mikheenko,
343 A.; Vollger, M.R.; Altemose, N.; Uralsky, L.; Gershman, A.; et al. The
344 Complete Sequence of a Human Genome. *Science* **2022**, *376*, 44–53,
345 doi:10.1126/science.abj6987.

346 2. Lane, W.J.; Gleadall, N.S.; Aeschlimann, J.; Vege, S.; Sanchis-Juan, A.;
347 Stephens, J.; Cone Sullivan, J.; Mah, H.H.; Aguad, M.; Smeland-Wagman,
348 R.; et al. Multiple GYPB Gene Deletions Associated with the U−
349 Phenotype in Those of African Ancestry. *Transfusion* **2020**, *60*, 1294–
350 1307, doi:10.1111/trf.15839.

351 3. Zhang, Z.; An, H.H.; Vege, S.; Hu, T.; Zhang, S.; Mosbruger, T.; Jayaraman,
352 P.; Monos, D.; Westhoff, C.M.; Chou, S.T. Accurate Long-Read Sequencing
353 Allows Assembly of the Duplicated RHD and RHCE Genes Harboring
354 Variants Relevant to Blood Transfusion. *Am. J. Hum. Genet.* **2022**, *109*,
355 180–191, doi:10.1016/j.ajhg.2021.12.003.

356 4. Montemayor, C.; Simone, A.; Long, J.; Montemayor, O.; Delvadia, B.;
357 Rivera, R.; Lewis, K.L.; Shahsavari, S.; Gandla, D.; Dura, K.; et al. An Open-
358 source Python Library for Detection of Known and Novel Kell, Duffy and
359 Kidd Variants from Exome Sequencing. *Vox Sang.* **2021**, *116*, 451–463,
360 doi:10.1111/vox.13035.

361 5. Thun, G.A.; Gueuning, M.; Sigurdardottir, S.; Meyer, E.; Gourri, E.;
362 Schneider, L.; Merki, Y.; Trost, N.; Neuenschwander, K.; Engström, C.; et
363 al. Novel Regulatory Variant in ABO Intronic RUNX1 Binding Site Inducing
364 A3 Phenotype. *bioRxiv* **2023**, 1–11, doi:10.1101/2023.05.04.539366.

365 6. Isa, K.; Takada, S.; Takeda, H.; Tsuneyama, H.; Ogasawara, K.; Takahashi,
366 D.; Miyazaki, T.; Miyata, S.; Satake, M. Two New <scp> *JK* </Scp> Silencing
367 Alleles Identified by Single Molecule Sequencing with <scp>20-Kb</Scp>
368 Long-reads. *Transfusion* **2023**, *63*, 1441–1446, doi:10.1111/trf.17397.

369 7. Meyer, S.; Vollmert, C.; Trost, N.; Brönnimann, C.; Gottschalk, J.; Buser,
370 A.; Frey, B.M.; Gassner, C. High-Throughput Kell, Kidd, and Duffy Matrix-
371 Assisted Laser Desorption/Ionization, Time-of-Flight Mass Spectrometry-
372 Based Blood Group Genotyping of 4000 Donors Shows Close to Full
373 Concordance with Serotyping and Detects New Alleles. *Transfusion* **2014**,
374 *54*, 3198–3207, doi:10.1111/trf.12715.

375 8. Thun, G.A.; Gueuning, M.; Mattle-Greminger, M. Long-Read Sequencing
376 in Blood Group Genetics. *Transfus. Med. Hemotherapy* **2023**, *50*, 184–
377 197, doi:10.1159/000530652.

378 9. Gueuning, M.; Thun, G.A.; Wittig, M.; Galati, A.-L.; Meyer, S.; Trost, N.;
379 Gourri, E.; Fuss, J.; Sigurdardottir, S.; Merki, Y.; et al. Haplotype Sequence
380 Collection of ABO Blood Group Alleles by Long-Read Sequencing Reveals
381 Putative A1 -Diagnostic Variants. *Blood Adv.* **2023**, *7*, 878–892,
382 doi:10.1182/bloodadvances.2022007133.

383 10. Wang, Y.; Zhao, Y.; Bollas, A.; Wang, Y.; Au, K.F. Nanopore Sequencing
384 Technology, Bioinformatics and Applications. *Nat. Biotechnol.* **2021**, *39*,
385 1348–1365, doi:10.1038/s41587-021-01108-x.

386 11. Makałowski, W.; Shabardina, V. Bioinformatics of Nanopore Sequencing.
387 *J. Hum. Genet.* **2020**, *65*, 61–67, doi:10.1038/s10038-019-0659-4.

388 12. Olson, N.D.; Wagner, J.; McDaniel, J.; Stephens, S.H.; Westreich, S.T.;

389 Prasanna, A.G.; Johanson, E.; Boja, E.; Maier, E.J.; Serang, O.; et al.
390 PrecisionFDA Truth Challenge V2: Calling Variants from Short and Long
391 Reads in Difficult-to-Map Regions. *Cell Genomics* **2022**, *2*, 100129,
392 doi:10.1016/j.xgen.2022.100129.

393 13. International Society of Blood Transfusion Available online:
394 https://www.isbtweb.org/isbt-working-parties/rcibgt.html (accessed on
395 10 November 2023).

396 14. da Cunha Gomes, E.G.; Machado, L.A.F.; de Oliveira, L.C.; Neto, J.F.N. The
397 Erythrocyte Alloimmunisation in Patients with Sickle Cell Anaemia: A
398 Systematic Review. *Transfus. Med.* **2019**, *29*, 149–161,
399 doi:10.1111/tme.12543.

400 15. Dean, L. Chapter10: The Kidd Blood Group. In *Blood groups and red cell*
401 *antigens*; 2005.

402 16. Gassner, C.; Meyer, S.; Frey, B.M.; Vollmert, C. Matrix-Assisted Laser
403 Desorption/Ionisation, Time-of-Flight Mass Spectrometry–Based Blood
404 Group Genotyping—The Alternative Approach. *Transfus. Med. Rev.* **2013**,
405 *27*, 2–9, doi:10.1016/j.tmrv.2012.10.001.

406 17. Meyer, S.; Trost, N.; Frey, B.M.; Gassner, C. Parallel Donor Genotyping for
407 46 Selected Blood Group and 4 Human Platelet Antigens Using High-
408 Throughput MALDI-TOF Mass Spectrometry BT - Molecular Typing of
409 Blood Cell Antigens. In; Bugert, P., Ed.; Springer New York: New York, NY,
410 2015; pp. 51–70 ISBN 978-1-4939-2690-9.

411 18. Irshaid, N.M.; Eicher, N.I.; Hustinx, H.; Poole, J.; Olsson, M.L. Novel Alleles
412 at the JK Blood Group Locus Explain the Absence of the Erythrocyte Urea

Transporter in European Families*. *Br. J. Haematol.* **2002**, *116*, 445–453, doi:10.1046/j.1365-2141.2002.03238.x.

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, doi:10.1046/j.1537-2995.2000.40010069.x.

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

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 <scp>Kidd, Duffy, and Kell</Scp> Antigens in Individuals with Sickle Cell Disease Using Whole Genome Sequencing Data from the <scp>NHLBI TOPMed Program</Scp>. *Transfusion* **2021**, *61*, 603–616, doi:10.1111/trf.16204.

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, doi:10.1159/000525326.

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.

24. Gaur, L.; Posadas, J.; Teraaur, G.; Hamura, A.; Gile, P.; Nakaya, S. Molecular Diversity of the JK Null Phenotype.; Vox Sang, 2010; p.

437 99(Suppl 1):371.

438 25. Chen, S.; Francioli, L.C.; Goodrich, J.K.; Collins, R.L.; Kanai, M.; Wang, Q.;
439 Alföldi, J.; Watts, N.A.; Vittal, C.; Gauthier, L.D.; et al. A Genome-Wide
440 Mutational Constraint Map Quantified from Variation in 76,156 Human
441 Genomes. *bioRxiv* **2022**, 2022.03.20.485034,
442 doi:10.1101/2022.03.20.485034.

443 26. Ramsey, G.; Sumugod, R.D.; Lindholm, P.F.; Zinni, J.G.; Keller, J.A.; Horn,
444 T.; Keller, M.A. A Caucasian JK*A/JK*B Woman with Jk(A+b-) Red Blood
445 Cells, Anti-Jkb, and a Novel JK*B Allele c.1038delG. *Immunohematology*
446 **2016**, *32*, 91–95.

447 27. Lucien, N.; Chiaroni, J.; Cartron, J.-P.; Bailly, P. Partial Deletion in the JK
448 Locus Causing a Jknull Phenotype. *Blood* **2002**, *99*, 1079–1081,
449 doi:10.1182/blood.V99.3.1079.

450 28. Wester, E.S.; Johnson, S.T.; Copeland, T.; Malde, R.; Lee, E.; Storry, J.R.;
451 Olsson, M.L. Erythroid Urea Transporter Deficiency Due to Novel JKnull
452 Alleles. *Transfusion* **2008**, *48*, 365–372, doi:10.1111/j.1537-
453 2995.2007.01532.x.

454 29. Liwski, R.; Clarke, G.; Cheng, C.; Abidi, S.S.R.; Abidi, S.R.; Quinn, J.G.
455 Validation of a Flow- <scp>cytometry-based</Scp> Red Blood Cell
456 Antigen Phenotyping Method. *Vox Sang.* **2023**, *118*, 207–216,
457 doi:10.1111/vox.13401.

458 30. Collins, R.L.; Brand, H.; Karczewski, K.J.; Zhao, X.; Alföldi, J.; Francioli, L.C.;
459 Khera, A. V.; Lowther, C.; Gauthier, L.D.; Wang, H.; et al. A Structural
460 Variation Reference for Medical and Population Genetics. *Nature* **2020**,

*581*, 444–451, doi:10.1038/s41586-020-2287-8.

31. 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 3,622 Icelanders Provides Insight into the Role of Structural Variants in Human Diseases and Other Traits. *Nat. Genet.* **2021**, *53*, 779–786, doi:10.1038/s41588-021-00865-4.